**The Sourcebook of Listening Research:
Methodology and Measures**

The Sourcebook of Listening Research: Methodology and Measures

Edited by

Debra L. Worthington

and

Graham D. Bodie

WILEY Blackwell

This edition first published 2018

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office
350 Main Street, Malden, MA 02148-5020, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Worthington, Debra L., editor. | Bodie, Graham D., editor.
Title: The sourcebook of listening research : methodology and measures /
edited by Debra L. Worthington and Graham D. Bodie.
Description: Hoboken, NJ : John Wiley & Sons, Inc., [2017] | Includes bibliographical
references and index.
Identifiers: LCCN 2017015025 (print) | LCCN 2017027033 (ebook) | ISBN 9781119102939 (epdf) |
ISBN 9781119102960 (epub) | ISBN 9781119103073 (cloth)
Subjects: LCSH: Listening. | Listening comprehension.
Classification: LCC BF323.L5 (ebook) | LCC BF323.L5 S68 2017 (print) | DDC 153.7/33–dc23
LC record available at https://lccn.loc.gov/2017015025

Cover image: © helenecanada/Gettyimages
Cover design: Wiley

Set in 10/12pt Warnock by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

Contents

Notes on Contributors *xi*
Preface *xxii*

Section One Decisions for the Measurement of Listening *1*

Chapter 1 **Defining Listening: A Historical, Theoretical, and Pragmatic Assessment** *3*
Debra L. Worthington and Graham D. Bodie

Section Two Methodological Trends, Challenges, and Opportunities in Listening Research *19*

Chapter 2 **Measuring Listening** *21*
Graham D. Bodie and Debra L. Worthington

Chapter 3 **Qualitative Approaches and Listening Research** *45*
Barbara Cook Overton

Chapter 4 **Modeling and Measuring Cognitive Components of Listening** *70*
Debra L. Worthington

Chapter 5 **Measuring Affective Components of Listening** *97*
Graham D. Bodie and Susanne M. Jones

Chapter 6 **Measuring Behavioral Components of Listening** *123*
Graham D. Bodie

Section Three Measurement Profiles *151*

Profile 1 **The Academic Listening Self-rating Questionnaire (ALSA)** *153*
Vahid Aryadoust and Christine C. M. Goh

Profile 2 **Active - Empathic Listening Scale (AELS)** *161*
Shaughan A. Keaton

Profile 3 **Active Listening Attitude Scale (ALAS)** *167*
Brock Bybee and Jonathon Frost

Profile 4 **Active Listening Observation Scale (ALOS)** *174*
Andrea J. Vickery

Profile 5 **Affectionate Communication Scale (TAS) (Trait-Given & Trait-Received)** *180*
Kory Floyd and Mark Alan Generous

Profile 6 **Affectionate Communication Index (ACI)** *186*
Kory Floyd and Lisa J. van Raalte

Profile 7 **Attributional Complexity Scale (ACS)** *191*
Michael Navarro

Profile 8 **Audio Message Complexity: Audio Content Change (Acc) and Audio Information Introduced (Aii)** *198*
Robert F. Potter and Annie Lang

Profile 9 **Communicative Adaptability Scale (CAS)** *204*
Gwen A. Hullman

Profile 10 **Communication Competency Assessment Instrument (CCAI)** *212*
Phil Backlund

Profile 11 **Communication Functions Questionnaire (CFQ-30)** *217*
Graham D. Bodie

Profile 12 **Conversational Listening Span (CLS)** *224*
Debra L. Worthington and Courtney Edwards

Profile 13 **Conversational Sensitivity Scale (CSS)** *232*
Debra L. Worthington

Profile 14 **The Couples Helping Exercise** *239*
Andrew C. High, Jenny L. Crowley, and Emily M. Buehler

Profile 15 **Doctors' Interpersonal Skills Questionnaire (DISQ)** *246*
Helen Meldrum

Profile 16 **Effective Listening and Interactive Communication Scale (ELICS)** *252*
Gillian King and Michelle Servais

Profile 17 **Empathic Accuracy: Standard Stimulus Paradigm (EA-SSP)** *259*
Vivian Ta and William Ickes

Profile 18 **Empathic Accuracy: Unstructured Dyadic Interaction Paradigm (EA-UDIP)** *265*
Vivian Ta and William Ickes

Profile 19 **Facilitating Listening Scale (FLS)** *272*
Avraham N. Kluger and Osnat Bouskila-Yam

Profile 20 **Five Facet Mindfulness Questionnaire (FFMQ)** *281*
Carly M. Danielson and Susanne M. Jones

Profile 21 **Functional Magnetic Resonance Imaging (fMRI)** *290*
Iballa Burunat and Elvira Brattico

Profile 22 **Highly Sensitive Person Scale (HSPS)** *299*
Christopher C. Gearhart

Profile 23 **HURIER Listening Profile** *306*
Margarete Imhof

Profile 24 **Informational Reception Apprehension Test (IRAT)** *313*
Shaughan A. Keaton

Profile 25 **Interaction Involvement Scale (IIS)** *319*
Debra L. Worthington

Profile 26 **Feffer's Interpersonal Decentering** *327*
Molly S. Tucker and Sharon Rae Jenkins

Profile 27 **Interpersonal Process Recall (IPR)** *334*
Christopher T. Belser

Profile 28 **Interpersonal Reactivity Index (IRI)** *340*
Shaughan A. Keaton

Profile 29 **Language Style Matching (LSM)** *348*
Kaitlin Cannava

Profile 30 **Leader-Member Exchange 7 Questionnaire (LMX-7)** *354*
Lisa K. Hanasono

Profile 31 **Listenability Style Guide (LSG)** *361*
Graham D. Bodie

Profile 32 **Listening Concepts Inventory (LCI and LCI-R)** *372*
Debra L. Worthington

Profile 33 **Listening Fidelity (LF)** *379*
Margarete Imhof

Profile 34 **Listening Practices Feedback Report (LPFR)** *387*
Debra L. Worthington

Profile 35 **Listening Span Tests** *394*
Margarete Imhof

Profile 36 **Listening Styles Profile-Revised (LSP-R)** *402*
Graham D. Bodie and Debra L. Worthington

Profile 37 **Medical Communication Competence Scale (MCCS)** *410*
Danielle Blanch-Hartigan

Profile 38 **Memory for Conversation** *419*
Jonathon Frost and Brock Bybee

Profile 39 **Metacognitive Awareness Listening Questionnaire (MALQ)** *430*
Christine C. M. Goh

Profile 40 **Metacognitive Listening Strategies Instrument (MLSI)** *438*
Laura A. Janusik

Profile 41 **Microanalysis of Face-to-Face Dialogue (MFD)** *445*
Janet Bavelas, Jennifer Gerwing, Sara Healing, and Christine Tomori

Profile 42 **Multidimensional Evaluation of Enacted Social Support (MEESS)** *453*
Daena J. Goldsmith and Abbey Griscom

Profile 43 **Multitasking While Listening** *458*
Jonathon Frost and Brock Bybee

Profile 44 **Narrative Believability Scale (NBS-12)** *465*
Graham D. Bodie

Profile 45 **Narrative Engagement Measure (NEM)** *469*
Jenny L. Crowley and Jennifer A. Jackl

Profile 46 **Weinstein Noise Sensitivity Scale (WNSS)** *475*
Debra L. Worthington

Profile 47 **Nonverbal Immediacy Measures** *482*
Jane B. Teel

Profile 48 **Normative Message Processing Scale (NMPS)** *496*
R. Kelly Aune and Rodney A. Reynolds

Profile 49 **Ordinary Conversation Scale (OCS)** *504*
Brian Lakey and Travis Sain

Profile 50 **Organizational Listening Survey (OLS)** *509*
Laura A. Janusik

Profile 51 **Perceived Partner Responsiveness Scale (PPRS)** *516*
Harry T. Reis, Dev Crasta, Ronald D. Rogge, Michael R. Maniaci, and Cheryl L. Carmichael

Profile 52 **Profile of Nonverbal Sensitivity (PONS)** *522*
Laura A. Janusik

Profile 53 **Rational-Experiential Inventory–40 (REI-40)** *530*
Shaughan A. Keaton

Profile 54 **Relational Framing** *537*
Denise Haunani Solomon and Sara K. Salmon

Profile 55 **Rhetorical Sensitivity Scale (RHETSEN)** *545*
Shaughan A. Keaton

Profile 56 **Role Category Questionnaire (RCQ)** *552*
Andrea J. Vickery

Profile 57 **Self-Perceived Listening Competence Scale (SPLCS)** *560*
Margarete Imhof

Profile 58 **Talkaholic Scale (Compulsive Communication Scale) (TAS)** *566*
Debra L. Worthington

Profile 59 **Team Listening Environment (TLE)** *573*
Michelle K. Johnston and Kendra Reed

Profile 60 **Time Studies** *578*
Graham D. Bodie

Profile 61 **The Listening Test of the Internet-Based Test of English as a Foreign Language (TOEFL iBT)** *592*
Vahid Aryadoust

Profile 62 **Verbal Aggressiveness Scale (VAS)** *599*
Timothy R. Levine

Profile 63 **Taxonomy of Verbal Response Modes (VRM)** *605*
William B. Stiles

Profile 64 **Watson-Barker Listening Test (WBLT)** *612*
Debra L. Worthington

Profile 65 **Willingness to Listen (WTL)** *617*
Andrea J. Vickery

Index *624*

Notes on Contributors

Vahid Aryadoust (PhD, Nanyang Technological University, Singapore) is assistant professor at the National Institute of Education of Nanyang Technological University, Singapore. His research focuses on the application of structural equation modeling, data mining, latent trait models, and complexity. He has been a frequent presenter in international conferences and has published papers in educational journals including *Language Testing, Language Assessment Quarterly, Assessing Writing, International Journal of Testing, International Journal of Listening, Educational Psychology*, and *Educational Assessment*.

R. Kelly Aune (PhD, University of Arizona) is professor and chair of the Department of Communicology, University of Hawai'i at Mānoa, where he teaches message processing, verbal and nonverbal communication, and research methods. His research is in message processing and natural language processing. He looks at inference making and implicature in conversations, and the evolution of idiosyncratic communication across dyads and groups. Dr. Aune also examines how the cooperative nature of communication can lead to misunderstandings and deception.

Phil Backlund (PhD, University of Denver) has taught communication studies courses at Central Washington University since 1979. He is and has been active in the National Communication Association, publishing, and international teaching.

Janet Beavin Bavelas was educated at Stanford (BA, psychology; MA, communication research; PhD, psychology) and has spent her academic career at the University of Victoria, where she is currently an active emeritus professor of psychology. She has coauthored *Pragmatics of Human Communication* (1967) and *Equivocal Communication* (1990) as well as over 90 journal articles or book chapters. Bavelas and her research team conduct experiments on the unique features of face-to-face dialogue, specifically the integration of words and co-speech acts (hand and facial gestures, and gaze) and the moment-by-moment collaboration between speakers and addressees. Microanalysis of face-to-face dialogue (MFD) developed out of this program of research. More recently, the team have been applying this method and their findings to dialogues outside the lab: in psychotherapy, medicine, parent–infant interaction, and computer-mediated interaction. Academic honors include fellowships in the Royal Society of Canada, the International Communication Association, and the Canadian Psychological Association.

Christopher T. Belser (PhD, University of Central Florida) is a graduate of the Counselor Education and Supervision program at the University of Central Florida. He is a graduate teaching associate and research assistant on a federally funded research project seeking to increase undergraduate participation in science, technology, engineering, and mathematics (STEM). His primary areas of research and scholarship are career development, STEM engagement, and school counselor training and practice.

Danielle Blanch-Hartigan (PhD, Northeastern University; MPH, Harvard School of Public Health) is assistant professor of health studies in the Department of Natural and Applied Sciences at Bentley University. Her interdisciplinary research in psychology and public health aims to improve the patient healthcare experience and foster patient-centered care through better communication. Her research explores healthcare providers' ability to accurately recognize patients' verbal and nonverbal cues of emotion and pain, as well as the patient-centered care experience of cancer survivors. Danielle has published over 30 articles in peer-reviewed journals, including *Patient Education and Counseling*, *Health Communication*, *Journal of Clinical Oncology*, and *Basic and Applied Social Psychology*.

Graham D. Bodie (PhD, Purdue University), formerly Professor of Communication Studies at Louisiana State University and Agricultural & Mechanical College, is now on faculty in the Meek School of Journalism and New Media at The University of Mississippi. His research on listening and the social cognitive underpinnings of human communicative behavior has been published in *Human Communication Research*, *Communication Monographs*, *Communication Research*, *Communication Yearbook*, and the *International Journal of Listening*, among other outlets. Graham's research has been funded by the Louisiana Board of Regents and the National Science Foundation, and he regularly appears in local and national media outlets on issues relevant to listening in close relationships.

Osnat Bouskila-Yam (MA, Hebrew University) is a doctoral candidate at the School of Business at Ben-Gurion University of the Negev, under the supervision of Prof. Nurit Zaidman. Her doctoral research addresses "Listening in Teams Operating in the High Tech Industry and Its Effect on Knowledge Transfer." Osnat is the founder and CEO of Osnat Bouskila-Yam Ltd.—Management Consultancy and Organizational Development. She lectures at The Interdisciplinary Center, Hertzelia, in the Psychology School (BA) and the Organizational Behavior & Development program (MA).

Elvira Brattico (PhD, University of Helsinki) worked as a postdoctoral researcher at the EU-funded research project "Tuning the Brain for Music" (2007–2009). From 2009 to 2013, she directed the Aesthetics module of the Finnish Centre of Excellence in Interdisciplinary Music Research. In 2011, she became head of the Neuroaesthetics of Music Group at the University of Helsinki. She is currently a full professor of neuroscience, music, and aesthetics at the Department of Clinical Medicine, Aarhus University, Denmark, and co-director of the Center for Music in the Brain (MIB), a center of excellence of the Danish National Research Foundation. Her main interests cover the neural processing of perceptual and emotional features in music, and their intra- and interindividual variations.

Emily M. Buehler (MA, Wake Forest University) is a doctoral student and teaching assistant in the Department of Communication Studies at the University of Iowa. Her research explores interpersonal communication processes as they operate in computer-mediated environments. She is particularly interested in investigating supportive communication processes online. Emily has presented her research at the annual conferences of both the National Communication Association and the Central States Communication Association. Emily teaches relational communication, interpersonal communication, and privacy management courses.

Iballa Burunat (MA, University of Jyväskylä) has a background in systematic musicology and cognitive neuroscience of music. She is currently a postdoctoral researcher funded by the Academy of Finland project "Dynamics of Music Cognition." Her research focuses on how music training modulates the brain dynamics of music listening in naturalistic conditions. Recent publications examine the reliability of naturalistic functional magnetic resonance imaging (fMRI) and the functional symmetry of brain responses to music listening.

Brock Bybee (MA, Southern Utah University) is a faculty member in the Department of Professional Sales at Weber State University, and is completing his doctorate in Interpersonal Communication from Louisiana State University and Agricultural & Mechanical College. His research interests are centered on interpersonal communication deficiencies and conflicts associated with the growing influence of social media.

Kaitlin Cannava (PhD, Louisiana State University A&M) is a visiting assistant professor at Xavier University. With a background in communication studies, psychology, and linguistics, she has conducted interdisciplinary research on multiple facets of language and conversation using a range of methods from computerized textual analysis software programs to discourse analysis. Her main interests are linguistic measures of repetition, coordination, and alignment.

Cheryl L. Carmichael (PhD, University of Rochester) earned her doctorate in social-personality psychology and is assistant professor of psychology at Brooklyn College. Her research aims to better understand the ways in which social interactions between close relationship partners influence health and well-being.

Dev Crasta (MA, clinical psychology, Carnegie Mellon University) is a PhD student in clinical psychology at the University of Rochester and a National Science Foundation (NSF) Graduate Research Fellow. His research aims to improve accessibility of family therapy to underserved populations. His translational approach combines basic research into how neighborhood contexts impact relationships with clinical research investigating accessible online tools.

Jenny L. Crowley (PhD, University of Iowa) is an assistant professor at the University of Tennessee. Her research interests revolve around the experience and social support of stigmatized identities. In particular, her research examines how features of communication contribute to perceptions of stigma as well as how supportive communication facilitates coping with stigma. Her research has been published in *Communication Theory*, *Communication Research*, and the *Journal of Family Communication*.

Carly M. Danielson (MA, University of Nevada, Las Vegas) is a Ph.D. candidate at the University of Minnesota, Twin Cities. Her research focuses on bullied students' coping processes and supportive communication, as well as bullying bystander interventions. She has published research in *Health Communication and the Journal of Health Communication*.

Courtney Edwards (MA, Auburn University) is a media specialist with Auburn Global, Auburn University's pathway program for international students. She has previously taught in the School of Communication and Journalism, and her research interests include public relations, social media, and intercultural communication.

Kory Floyd (PhD, University of Arizona) is professor of communication at the University of Arizona. His research focuses on the communication of affection in personal relationships and on the interplay between interpersonal interaction and physical health. He has authored 14 books and more than 100 journal articles and book chapters on the topics of affection, psychophysiology, family communication, nonverbal behavior, and health. He is a former editor of *Journal of Family Communication* and *Communication Monographs*.

Jonathon Frost (MA, Auburn University) is a doctoral candidate in interpersonal communication and a teaching assistant at Louisiana State University A&M. His research interests include message interpretation, computer-mediated communication, communication across the lifespan, and signal detection.

Christopher C. Gearhart (PhD, Louisiana State University A&M) is an assistant professor in the Department of Communication Studies at Tarleton State University in Stephenville, Texas. Recent articles have been published in *Western Journal of Communication*, the *International Journal of Listening*, and *Communication Reports*.

Mark Alan Generous (PhD, Arizona State University) is assistant professor of communication studies at Saint Mary's College of California. His research focuses on the effects of interpersonal communication processes on relationship development and dissolution within a variety of relational contexts, including romantic relationships, casual-sex relationships, families, and student–teacher bonds. He has published several book chapters and articles in journals such as *Journal of Family Communication*, *Communication Reports*, *Death Studies*, and *Journal of Social Work in End-of-Life & Palliative Care*.

Jennifer Gerwing (PhD, Psychology, University of Victoria, Canada) uses microanalysis in both lab experiments and applied settings. Her research focuses primarily on the effects of face-to-face dialogue on hand and facial gestures. She applies microanalysis in hospital settings, including her present position as a senior researcher at the Health Services Research Center at Akershus University Hospital, Oslo, Norway. She is particularly interested in multimodality, specifically the semantic role that hand and facial gestures play in clinical interactions and how these modalities integrate with speech to present essential information.

Christine C. M. Goh (PhD, Lancaster University) is a professor of linguistics and language education and holds a concurrent appointment as Dean of the Graduate Studies and Professional Learning at the National Institute of Education, Nanyang Technological University, Singapore. She contributes actively to academic and professional discussions on the teaching and assessment of listening and speaking, and has a strong interest in the role of metacognition in language development and use. She has researched and published extensively in these areas over the past 20 years.

Daena J. Goldsmith (PhD, University of Washington) is a professor of rhetoric and media studies at Lewis & Clark College. She studies how we enact identities and relationships in everyday interactions, including conversations between spouses or partners, advice giving among friends and family, and storytelling face-to-face and online. Her two books and many of her recent articles focus on how couples or families talk about illness or disability.

Abbey Griscom (B.A., Lewis & Clark) has research interests in the areas of rhetoric, media studies and gender studies.

Lisa K. Hanasono (PhD, Purdue University) is an assistant professor in the School of Media and Communication at Bowling Green State University. Her research examines how people communicate and cope with discrimination in interpersonal, organizational, intercultural, and computer-mediated contexts. She also examines how supportive communication, community building, and institutional change can help targets of discrimination. Her work has been published in journals such as *Human Communication Research*, the *Journal of International and Intercultural Communication*, *Communication Research*, and *Sex Roles*.

Sara Healing's (BA, hons., MSc, Psychology, University of Victoria, Canada) honors thesis addressed the effects of two different lines of questioning about the same task. For her MSc thesis, she developed a microanalysis that identified the unique information that an individual patient can contribute to oncology consultations. Her primary research interests are using microanalysis to study face-to-face dialogue, and publications have addressed hand and facial gestures and bad-news delivery. As part of International Microanalysis Associates, she teaches international professional workshops on microanalysis and communication research.

Andrew C. High (PhD, The Pennsylvania State University) is an assistant professor in the Department of Communication Studies at the University of Iowa. His research interests focus on the interconnections between supportive communication and computer-mediated communication. In particular, his research investigates whether and how different channels or features of computer-mediated communication influence the processes of seeking, providing, and receiving support. His recent research has been published in outlets such as *Human Communication Research*, *Communication Monographs*, *Communication Research*, and *Computers in Human Behavior*.

Gwen A. Hullman (PhD, Kent State University) is associate professor and Chair of the Department of Communication Studies at the University of Nevada, Reno. She teaches courses in interpersonal communication, conflict resolution, alternative dispute resolution, listening, and research methods. Her research interests include mediation, message planning, cognitive aspects of interpersonal communication, and conversational goals. She has been a volunteer mediator for the past 7 years and is a member of the Nevada Dispute Resolution Coalition.

William Ickes (PhD, University of Texas at Arlington) is Distinguished Professor of Psychology at the University of Texas at Arlington. He is the editor of *Empathic Accuracy* (1997) and the coeditor (with Jean Decety) of the *Social Neuroscience of Empathy* (2009). He is the author of *Everyday Mind Reading: Understanding What Other People Think and Feel* (2003) and *Strangers in a Strange Lab: How Personality Shapes Our Initial Encounters With Others* (2009). His research on empathic accuracy has been recognized by three international research awards. His book *Strangers in a Strange Lab* received the International Association for Relationship Research Book Award in 2012.

Margarete Imhof (PhD, University of Bamberg) is a professor of psychology in education at the Johannes Gutenberg University in Mainz, Germany. Her primary research interest is in listening in educational contexts, such as instruction and teacher training. She has published in the *International Journal of Listening* and the *Journal of Nonverbal Behavior*, among others. She won the Nichols Award for her paper "What Makes a Good Listener?" She served as the president of the International Listening Association (2007–2008) and has been the editor of the *International Journal of Listening* since 2012.

Jennifer A. Jackl (PhD, University of Iowa) is a visiting assistant professor at Roanoke College. She studies the communication of narratives within the context of interpersonal and family relationships. Her current research focuses on the content of and communicative processes surrounding (not) telling untellable tales and how these stories impact the ways in which individuals make sense of their identity, cope with challenges, and make meaning of their life and relationships.

Laura A. Janusik (PhD, University of Maryland; MBA, Rockhurst University), associate professor and the McGee Chair of Communication, Rockhurst University, has received top paper awards from divisions of the National Communication Association, International Listening Association (ILA), and Central States Communication Association. She is a recipient of the Nichols Award and the Brennar Award, presented for scholarly achievement by the ILA and Rockhurst University, respectively. Her research interests include listening cognition, listening metacognitions, pedagogy, and intercultural listening. She is a past president of the ILA and is published nationally and internationally.

Sharon Rae Jenkins (PhD, Boston University), a professor of clinical psychology at the University of North Texas, specializes in personality assessment in the context of culture. She has a PhD in personality psychology from Boston University and a postdoctoral certificate in clinical psychology from the California School of Professional Psychology, Berkeley/Alameda (CSPP-B/A). She was a National Institute of Mental Health postdoctoral fellow, and was an assistant research psychologist at the University of California, Berkeley, Institute of Human Development. She has taught at the University of California, Santa Cruz, and at CSPP-B/A. She recently edited the *Handbook of Clinical Scoring Systems for Thematic Apperceptive Techniques* (Erlbaum, 2008).

Michelle K. Johnston (PhD, Louisiana State University A&M) is a Chase Minority Entrepreneurship Professor of Management at Loyola University New Orleans. She teaches undergraduate and graduate courses in leadership and strategic communication and has published numerous articles in listening, particularly focused on listening environments. Her work can be found in the *Journal of Business Research*, *Journal of Business Communication*, and *International Journal of Listening*.

Susanne M. Jones (PhD, Arizona State University) is an associate professor of communication studies at the University of Minnesota, Twin Cities. Her research focuses on the influence of verbal and nonverbal supportive communication on cognition and emotion regulation. Her research has appeared in *Communication Monographs*, *Communication Research*, and *Human Communication Research*; in interdisciplinary journals, such as *Mindfulness*; and in several edited volumes.

Shaughan A. Keaton (PhD, Louisiana State University A&M) is an assistant professor of communication studies at Young Harris College in Young Harris, Georgia. His research specializes in listener goals and communicator styles, intergenerational and intercultural communication, and social psychology and sport communication. Publications feature an edited volume, *The Influence of Communication in Physiology and Health*.

Gillian King (PhD, University of Western Ontario) is senior scientist at Bloorview Research Institute in Toronto, Canada, with an appointment in the Department of Occupational Science and Occupational Therapy at the University of Toronto. She holds the Canada Research Chair in Optimal Care for Children with Disabilities. Her research interests concern the psychosocial aspects of pediatric services, including child and parent wellness, the participation of children with disabilities, client engagement, models of service delivery, service providers' expertise, and interprofessional practice.

Avraham N. Kluger (PhD, Stevens Institute of Technology) is a professor of organizational behavior at the School of Business at the Hebrew University. His research on the detrimental effects of feedback on performance won him awards both from the Organizational Behavior division of the Academy of Management and from the Society of Industrial and Organizational Psychology. Currently, he is focusing on employing meta-analytic techniques on listening-related research, including forces that may interfere with listening (loss of dominance, secondhand trauma, and avoidance-attachment style). He is also working to identify and develop tools to teach effective-listening practices.

Brian Lakey (PhD, Indiana University) is a professor of psychology at Grand Valley State University. Trained as a clinical psychologist, he has devoted his career to understanding basic mechanisms in social support that can be used to develop more effective interventions. He is the author of relational regulation theory with Ed Orehek.

Annie Lang (PhD, University of Wisconsin–Madison) is a Distinguished Professor of Communication Science in the Media School at Indiana University. Her research focuses on the dynamic cognitive processing of mediated messages. She is a past editor of *Media Psychology*, fellow of the International Communication Association (ICA), and recipient of the Association for Education in Journalism and Mass Communication (AEJMC) Under-40 Award and the ICA Career Productivity Award.

Timothy R. Levine (PhD, Michigan State University) is a Distinguished Professor and Chair of Communication Studies at the University of Alabama at Birmingham and Global Professor of Communication and Media at Korea University, Seoul, South Korea. Research areas include interpersonal communication, persuasion and social influence, cross-cultural communication, and social scientific research methods. Levine has published more than 100 refereed articles related to communication, and he is an internationally recognized leader in deception research. He is the author or coauthor of information manipulation theory, truth default theory, the veracity effect, the probing effect, and the Park–Levine probability model. His research on deception has been funded by the National Science Foundation, US Department of Defense, and FBI.

Michael R. Maniaci (PhD, University of Rochester) received his PhD in social-personality psychology and is currently assistant professor of psychology at Florida Atlantic University. His research focuses on interpersonal attraction and the formation and maintenance of close relationships.

Helen Meldrum (EdD., Clark University) is an associate professor of psychology in the Program in Health Sciences and Industry, Department of Natural and Applied Sciences at Bentley University. Prior to coming to Bentley, she was associate professor of communication and psychology at Massachusetts College of Pharmacy and Health Sciences. Most of her research has focused on human relations in healthcare settings.

Michael Navarro (MA, Louisiana State University A&M) is a doctoral student in the Department of Communication Studies at LSU A&M. His research centers on team communication and leadership roles in competitive video games, as well as the philosophical foundations for the study of computer-mediated communication.

Barbara Cook Overton (PhD, Louisiana State University A&M) is an assistant professor in the Department of Languages and Communication at Southeastern Louisiana University. Her research interests include emergency room communication, electronic medical records systems, organizational change, communication and aging, ageism and health outcomes, persuasion, mass media, public health campaigns, pharmaceutical advertising, sexual health, patient education, and patient–provider interactions.

Robert F. Potter (PhD, Indiana University) is professor of communication science at The Media School at the University of Indiana. His research focuses on the impact of auditory elements on information processing of media; advanced communication science research methods, including psychophysiological and neuroscience measures as indicators of cognitive and emotional responses to media; and the concept of advertising clutter and its influence on information processing. His research has been published in *Media Psychology*, *Communication Research*, *Cognition and Emotion*, *Journal of Broadcasting & Electronic Media*, *Journal of Advertising*, *Journal of Advertising Research*, and *Journal of Management Information Systems*.

Kendra Reed (PhD, University of Nebraska–Lincoln) is an associate professor of management and the Barry and Teresa LeBlanc Professor in Business Ethics at Loyola University New Orleans. Her research and teaching interests focus on factors impacting employee performance, including climate, efficacy, values, and change, with articles published in the *Journal of Business Research*, *Journal of Managerial Issues*, *Journal of*

Leadership Studies, and *Journal of Business Communication*. Prior to an academic career, she spent 7 years working in the human resource management profession.

Harry T. Reis (PhD, New York York University) earned his doctorate in social-personality psychology and has taught at the University of Rochester since 1974. He has published more than 200 empirical articles and chapters. Among many honors, he is recipient of the Distinguished Research Career Award from the International Association for Relationship Research and from the Society for Personality and Social Psychology.

Rodney A. Reynolds (PhD, Michigan State University) is the accreditation liaison officer (ALO) and director of educational effectiveness and institutional research at California Lutheran University. His research publications center on message creation and processing. He has a number of publications on communication measurement issues and instruments.

Ronald D. Rogge (PhD, UCLA) received his PhD in clinical psychology. He is an associate professor at the University of Rochester, where he studies factors that contribute to marital discord and breakup. He is a co-developer of the Compassionate and Accepting Relationships through Empathy (CARE) program.

Travis Sain (MA, Southern Illinois University) is a PhD student in clinical psychology at Southern Illinois University. His research examines the efficacy of computer-based Acceptance and Commitment Therapy (ACT) interventions.

Sara K. Salmon (MA, The Pennsylvania State University) supports nonprofit agencies through philanthropic giving in the Greater Los Angeles Area. Her research focuses on sexual disclosures in intimate relationships.

Michelle Servais (PhD, University of Western Ontario) is a researcher, educator, and knowledge exchange professional on the Thames Valley Children's Centre's Quality Management Team, which focuses on enhancing service delivery and the development of professional expertise. Michelle coaches clinicians to develop their research capacity and to use evidence to inform their daily practices. Research interests include family-centered care, relationship-centered practice, solution-focused coaching, interdisciplinary collaboration and learning, clinical listening, professional development and educational interventions, and measure development.

Denise Haunani Solomon (PhD, Northwestern University), a liberal arts research professor of communication arts and sciences at The Pennsylvania State University, researches communication experiences in personal relationships, such as support and conflict, which promote or erode well-being. She and her colleagues developed the relational turbulence model to describe how transitions in relationships complicate emotions, cognitions, and communication. Her work examines how transitions in romantic relationships promote relationship qualities that polarize reactions to both ordinary and extraordinary experiences. She and Jennifer Theiss coauthored *Interpersonal Communication: Putting Theory Into Practice.*

William B. Stiles (PhD, University of California, Los Angeles) is professor emeritus of psychology, Miami University, Oxford, Ohio, and adjunct professor of psychology at

Appalachian State University, Boone, North Carolina. He received his PhD from UCLA in 1972, has been president of Division 29 (Psychotherapy) of the American Psychological Association and of the Society for Psychotherapy Research, and has served as editor of *Psychotherapy Research* and *Person-Centered and Experiential Psychotherapies*. He has published more than 300 journal articles and book chapters, most dealing with psychotherapy, verbal interaction, and research methods (http://www.users.miamioh.edu/stileswb/).

Vivian Ta (MSc, University of Texas at Arlington) is a PhD candidate in experimental psychology at the University of Texas at Arlington, a fellow in the National Science Foundation's Louis Stokes Alliances for Minority Participation (LSAMP) Bridge to the Doctorate program, and has received recognition for her research from various psychological associations. Her research focuses on investigating the role of latent semantic similarity in face-to-face and computer-mediated communication. Other research interests include close relationships, personality, online dating, and discourse analysis.

Jane B. Teel (PhD, Auburn University) is an assistant clinical professor in the Department of Educational Foundations, Leadership, and Technology at Auburn University. Her research focuses on nonverbal communication, organizational communication, leadership, learning styles, videoconferencing, and blended learning.

Christine Tomori's (MSc, University of Victoria, Canada) honors thesis used microanalysis to compare the communication patterns of therapists using two different approaches. Her MSc thesis developed the Patient-Centred Assessment of Symptoms and Activities (P-CASA). As part of Victoria Microanalysis Associates, Tomori has developed and taught international professional workshops on microanalysis and contemporary communication research for practitioners, including psychotherapists, coaches, and physicians. She is currently the principal of Tomori Solutions, Ltd. (http://www.tomorisolutions.com), which provides research, project management, and professional development within the fields of healthcare administration and communicative processes in applied settings (workplace, health, and social services).

Molly S. Tucker (MS, University of North Texas) is a doctoral student in the Clinical Psychology program at the University of North Texas. Research interests include the value of romantic relationships, and the physiological and psychological consequences of relationship dissolution. Publications have focused on maternal self-efficacy in Mexican women, parenting among Latino immigrant families, the impact of home visiting programs, and the relationship between perspective taking and marital satisfaction.

Lisa J. van Raalte (PhD, Arizona State University) is assistant professor of communication studies at Sam Houston State University. Her research focuses on how communication in close personal relationships influences one's psychological health, physiological health, and relational health. Lisa is an interpersonal communication scholar with expertise in quantitative methodology and analysis. Using survey and experimental design, her research interests focus on romantic relationships such as friends-with-benefits relationships and exclusive committed romantic relationships.

Andrea J. Vickery (PhD, Louisiana State University A&M) is presently on faculty at University of Richmond in the department of Rhetoric and Communication Studies. Her teaching and research interests include interpersonal communication, research methods, supportive communication, and close relationships.

Debra L. Worthington (PhD, University of Kansas), professor of communication at Auburn University, has published numerous articles in the fields of listening and legal communication. She is lead author of *Listening: Processes, Functions, and Competency*. Her research has been recognized by the American Society of Trial Consultants, the Burton Foundation, the European Communication Research and Education Association, the International Listening Association (ILA), and the National Communication Association. She is a past president of the International Listening Association.

Preface

Listening is a multidisciplinary field. The methods and measures profiled in this book were developed and continue to be employed by scholars in a variety of fields, among them: communication, education, psychology, linguistics, management, health, and neuroscience. The multidisciplinary nature of the field has resulted in a variety of methodological and theoretical approaches and a range of definitions. As you'll see in Chapter 1, listening has been defined in myriad ways, and models of listening utilize a number of distinct terms and concepts, proposing that listening consists of anywhere between three to more than a dozen separate processes (see Wolvin, 1989). Such diversity presents benefits and challenges—it has revealed the complex nature of listening, allowing researchers and practitioners more and better ways to improve this important life skill; at the same time, it has resulted in a rather fragmented field.

Although we do not advocate a one-size-fits-all approach to defining (or measuring) listening, there are distinct advantages to working within a limited number of conceptual boundaries. For one, profound differences regarding central conceptual matters can threaten the potential for scientific coherence and stymie progress. As a result, the vision of a unified field of listening becomes blurry at best—and imperceptible at worst. In 2008, we, along with Margarete Imhof and Lynn Cooper, were invited to coauthor a review for the *International Journal of Listening*. In that article, we asked, "What would a unified field of listening look like?" Among the goals outlined in that article were to begin establishing a common language for the field as well as to identify important areas of research in need of further development (Bodie, Worthington, Imhof, & Cooper, 2008). This book furthers our original goals in three important ways: (a) It provides a history of the field to help novice scholars understand its current state, (b) it begins to establish a common language (or, better, a set of common languages) for the field, and (c) it outlines the strengths and weaknesses of common (and not-so-common) methodological approaches to the study of listening.

Throughout our discussions and the process of editing the *Sourcebook*, our vision was simple: to provide a comprehensive, go-to resource for listening researchers, practitioners, and students. Although there are similar resources available for scholars of nonverbal communication (Manusov, 2004) and family communication (Turner & West, 2006), as well as more general material for those interested in self-report measures of communication (Rubin, Rubin, Graham, Perse, & Seibold, 2009), listening scholars are left to comb the pages of journals in search of ways to operationalize key listening constructs. We hope this book brings organization to what may seem at first like a daunting and arduous task of thinking through conceptual and operational decisions. This book was not written to sit pristinely on a library or office shelf. We hope that it will become a well-used, dog-eared, and highlighted (or underlined, as your preference may be) essential resource for those new to the field and for the well-established listening scholar. If so, then our goal of writing the first comprehensive sourcebook of listening measures and measurement will have become a reality.

We would like to express our deepest gratitude to everyone who directly or indirectly helped bring this *Sourcebook* to fruition. We thank our reviewers for their insights, our contributors for their knowledge and expertise, and the Wiley team for their help bringing our ideas to the page.

Finally, we thank our colleagues and students for helping us to think more creatively about measuring listening and about novel ways to conceptualize its role in daily life. Most of all, thanks to our families for putting up with the fact that we do not always practice what we preach—and for their patience as we try to become the kind of listeners they deserve.

Debra L. Worthington
Graham D. Bodie

References

Bodie, G. D., Worthington, D. L., Imhof, M., & Cooper, L. (2008). What would a unified field of listening look like? A proposal linking past perspectives and future endeavors. *International Journal of Listening, 22*, 103–122. doi: 10.1080/10904010802174867

Manusov, V. (Ed.). (2004). *The sourcebook of nonverbal measures: Going beyond words.* New York: Psychology Press.

Rubin, R. B., Rubin, A. M., Graham, E. E., Perse, E. M., & Seibold, D. R. (Eds.). (2009). *Communication Research Measures II: A sourcebook.* New York: Routledge.

Turner, L. H., & West, R. (Eds.). (2006). *The family communication sourcebook.* Thousand Oaks, CA: Sage.

Wolvin, A. D. (1989). Models of the listening process. In C. W. Roberts & K. W. Watson (Eds.), *Intrapersonal communication processes: Original essays* (pp. 508–527). New Orleans, LA: SPECTRA.

Section One

Decisions for the Measurement of Listening

If you come to this book for ready-made solutions for your research questions about listening, it will likely disappoint. Instead, we compiled this *Sourcebook* to be a valuable resource, something you can pick up to help guide decisions regarding how to study listening. The research process is full of decisions, big and small. This book seeks to provide, in a readily available and accessible format, a discussion of the important decisions you will make when studying listening.

This section includes one chapter that provides a brief history of the field of listening; addresses a primary conceptual issue, namely, how to define listening; discusses various theoretical frameworks that help situate scholarship; and provides an overview of the book.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

1

Defining Listening: A Historical, Theoretical, and Pragmatic Assessment

Debra L. Worthington and Graham D. Bodie

This chapter provides a brief historical background to illustrate how the history of listening research has affected the conceptualization and measurement of listening as well as how this skill has been taught. Like all history, we write this one from a particular perspective. Both of us have doctorates in communication studies and have spent most of our academic careers in this discipline. We have both been heavily involved in both the International Listening Association (ILA), a scholarly society devoted to the teaching, practice, and research of listening, and the National Communication Association, the largest academic organization for the advancement of communication research and practice. Thus, even aside from space limitations, our review is necessarily partial and incomplete. Our purpose here is not to provide a full history of the field, but to provide enough of a background so that readers can understand the development of measurement practices that have shaped our understanding of listening.

We first trace the research as it advanced an expanding definition of listening, directing you to profiles of measures contained in Section Three of this volume when appropriate. We then outline the growing recognition that taking listening seriously requires constructing and empirically testing theories of its essential components and consequential roles for individual and relational health and well-being. We conclude with an overview of the book.

Listening: Distinctions and Definitions

Whereas “to listen” is rooted in terms that connote attention and silent obedience, “to hear” has more to do with the perception of sound and the faculties of the ear (see Lipari, 2010; and the response by Bodie & Crick, 2014). This distinction often helps separate the focus of work by audiologists who study the physiological components of hearing from those, like communication scholars, who study the individual and relational components of listening.

In this latter work, listening is recognized as a multidimensional construct that consists of complex (a) affective processes, such as being motivated to attend to others; (b) behavioral processes, such as responding with verbal and nonverbal feedback; and (c) cognitive processes, such as attending to, understanding, receiving, and interpreting content and relational messages (Halone, Cunconan, Coakley, & Wolvin, 1998). As seen

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Table 1.1 Sample of Listening Definitions.

Author	Year	Definition
Tucker	1925	An analysis of the impressions resulting from concentration where an effort of will is required
Rankin	1926	The ability to understand spoken language
Nichols	1948	The comprehension of expository materials presented orally in a classroom situation
Barbe & Meyers	1954	The process of reacting to, interpreting, and relating the spoken language in terms of past experiences and a future course of action
Brown & Carlson	1955	The aural assimilation of spoken symbols in a face-to-face speaker–audience situation, with both oral and visual cues present
Barbara	1957	A definite, usually voluntary, effort to apprehend acoustically
Spearritt	1962	The active process involved in attaching meaning to sounds
Barker	1971	The selective process of attending to, hearing, understanding, and remembering aural symbols
Weaver	1972	A process that takes place when a human organism receives oral data; the selection and retention of aurally received data
Kelly	1975	A rather definite and deliberative ability to hear information, to analyze it, to recall it at a later time, and to draw conclusions from it
Steil *et al.*	1983	Consists of four connected activities – sensing, interpreting, evaluating, and responding
Wolff *et al.*	1983	A unitary-receptive communication process of hearing and selecting, assimilating and organizing, and retaining and covertly responding to aural and nonverbal stimuli
Wolvin & Coakley	1988	The process of receiving, attending to, and assigning meaning to aural stimuli
Brownell	1994	An overt behavior that conceptualizes the teaching and training process
ILA	1996	The process of receiving, constructing meaning from, and responding to spoken and/or nonverbal messages
Cooper	1997	Listening competency means behavior that is appropriate and effective. *Appropriateness* means that the content is understood, and *effectiveness* deals with the achievement of interactive goals.
de Ruyter & Wetzels	2000	(As perceived by customers) A set of interrelated activities, including apparent attentiveness, nonverbal behaviors, verbal behavior, perceived attitudes, memory, and behavioral responses.
Bostrom	2011	The acquisition, process, and retention of information in the interpersonal context

Source: Glenn (1989) and Wolvin and Coakley (1988).

in Table 1.1, scholars often stress one of these three categories, although examples do exist that synthesize all three (e.g., ILA; Steil, Barker, & Watson, 1983).

Affective components of listening include how individuals think about listening and their motivation and enjoyment of the activity. Individuals' views about listening and their (often idiosyncratic) barriers to attending to others can have profound effects on comprehension and understanding as well as consequences for personal, professional, and relational success. Listening *behaviors* are actions such as eye contact and asking questions that serve to signal attention and interest to others. The responses that listeners enact while engaged with another are the only signals that listening is taking (or has taken) place. Finally, *cognitive* elements of listening are those internal processes that operate to enable individuals to attend to, comprehend, interpret, evaluate, and make sense of spoken language. The notion that listening is an information-processing activity consisting of a stable set of practices that can be trained and improved is the most popular way to conceptualize the term and one that has framed all listening research at least since the early 1940s.

Cognitive Components

Whether sleeping or awake, humans are constantly processing sound; that is, vibrations pass through our ears and are processed in our brains continuously (Antony, Gobel, O'Hare, Reber & Paller, 2012). Not all of these sounds, however, are attended to consciously. Most sounds we hear are not "listened to" cognitively, that is, comprehended, understood, and stored in memory for later retrieval and use. Although communication scientists recognize the importance of hearing, most of the work on listening as a cognitive phenomenon has focused on how attended sounds are parsed into words and phrases that are comprehended, understood, interpreted, evaluated, remembered, and recalled (e.g., Burleson, 2011).

The biggest emphasis in the literature has been placed on the factual recall of large chunks of spoken monologue, particularly in the classroom setting. The study that many cite as the catalyst for contemporary listening research was published in 1948 by Ralph Nichols. In that study, Nichols played six 10-minute audio-recorded lectures to a sample of undergraduate students who were asked to answer 10 multiple-choice questions after each. Items on the tests were designed to assess the amount of material from the lectures that students could recall without the assistance of note taking. Student participants recalled an average of 68% of the lecture material, with higher scores related to both individual (e.g., intelligence) and situational (e.g., listener fatigue) factors. Subsequent interviews with instructors of students scoring in the top and bottom tertiles of the test revealed that good, compared to poor, listeners were "more attentive during classroom activities and more conscientious in their … work habits" (Nichols, 1948, p. 160). Nichols spent the remainder of his career attempting to convince others of the power of listening, largely through the publication of his "Listening Is a 10-Part Skill" (Nichols, 1975) and his involvement in both the ILA and the International Communication Association. Nichols's work was instrumental toward motivating serious scholarly attention to factors likely to discriminate among good and poor listeners and to instructional efforts aimed at improving student ability to comprehend aural input. His approach to defining listening as a set of discernable skills (e.g., listening for main ideas, and inference making) remains with us today.

The focus on how students comprehend aural information was shared by early listening scholars who emphasized the importance of comprehending and recalling lecture-based information for student success (e.g., Beatty & Payne, 1984; Beighley, 1952; Goldhaber & Weaver, 1968; McClendon, 1958). From a research standpoint, it is instructive to note that short-term recall of information was the focus of the earliest measures of listening (Gilkinson, 1944) and remained a standard in major listening measures developed from 1950 until the 1970s (Brown & Carlsen, 1955; Dow, 1955). The more you retained, the better, more competent listener you were believed to be. The format of these early tests—multiple-choice with one correct and three or more incorrect answers—remains standard practice.

Issues related to retention and recall remained a strong component of listening research for many years. Nichols's research suggested that listening (as measured by recall) was associated with individual intelligence, vocabulary size, and one's ability to identify the organizational elements of a message. This focus led early scholars to view listening ability as a separate, unitary skill and reduced listening to an activity of information acquisition (Bostrom, 1990). Kelly's (1965, 1967) research suggested otherwise, however. His finding that early listening measures were more highly correlated with tests of intelligence than with each other led listening scholars to reevaluate listening and its facets in terms of a complex, multifaceted process.

Kelly's criticism of early listening tests suggested that cognitive ability contributed to listening ability, and later work has supported this perspective (Thomas & Levine, 1994). A primary cognitive component that entered into listening research around the time of Kelly was the role of memory. The relationship between listening and memory was most extensively theorized by Bostrom and Waldhart (1980), who suggested that the separation of short- and long-term memory could be usefully applied to the development of measures of listening comprehension. Their Kentucky Comprehensive Listening Test (KCLT), which is now out of production, was designed to measure five components of listening comprehension: (a) short-term listening, (b) listening with rehearsal, (c) interpretive listening, (d) lecture listening, and (e) short-term listening with distractions (Bostrom & Waldhart, 1983). By incorporating memory models into a conceptualization of listening, Bostrom and his colleagues were able to tease apart relations among certain types of listening and particular individual predispositions. The relation between listening and memory (and thus recall), however, remains unclear (Bostrom, 1990, 2011; Glenn, 1989; Thomas & Levine, 1994).

The emphasis on retention and comprehension ultimately begs the question of how much retention and recall are necessary. Even in Nichols's work, the average recall score hovered around two thirds. Moreover, recall of lecture material is qualitatively different from recall in an interpersonal context, where emotional overtones may affect retention. Recognitions such as these led to the development of other measures of listening comprehension, including the Watson-Barker Listening Test (Watson & Barker, 1988; Watson, Barker, Roberts, & Roberts, 2001; see Profile 64) and rubrics designed to assess memory for conversation (Stafford, 1982; Stafford, Burggraf, & Sharkey, 1987; Stafford & Daly, 1984; see Profile 38). But even these later measures suffer from insufficient evidence of validity (Bodie, Worthington, & Fitch-Hauser, 2011) and perhaps even a misunderstanding of how people remember conversational details and themes (Janusik, 2005, 2007).

Affective Components

The focus on retention drove listening research for a number of years, with scholars focusing on the relation between comprehension and other cognitive constructs (e.g., Kelly, 1965; Spearritt, 1962). With advances in measurement techniques such as videotaped presentations (as opposed to simply audiotaped ones) and the recognition that elements of the voice and characteristics of speakers and messages can influence comprehension, others started to turn attention to inference making and evaluation rather than simple regurgitation (Fitch-Hauser, 1984, 1990). How and why individuals come to the conclusions they do as they listen have been studied under the auspices of message interpretation (Edwards, 2011), relational framing (Dillard, Solomon, & Palmer, 1999), and other research programs like constructivism (Bodie & Jones, 2016) and schema theory (Edwards & McDonald, 1993)—all assume comprehension of aural information is more complex than simply remembering uttered speech. Research in psychology seems to confirm that memory is not as simple as repeating what is seen or heard and that people have "false memories" even with short lists of words or phrases (Loftus & Palmer, 1974). Extrapolating to interactive contexts, individuals often come away from the same oral event with different information or at least different interpretations and evaluations of that information (Edwards, 2011).

A significant portion of research on affective components of listening has focused on associations between listening and trait-like personality factors that may affect individual motivation. A focus on individual predispositions and their influence on how people interpret and process aural information was implicit in the work of Nichols but was not formally included in cognitive models of listening until the 1972 publication of *Human Listening: Process and Behavior* by Carl Weaver. In his book, Weaver argued that a listener's "attitudes" should be incorporated as part of a "selective perception" model of listening. For the first time, a listener's willingness to or attitude toward listening was identified as a separate component of the listening process (see also Barker, 1971). In other words, individual choice was seen as a key element of listening—we choose to listen (or to avoid it).

Personal experience and academic research suggest that all listeners are not created equal. The central question is, why are some individuals more proficient (or at least more likely to put forth effort) at listening than other individuals? Much research energy has been devoted to the discovery of an overarching profile for good listening (see Bodie, St. Cyr, Pence, Rold, & Honeycutt, 2012). Scholars have investigated how differences in memory (Bostrom & Waldhart, 1988; Janusik, 2005), schema formation (Fitch-Hauser, 1984, 1990), anxiety (Schrodt, Wheeless, & Ptacek, 2000; Wheeless, Preiss, & Gayle, 1997; see the Informational Reception Apprehension Test, Profile 24), and individual preferences for (Bodie, Worthington, & Gearhart, 2013; Worthington, 2003, 2005, 2008; see the Listening Styles Profile-Revised, Profile 36) and conceptualizations of (Imhof & Janusik, 2006; see the Listening Concepts Inventory, Profile 32) listening potentially affect how listeners enact their role. Other examples of research into individual differences include studies between various listening concepts and empathic tendencies (Bodie, 2011a; Bommelje, Houston & Smither, 2003; Chesebro, 1999; Weaver & Kirtley, 1995), noise sensitivity (Worthington, Keaton, Imhof, & Valikoski, 2016; see Profile 46), and related social skills (Gearhart & Bodie, 2011; see the Active-Empathic Listening Scale, Profile 2).

As seen here, this area of study is quite varied. Unfortunately, comparisons of study findings are difficult due to conceptual and methodological differences. For example, studies examining listening style and personality have used the Eysenck Personality Questionnaire, the Kiersey Temperament Sorter, and the Hogan Personality Inventory (Pearce, Johnson, & Barker, 2003; Watson, Barker, & Weaver, 1995; Worthington, 2003). As noted by Bodie, Worthington, Imhof, and Cooper (2008), the research on individual differences presents a difficulty: The unique contribution of any single variable to the listening process is blurred primarily because each is typically studied in isolation. Rarely do studies include measures from a diverse population of constructs; instead, studies are often designed with a limited set of variables in mind. Of course, "People do not possess a 'single' individual difference; they are multifaceted, reflecting numerous personality, temperament, and learned qualities" (Bodie *et al.*, 2008, p. 111). Researchers are urged to use more sophisticated techniques to better understand the unique contributions that "various individual differences may make to both general listening processes and to specific listening skills" (Bodie *et al.*, 2008, p. 111). Systematic and quantitatively oriented reviews of the literature should follow this work.

Behavioral Components

Although placing an emphasis on a listener's motivation and willingness to listen in particular ways, Weaver's book set aside the listening response as a viable research trajectory. It was not until the mid-1980s and the push to develop "speaking and listening competencies" in high school and college students that listening scholars were able to legitimately focus on behaviors. Prior to this time, the response phase was thought to begin a new process, one that was more speaking-focused in nature (see Ridge, 1993, for a discussion of this trend; Berger, 2011, addresses problems associated with this view).

The listening competency model that emerged from the research begun in the 1980s followed closely the communicative competency model made popular by Wiemann, Spitzberg, Rubin, and others (Morreale, Rubin, & Jones, 1998; Spitzberg & Cupach, 1984; Wiemann & Backlund, 1980; Wilson & Sabee, 2003). Models of listening competency placed the overt behaviors of listeners as central to whether a person could be deemed a "good" (or poor) listener. This focus was a natural outgrowth of previous research emphasizing outcomes of retention and recall. Nichols's work was used to justify the need for training in cognitive elements of listening, where observations made by educators were based solely on outward signs of attention and engagement within the classroom setting (i.e., listening behaviors). Even before Nichols, work by Rankin (1926) that asked adults to chart their waking hours devoted to several communication activities (including listening) suggested that what we *do* as listeners (i.e., how we behave as listeners) is important (see Profile 60, "Time Studies"). Even so, a behavioral view of listening was not mainstreamed until the movement toward assessment and measurement was tied to federal funding initiatives (see Beard & Bodie, 2014).

Fundamental to the "listening as competent behavior" perspective is "the view that an identifiable set of skills, attitudes, and abilities can be formulated and taught to improve individual performance" (Bodie *et al.*, 2008, p. 107). What the research from the latter part of the 1980s to throughout the 1990s accomplished was to shift the focus from covert mental processes to overt behavioral ones. Two claims are central in this shift: (a) that our behavioral choices are moderated by our relationships and (b) that competency resides in the eye of the beholder. In other words, our listening competency is judged by others, and this judgment (or at least what is relevant for that judgment) varies with the context. As our listening competency is judged and as we judge the listening of others, we assess the appropriateness and effectiveness of the listening behaviors in specified contexts (Cooper & Husband, 1993; Spitzberg & Cupach, 2002).

Along with a conceptual shift, the behavioral perspective inspired new measurement techniques. Competency expanded beyond multiple-choice assessments of comprehension to include multi-item scales that could be completed by listeners, their interlocutors, and their peers, coworkers, friends, and family members. Along with traditional self-report measures used to assess affective components of listening, researchers began utilizing a variety of other reporting techniques, including third-party and critical-incident techniques (Rubin & Feezel, 1986; Wellmon, 1988). Moreover, there was a growing acknowledgment that listening competency was contextual, with researchers exploring listening competency in the areas of business, education, and health. Researchers in these areas have tied listening competency (measured in multiple ways) to attentiveness, memory, and understanding, as well as employee motivation, upward mobility in the workplace, and job and class performance (Brownell, 1985; Rubin & Feezel, 1986; Sypher, Bostrom, & Seibert, 1989; Wanzer, Booth-Butterfield, & Gruber, 2004; Worthington, 2001).

Even with all the advances afforded by a focus on behavior, there are at least two important limitations (Bodie *et al.*, 2008). First, the emphasis on skills and training directed research attention away from identifying elements of listening competency. Thus, the skills that are taught in classrooms and in training programs may or may not capture the primary elements that should be taught. Second, the majority of studies in this area have been atheoretical in nature (Wolvin, Halone, & Coakley, 1999). No unified framework currently exists to organize and evaluate competency skills, and some even take issue with the need for theoretically oriented research more generally (Purdy, 2011; see response by Bodie, 2011b). Theory, however, is what provides measures with focus and what allows more sophisticated interpretation of results.

The Role of Theory in Listening Research

The early history of the listening field and the drive to develop a single, mutually agreeable definition influenced the nature and type of research that was conducted by scholars. In this section, we highlight and explore the impact of these two forces on both theory development and the type of research that scholars conducted.

Searching for "The" Definition of Listening

Our review may leave the impression that listening scholarship is completely void of theory. This is not the case. What is true is that scholars were slow to expand beyond an initial emphasis on lecture comprehension, and this emphasis drove a felt need to develop a single, all-encompassing definition of listening.

As we detailed in this chapter, early listening research focused on comprehension of orally delivered information in educational contexts, a narrow focus that restricted listening to a kind of information processing sans its broader connection to human communication and relational experiences (Bostrom, 2011). As a result, almost all early listening measures (e.g., the Brown-Carlsen Listening Test [Brown & Carlsen, 1955]; STEP [Educational Testing Services, 1957]; and the Watson-Barker Listening Test, Profile 64) emphasized listening comprehension and recall, a trend that continued for several decades. Moreover, test responses typically took the form of multiple-choice questions with absolute answers. This "right-or-wrong" scoring conflicts with our common experience of partially understanding or comprehending messages (Janusik, 2005, 2007). The work of Kelly (1967) and others (Caffrey, 1955; Lundsteen, 1966; Weaver, 1972) pointed out the limitations of this approach and laid the groundwork for conceptualizing listening as a set of complex skills and abilities.

The set of skills and abilities that should be included in a definition of listening was debated furiously in the listening literature over the course of three decades (1970s–1990s). Some work was grounded in one or more theoretical perspectives, largely borrowed from cognitive psychology. The work of Bostrom and colleagues, for instance, stressed the role of memory, and several others proposed models grounded in human information-processing approaches (Fitch-Hauser, 1990; Goss, 1982). During this time, models of listening proliferated, with most stressing the internal, working apparatus thought to be necessary to process spoken language (see Chapter 4). A common theme in these approaches was the effort to streamline listening scholarship toward the construction of a single, unified definition upon which all could agree; the ultimate goal was to develop universal tests of competence. Emblematic of this view are the words used by Barker, Barker, and Fitch-Hauser (1987): "in order to develop a theory [of listening] we must first agree upon a definition" (p. 15).

Searching for the one, all-encompassing definition of listening, although admired by some, is ultimately rather like putting the proverbial cart before the horse. Instead of theories following definitions, definitions follow theory. Moreover, in line with our view that listening is a multidisciplinary endeavor, and given that there are myriad theoretical frameworks appropriate for the study of listening, there too should be myriad definitions that help shape the field (Bodie, 2010, 2012). When viewed as a theoretical term, listening derives its meaning from the surrounding theoretical structure. And because different theoretical structures propose different terms and processes, *definitions*—instead of a single definition—are the goal. When various meanings of listening are allowed, each of which depends "on the practical purpose pursued by an individual or team of scholars" (Bodie, 2012, p. 114), our goal as listening scholars moves away from the pursuit of definitional consensus and toward exploring the many complexities of the listening process.

Revisiting "The" Definition of Listening

We began this chapter by asserting that listening is a multidisciplinary field. Indeed, the term itself is much broader than past work suggests, and our review here illustrates at least three facets of the term that can guide both empirical work and theory-building efforts. Definitions by their very nature tell us what something is and, by extension, what it is not. Early disagreement over how to correctly define listening reflects the history of the field, including debate surrounding what exactly constitutes listening, the differing philosophies in how it should be defined, and differing views on whether a single definition helps or hurts the field.

Conceptual definitions of a communication, psychological, or related construct serve two important purposes: (a) They describe internal processes and external behaviors that compose the construct, and (b) they delineate its relations with other variables. Unfortunately, many of the definitions provided in Table 1.1 lack the scientific rigor that should undergird a conceptual definition. Conceptual definitions should be grounded in theoretical frameworks and revised over time. The fact is that although the field of listening is over a half century old, much of what contributes to the listening process is not well understood. Such misunderstanding is exacerbated when we also consider differences between scholarly and lay definitions of the construct.

When the average individual refers to listening, they are referencing a state of interpersonal connection with and presence of others (Purdy, 2006). These implicit theories of listening are an important part of the cognitive and affective components of listening and may well determine how people judge others who enact particular behaviors (Bodie, 2010; Bodie *et al.*, 2015). Conversely, when scholars have defined listening, they have tended to focus on the cognitive processes responsible for understanding, comprehending, evaluating, and responding to spoken messages. To ask which of these views is correct is like asking which of the various definitions of any term proposed in a dictionary is correct. Definitions are functional, not right or wrong, but more or less useful for some particular purpose.

Thus, we end this section by *not* providing you with *the* definition of listening. We are not convinced that a single definition of listening is practical or even desirable. Although listening research has seen a resurgence in recent years, our understanding of key aspects of listening processes is woefully lacking. Instead, we suggest that researchers focus greater attention on first determining the key features of specific listening processes and/or behaviors of interest to their particular research project. We believe that the investigator's research goal(s) should be the guiding principle when choosing how listening and related concepts should be conceptualized and subsequently operationalized.

By not providing you with our definition of listening, we are encouraging you to explore the myriad theoretical frameworks appropriate for the study of listening. Bodie (2012) outlined several such frameworks drawn from the work of interpersonal communication scholars. Several measures profiled in this volume stem from one or more of these frameworks. For instance, Affection Exchange Theory views listening as an important way to communicate affection in close relationships (Floyd, 2014; see the Affectionate Communication Index and Affectional Communication Scale, Profiles 6 and 5, respectively). When we are "listened to" and feel understood and valued, our interpersonal needs are being fulfilled. Such a definition of listening is similar in many ways to the one drawn from interpersonal adaptation theory—that is, a behavior that signals involvement and engagement with a person or topic (Jones, 2011), often measured with one or more scales that assess nonverbal immediacy (see Profile 47). Other theories propose definitions that have more to do with how listeners work through understanding messages rather than how they show understanding or communicate intimacy (e.g., constructivism [Burleson, 2011] and relational framing theory [Dillard *et al.*, 1999]). These differences are not problematic, as suggested by "definition first" scholars; they simply illustrate the multidimensional nature of the concept.

Using This Book

This sourcebook was initially conceived as a means of aiding students and scholars in identifying areas of listening study and engaging in the best research practices. We also wanted to provide convenient access to a variety of listening and listening-related measures. Toward these goals, the first half of the book focuses on methodology and measurement issues. Chapter 2, primarily relevant for studies that utilize numerical data to make principled arguments, provides an introduction to measurement issues, including scale development and assessing standardized scales for reliability and validity. Chapter 3 focuses on so-called qualitative methods appropriate for listening research, with an emphasis on ethnographic methods. The remaining chapters outline the various ways in which scholars have operationalized the cognitive processes underlying listening (Chapter 4), have measured affective components (Chapter 5), and have assessed behavioral enactments of listening (Chapter 6).

The second half of the sourcebook offers 65 measurement profiles, tools for assessing the cognitive, affective, and behavioral facets of listening. Profiles utilize a standardized format and were selected from measures previously used in listening research as well as related measures that have implications for listening. Many of these profiles stem from a particular theoretical framework, and when conceptualizing this book, we took seriously the need to expand listening scholarship beyond the standard cognitive model. All measures, if used appropriately, can add to our knowledge regarding the importance and ubiquity of one of the most consequential of life's skills.

References

Antony, J. W., Gobel, E. W., O'Hare, J. K., Reber, P. J., & Paller, K. A. (2012). Cued memory reactivation during sleep influences skill learning. *Nature Neuroscience, 15*, 1114–1116. doi:10.1038/nn.3152

Barbara, D. A. (1971). *How to make people listen to you*. Springfield, IL: Charles C. Thomas.

Barbe, W., & Meyers, R. (1954, February). Developing listening ability in children. *Elementary English, 31*, 82–84.

Barker, D. R., Barker, L. L., & Fitch-Hauser, M. (1987, March). *Listening as a hypothetical construct*. Paper presented at the annual meeting of the International Listening Association, New Orleans, LA.

Barker, L. L. (1971). *Listening behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Beard, D., & Bodie, G. D. (2014). Listening research in the communication discipline. In P. J. Gehrke & W. M. Keith (Eds.), *The unfinished conversation: 100 years of communication studies*. New York: Routledge.

Beatty, M. J., & Payne, S. K. (1984). Listening comprehension as a function of cognitive complexity: A research note. *Communication Monographs, 51*, 85–89. doi:10.1080/03637758409390186

Beighley, K. C. (1952). An experimental study of the effect of four speech variables on listener comprehension. *Speech Monographs, 19*, 249–258. doi:10.1080/03637755209375068

Berger, C. R. (2011). Listening is for acting. *International Journal of Listening, 25*, 104–110. doi:10.1080/10904018.2011.536477

Bodie, G. D. (2010). Treating listening ethically. *International Journal of Listening, 24*, 185–188. doi:10.1080/10904018.2010.513666

Bodie, G. D. (2011a). The Active-Empathic Listening Scale (AELS): Conceptualization and evidence of validity within the interpersonal domain. *Communication Quarterly, 59*, 277–295. doi:10.1080/01463373.2011.583495

Bodie, G. D. (2011b). Theory and the advancement of listening research: A reply to Purdy. *International Journal of Listening, 25*, 139–144. doi:10.1080/10904018.2011.537149

Bodie, G. D. (2012). Listening as positive communication. In T. Socha & M. Pitts (Eds.), *The positive side of interpersonal communication* (pp. 109–125). New York: Peter Lang.

Bodie, G. D., & Crick, N. (2014). Listening, hearing, sensing: Three modes of being and the phenomenology of Charles Sanders Peirce. *Communication Theory, 24*, 105–123. doi:10.1111/comt.12032

Bodie, G. D., & Jones, S. M. (2016). Constructivism. In C. R. Berger & M. E. Roloff (Eds.), *International encyclopedia of interpersonal communication*. Hoboken, NJ: Wiley-Blackwell.

Bodie, G. D., Pence, M. E., Rold, M., Chapman, M. D., Lejune, J., & Anzalone, L. (2015). Listening competency in initial interactions, II: Applying trait centrality to discover the relative placement of listening competence among implicit competency theories. *Communication Studies, 66*, 528–548. doi:10.1080/10510974.2015.1039657

Bodie, G. D., St. Cyr, K., Pence, M., Rold, M., & Honeycutt, J. (2012). Listening competence in initial interactions: Distinguishing between what listening is and what listeners do. *International Journal of Listening, 26*, 1–28. doi:10.1080/10904018.2012.639645

Bodie, G. D., & Worthington, D. L. (2010). Revisiting the Listening Style Profile (LSP-16): A confirmatory factor analytic approach to scale validation and reliability estimation. *International Journal of Listening, 24*, 69–88. doi:10.1080/10904011003744516

Bodie, G. D., Worthington, D. L., & Fitch-Hauser, M. (2011). A comparison of four measurement models for the Watson-Barker Listening Test (WBLT)-Form C. *Communication Research Reports, 28*, 32–42. doi:10.1080/08824096.2011.540547

Bodie, G. D., Worthington, D. L., & Gearhart, C. G. (2013). The Listening Styles Profile-Revised (LSP-R): A scale revision and evidence for validity. *Communication Quarterly, 61*, 75–93. doi:10.1080/01463373.2012.720343

Bodie, G. D., Worthington, D. L., Imhof, M., & Cooper, L. (2008). What would a unified field of listening look like? A proposal linking past perspectives and future endeavors. *International Journal of Listening, 22*, 103–122. doi:10.1080/10904010802174867

Bommelje, R. K., Houston, J., & Smither, R. (2003). Personality characteristics of effective listeners: A five factor perspective. *International Journal of Listening, 17*, 32–46. doi:10.1080/10904018.2003.10499054

Bostrom, R. N. (1990). Conceptual approaches to listening behavior. In R. N. Bostrom (Ed.), *Listening behavior: Measurement and application* (pp. 1–14). New York: Guilford.

Bostrom, R. N. (2011). Rethinking conceptual approaches to the study of "listening." *International Journal of Listening, 25*, 10–26. doi:10.1080/10904018.2011.536467

Bostrom, R. N., & Waldhart, E. S. (1980). Components in listening behavior: The role of short-term memory. *Human Communication Research, 6*, 221–227. doi:10.1111/j.1468-2958.1980.tb00142.x

Bostrom, R. N., & Waldhart, E. S. (1983). *The Kentucky Comprehension Listening Test.* Lexington, KY: The Kentucky Listening Research Center.

Bostrom, R. N., & Waldhart, E. S. (1988). Memory models and the measurement of listening. *Communication Education, 37*, 1–13. doi:10.1080/03634528809378699

Brown, J. I., & Carlsen, G. R. (1955). *Brown-Carlsen Listening Comprehension Test.* New York: Harcourt, Brace & World.

Brownell, J. (1985). A model for listening instruction: Management applications. *Business Communication Quarterly, 48*, 39–44. doi:10.1177/108056998504800312

Brownell, J. (1994, December). Teaching listening: Some thoughts on behavioral approaches. *The Bulletin,* 19–24.

Burleson, B. R. (2011). A constructivist approach to listening. *International Journal of Listening, 25*, 27–46. doi:10.1080/10904018.2011.536470

Caffrey, J. G. (1955). Auding. *Review of Educational Research, 25*, 121–138.

Chesebro, J. L. (1999). The relationship between listening styles and conversational sensitivity. *Communication Research Reports, 16*, 233–238. doi:10.1080/08824099909388722

Cooper, L. O. (1997). Listening competency in the workplace: A model for training. *Business Communication Quarterly, 60*, 75–84. doi:10.1177/108056999706000405

Cooper, L., & Husband, R. L. (1993). Developing a model of organizational listening competency. *International Journal of Listening 7*, 6–34. doi:10.1080/10904018.1993.10499112

de Ruyter, K., & Wetzels, M. (2000). Customer equity considerations in service recovery: A cross-industry perspective. *International Journal of Service Industry Management, 11*, 91–108. doi:10.1108/09564230010310303

Dillard, J. P., Solomon, D. H., & Palmer, M. T. (1999). Structuring the concept of relational communication. *Communication Monographs, 66*, 49–65. doi:10.1080/03637759909376462

Dow, C. W. (1955). Listening instruction at Michigan State: 1954–55. *Journal of Communication, 5*, 110–112. doi:10.1111/j.1460-2466.1955.tb01107.x

Educational Testing Services. (1957). *Sequential tests of educational progress*. Princeton, NJ: Author.

Edwards, R. (2011). Listening and message interpretation. *International Journal of Listening, 25*, 47–65. doi:10.1080/10904018.2011.536471

Edwards, R., & McDonald, J. L. (1993). Schema theory and listening. In A. D. Wolvin & C. G. Coakley (Eds.), *Perspectives on listening* (pp. 60–77). Norwood, NJ: Ablex.

Fitch-Hauser, M. (1984). Message structure, inference making, and recall. *Communication Yearbook, 8*, 378–392.

Fitch-Hauser, M. (1990). Making sense of data: Constructs, schemas, and concepts. In R. N. Bostrom (Ed.), *Listening behavior: Measurement and application* (pp. 76–90). New York: Guilford.

Floyd, K. (2014). Empathic listening as an expression of interpersonal affection. *International Journal of Listening, 28*, 1–12. doi:10.1080/10904018.2014.861293

Gearhart, C. C., & Bodie, G. D. (2011). Active-empathic listening as a general social skill: Evidence from bivariate and canonical correlations. *Communication Reports, 24*, 86–98. doi:10.1080/08934215.2011.610731

Gilkinson, H. (1944). Experimental and statistical research in general speech: I. Effects of training and correlates of speech skill. *Quarterly Journal of Speech, 30*, 95–101. doi:10.1080/00335634409380958

Glenn, E. C. (1989). A content analysis of fifty definitions of listening. *Journal of the International Listening Association, 3*, 21–31. doi:10.1207/s1932586xijl0301_3

Goldhaber, G. M., & Weaver, C. H. (1968). Listener comprehension of compressed speech when the difficulty, rate of presentation, and sex of the listener are varied. *Speech Monographs, 35*, 20–25. doi:10.1080/03637756809375562

Goss, B. (1982, April). Listening to language: An information processing perspective. Paper presented at the Southern Communication Association convention, Hot Springs, AR. Retrieved from http://files.eric.ed.gov/fulltext/ED215379.pdf

Halone, K., Cunconan, T. M., Coakley, C. G., & Wolvin, A. D. (1998). Toward the establishment of general dimensions underlying the listening process. *International Journal of Listening, 12*, 12–28. doi:10.1080/10904018.1998.10499016

Imhof, M., & Janusik, L. (2006). Development and validation of the Imhof-Janusik Listening Concepts Inventory to measure listening conceptualization differences between cultures. *Journal of Intercultural Communication Research, 35*, 27–98. doi:10.1080/17475750600909246

International Listening Association (ILA). (1996). Home page. Retrieved from http://listen.org/.

Janusik, L. (2005). Conversational Listening Span: A proposed measure of conversational listening. *International Journal of Listening, 19*, 12–28.

Janusik, L. (2007). Building listening theory: The validation of the conversational listening span. *Communication Studies, 58*, 139–156. doi:10.1080/10510970701341089

Jones, S. M. (2011). Supportive listening. *International Journal of Listening, 25*, 85–103. doi:10.1080/10904018.2011.536475

Kelly, C. M. (1965). An investigation of the construct validity of two commercially published listening tests. *Speech Monographs, 32*, 139–143. doi:10.1080/03637756509375443

Kelly, C. M. (1967). Listening: Complex of activities—and a unitary skill? *Speech Monographs, 34*, 455–465. doi:10.1080/03637756709375556

Kelly, C. M. (1975). Empathic listening. In R. L. Applbaum, O. O. Jenson, & R. Carroll (Eds.), *Speech communication: A basic anthology* (pp. 115–123). New York: Macmillan.

Lipari, L. (2010). Listening, thinking, being. *Communication Theory, 20*, 348–362. doi:10.1111/j.1468–2885.2010.01366.x

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 585–589. doi:10.1016/S0022-5371(74)80011-3

Lundsteen, S. W. (1966). Teaching and testing critical listening: An experiment. *Elementary School Journal, 66*, 311–315. Retrieved from http://www.jstor.org/stable/41385715

McClendon, P. I. (1958). An experimental study of the relationship between the note-taking practices and listening comprehension of college freshmen during expository lectures. *Speech Monographs, 25*, 222–228. doi:10.1080/03637755809375236

Morreale, S., Rubin, R. B., & Jones, E. (Eds.). (1998). *Speaking and listening competencies for college students*. Washington, DC: National Communication Association.

Nichols, R. (1948). Factors in listening comprehension. *Speech Monographs, 15*, 154–163. doi:10.1080/03637754809374953

Nichols, R. G. (1975, July). Listening is a 10 part skill. *Nation's Business*, pp. 56–58, 60. Retrieved from http://digital.hagley.org/cdm/landingpage/collection/p16038coll3

Pearce, C. G., Johnson, I. W., & Barker, R. T. (2003). Assessment of the Listening Styles Inventory: Progress in establishing reliability and validity. *Journal of Business and Technical Communication, 17*, 84–113. doi:10.1177/1050651902238546

Purdy, M. (2006, April). Listening and gender: Stereotypes and explanations. Paper presented at the International Listening Association annual convention, Salem, OR.

Purdy, M. W. (2011). Grounding listening: The limitations of theory. *International Journal of Listening, 25*, 132–138. doi:10.1080/10904018.2011.537144

Rankin, P. T. (1926). *The measurement of the ability to understand spoken language. Dissertation Abstracts, 12*(1952), 847.

Ridge, A. (1993). A perspective on listening skills. In A. D. Wolvin & C. G. Coakley (Eds.), *Perspectives on listening* (pp. 1–14). Norwood, NJ: Ablex.

Rubin, R. B., & Feezel, J. D. (1986). A needs assessment of research in communication education. *Central States Speech Journal, 37*, 113–118. doi:10.1080/10510978609368210

Schrodt, P., Wheeless, L. R., & Ptacek, K. M. (2000). Informational reception apprehension, educational motivation, and achievement. *Communication Quarterly, 48*, 60–73. doi:10.1080/01463370009385580

Spearritt, D. (1962). *Listening comprehension: A factorial analysis* (ACER Research series no. 76). Melbourne: Australian Counsel for Educational Research (ACER).

Spitzberg, B., & Cupach, W. R. (1984). *Interpersonal communication competence*. Beverly Hills, CA: Sage.

Spitzberg, B., & Cupach, W. R. (2002). Interpersonal skills. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 564–612). Thousand Oaks, CA: Sage.

Stafford, L. (1982). *Conversational memory: Effects of instruction set and recall mode*. Unpublished master's thesis, University of Texas, Austin, TX.

Stafford, L., Burggraf, C. S., & Sharkey, W. F. (1987). Conversational memory: The effects of time, recall, mode and memory expectancies on remembrances of natural conversations. *Human Communication Research, 14*, 203–229. doi:10.1111/j.1468-2958.1987.tb00127.x

Stafford, L., & Daly, J. A. (1984). Conversational memory: The effects of recall mode and memory expectancies on remembrances of natural conversations. *Human Communication Research, 10*, 379–402. doi:10.1111/j.1468-2958.1984.tb00024.x

Steil, L. K., Barker, L. L., & Watson, K. W. (1983). *Effective listening: Key to your success*. New York: Random House.

Sypher, B. D., Bostrom, R. N., & Seibert, J. H. (1989). Listening, communication abilities, and success at work. *Journal of Business Communication, 26*, 293–303. doi:10.1177/002194368902600401

Thomas, L. T., & Levine, T. R. (1994). Disentangling listening and verbal recall. *Human Communication Research 21*, 103–129. doi:10.1111/j.1468-2958.1994.tb00342.x

Tucker, W. (1925, April). Science of listening. *19th Century*, *97*, 548–557.

Wanzer, M. B., Booth-Butterfield, M., & Gruber, K. (2004). Perceptions of health care providers' communication: Relationships between patient-centered communication and satisfaction. *Health Care Communication, 16*, 363–384. doi:10.1207/S15327027HC1603_6

Watson, K. W., & Barker, L. L. (1988). Listening assessment: The Watson-Barker listening test. *Journal of the International Listening Association, 2*, 20–32. doi:10.1080/10904018.1988.10499095

Watson, K. W., Barker, L. L., Roberts, C. V., & Roberts, J. D. (2001). *Watson-Barker listening test: Video version: Facilitator's guide*. Sautee, GA: SPECTRA.

Watson, K. W., Barker, L. L., & Weaver, J. B., III. (1995). The listening styles profile (LSP-16): Development and validation of an instrument to assess four listening styles. *International Journal of Listening, 9*, 1–13. doi:10.1080/10904018.1995.10499138

Weaver, C. (1972). *Human listening: Process and behavior*. Indianapolis, IN: Bobbs-Merrill.

Weaver, J. B., III, & Kirtley, M. D. (1995). Listening styles and empathy. *Southern Journal of Speech Communication, 60*, 131–140. doi:10.1080/10417949509372970

Wellmon, T. A. (1988). Conceptualizing organizational communication competence: A rules-based perspective. *Management Communication Quarterly, 1*, 515–534. doi:10.1177/0893318988001004004

Wheeless, L. R., Preiss, R. W., & Gayle, B. M. (1997). Receiver apprehension, informational receptivity, and cognitive processing. In J. A. Daly, J. C. McCroskey, J. Ayres, T. Hopf, & D. M. Ayres (Eds.), *Avoiding communication: Shyness, reticence, and communication apprehension* (2^{nd} ed., pp. 151–187). Cresskill, NJ: Hampton.

Wiemann, J. M., & Backlund, P. (1980). Current theory and research in communicative competence. *Review of Educational Research, 50*, 185–199. doi:10.3102/00346543050001185

Wilson, S. R., & Sabee, C. M. (2003). Explicating communicative competence as a theoretical term. In J. O. Greene & B. R. Burleson (Eds.), *Handbook of communication and social interaction skills* (pp. 3–50). Mahwah, NJ: Erlbaum.

Wolff, F., Marsnik, N., Tacey, W., & Nichols, R. (1983) *Perceptive listening*. New York: Holt, Rinehart & Winston.

Wolvin, A. D., & Coakley, C. (1988). *Listening*. Dubuque, IA: Wm. C. Brown.

Wolvin, A. D., Halone, K. K., & Coakley, C. G. (1999). An assessment of the "intellectual discussion" on listening theory and research. *International Journal of Listening, 13*, 111–129. doi:10.1080/10904018.1999.10499030

Worthington, D. L. (2001). Exploring juror's listening processes: The effect of listening style preference on juror decision making. *International Journal of Listening, 15*, 20–35. doi:10.1080/10904018.2001.10499043

Worthington, D. L. (2003). Exploring the relationship between personality type and listening style preference. *International Journal of Listening, 17*, 68–87. doi:10.1080/10904018.2003.10499056

Worthington, D. L. (2005). Exploring the relationship between listening style preference and personality traits: Verbal aggressiveness. *International Journal of Listening. 19*, 3–11. doi:10.1080/10904018.2005.10499069

Worthington, D. L. (2008). Exploring the relationship between listening style preference and need for cognition. *International Journal of Listening, 22*, 46–58. 10.1080/10904010701802154

Worthington, D. L., Keaton, S., Imhof, M, & Valikoski, T.-R. (2016). Impact of noise sensitivity on mobile phone attitudes and behaviors. *Mobile Media and Communication, 4*, 3–18. doi:10.1177/2050157915581435

Section Two

Methodological Trends, Challenges, and Opportunities in Listening Research

This section contains five chapters that outline various challenges, trends, and opportunities that should be considered prior to launching a listening research study. Chapters 2 and 3 provide an introduction to basic issues of measurement appropriate for studies using numerical data and narrative-based data (with a primary focus on ethnography), respectively. Given space limitations, neither chapter can do full justice to the research enterprise, and readers are advised to seek more thorough texts, referenced throughout the chapters.

The chapters that follow address the three primary conceptualizations of listening introduced in Chapter 1: cognitive, affective, and behavioral. Each chapter contains information regarding current best practices, suggests directions for future research, and introduces emerging methods of data collection and analysis. Together, they provide the framework for beginning researchers to understand the ins and outs of conducting listening research; for more advanced scholars, we hope this information provides new perspectives for considering the role of listening in our personal and professional lives. At the least, these chapters are appropriate for advanced undergraduate and beginning graduate-level courses that incorporate reading or conducting listening research.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

2

Measuring Listening

Graham D. Bodie and Debra L. Worthington

As we outlined in Chapter 1, how best to measure listening has been a focus of listening scholarship from the beginning. Following the first truly systematic study of listening comprehension (Nichols, 1948), Brown and Carlsen (1955) designed a measure to test factors underlying "good" listening: vocabulary, recognition of transitions, ability to follow directions, immediate recall, and the retention of facts from a lecture. Similar components were used to formulate tests of listening comprehension still used today (e.g., the Watson-Barker Listening Test [WBLT]; see Profile 64), and these same skills are listed as essential components of listening competence for K–12 and higher education students alike (Cooper, 1998; Wolvin & Coakley, 1994; also see Chapter 6).

As educators and practitioners, we are interested in teaching and training specific skills with the goal of helping people become better listeners. To achieve this goal, it is essential that we use tests that are valid indicators of what we are teaching. Well-developed measures are essential to both quality training efforts and theory-building efforts. Without measures that reliably approximate important theoretical constructs, efforts to provide abstract answers to real-world problems are doomed from the start. This chapter provides a review of basic scale development processes as outlined by DeVellis (2012) and framed by listening research. We focus on three broad areas:

- defining and operationalizing the phenomenon,
- developing the initial measure, and
- developing a validity portfolio (including evidence of score reliability).

We review each area, providing considerations related to developing and assessing measures, including appropriate statistical analysis, item analysis, norms, and scaling options.

Define and Operationalize the Phenomenon

The chief purpose of measurement in the social sciences is to estimate constructs. *Constructs* are the "postulated attribute[s] of people, assumed to be reflected in test performance" (Cronbach & Meehl, 1959, p. 283). As hypothetical variables, constructs cannot be directly observed but are assumed to exist as processes or entities. In other

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

words, a construct is "something that scientists put together from their own imaginations, something that does not exist as an isolated, observable dimension of behavior" (Nunnally, 1978, p. 96). Very few listening phenomena are ever directly measured, but we make claims about listening constructs based on their operationalizations.

Listening scholars are interested in a wide range of constructs, from how proficiently a student can comprehend a lecture to the degree of affiliation or immediacy within a conversation between a parent and child. Before researchers collect a single empirical observation of listening, therefore, they must first define how they are using the term. In Chapter 1, we presented a number of common definitions of listening, arguing that definitions should be situated within appropriate theoretical frameworks because theory provides both the focus and the boundary conditions necessary for developing measurement content. On the rare occasions when existing theories are neither useful nor appropriate for measurement development, tangential theories may serve to help develop a theoretically driven conceptual foundation. In these cases, DeVellis (2012) argued that researchers may rely on a clear definition of the phenomenon under study or outline the relation of the proposed construct to other established constructs.

Operationalization

Regardless of the conceptual starting point, researchers must eventually address how to measure the variables they wish to study. As Kerlinger and Lee (2000) wrote:

> operational definitions ... are indispensable ingredients of scientific research because they enable researchers to measure variables and because they are bridges between the theory-hypothesis-construct level and the level of observation. There can be no scientific research without observations, and observations are impossible without clear and specific instructions on what and how to observe. Operational definitions are such instructions. (p. 43)

Operationalization is the process of defining the measurement of a construct. Because conceptual definitions are abstract, any given operationalization will necessarily be selective, with selectivity based largely on logistical concerns. For instance, researchers can ask only a finite number of questions or ask participants to respond to a limited number of items. When observing listeners in action, the behaviors that are coded have to be selected and defined in specific ways. Researchers who study nonverbal communication might, for instance, make distinctions between types of eye contact (e.g., looking up and gazing), whereas others may code at a more molar level (see Microanalysis of Face-to-Face Dialogue and nonverbal immediacy in Profiles 41 and 47, respectively).

The number and types of items on a measurement instrument are functions of both what is being measured and administration time. As a rule, the measurement instrument should be generated with the population of interest in mind, and researchers should be aware that changing instructions, item or question wording, or other elements of a questionnaire will result in nonequivalent measures that should be compared both conceptually and empirically (Raju, Laffitte, & Byrne, 2002).

If you are new to research, you may be more familiar with another term that is closely associated with operationalization—*variable*. Technically, a *variable* is something that varies. It is "a property that takes on different values" (Kerlinger & Lee, 2000, p. 40). Multiple facets of listening vary, as do the antecedents and consequences of listening; thus, the number of variables of interest to listening scholars is vast. In general, we can identify two broad classes of operational definitions, measured and experimental, and two types of variables, dependent and independent.

According to Kerlinger and Lee (2000), "a *measured* operational definition describes how a variable will be measured" (p. 42). Listening comprehension is generally defined by some standard set of questions asked after the presentation of orally delivered information. When people answer more questions correctly, they are said to have a higher listening comprehension score. When conducting experimental work, researchers often use *experimental* operational definitions that "[spell] out the details (operations) of the investigator's manipulation of a variable" (Kerlinger & Lee, 2000, p. 43). For example, if you are interested in discovering how listening comprehension might be improved, you might operationally define message complexity as the *type–token ratio* (TTR), the number of different words in a message (types) divided by the total number of words (tokens). Messages with a high degree of complexity will have a high TTR, whereas messages with a low degree of complexity will have a low TTR. By manipulating the TTR in messages, researchers could ascertain whether message complexity (operationalized as TTR) influences how much information people retain from messages (see Listenability Style Guide, Profile 31).

When conducting an experimental study (like the one regarding message complexity), scholars often talk about dependent and independent variables. The *dependent variable* (DV) is the outcome of interest. A researcher interested in listening comprehension could operationalize this concept in several ways, such as the amount of material recalled after lecture exposure or the ability to construct an appropriate narrative regardless of whether explicit details are recalled. DVs involve measured operational definitions. An *independent variable* (IV) is any aspect of the environment or individual thought to have an influence on the DV. IVs can either be manipulated by the researcher (part of the environment and thus experimental operational definitions) or be a characteristic of the participants in a sample (and thus measured operational definitions). For instance, in your listening comprehension study, IVs might include manipulating the length of the lecture, the subject material presented in the lecture, and/or the credibility of the lecturer. Or you might be interested in how much information is recalled by male versus female students; freshmen, sophomores, juniors, or seniors of various academic majors; or students with different ACT scores or grade point averages (GPAs). These latter variables—those intrinsically associated with participants and out of direct control of the researcher—are called *classification*, *subject-characteristic*, or *attribute IVs*. The distinction between manipulated and classification variables is important because conclusions about causation are most strongly made when the researcher has direct control over manipulated variables; conclusions regarding causation with classification IVs are more speculative.

Choosing an Operationalization

The fact that even a single construct can have myriad operationalizations (that can be measured and/or manipulated) begs the question, how do you choose among them? The first consideration is to choose an operationalization that potential critics of your research, like editors and reviewers, will find convincing (Hayes, 2005).

Read the literature, and determine what measures other scholars are using for similar purposes. Typically, widely used scales that others have relied on to measure the construct are preferred. Of course, broad acceptance does not guarantee an established validity portfolio (see further in this chapter and Profile 64). The descriptions and critiques of the measures profiled in this volume will help you choose among operationalizations wisely.

Second, time and length constraints allowing, multiple operationalizations of a construct should be utilized. Although reporting results for one operationalization is typically enough to warrant publication, it is much more powerful to present results that converge for multiple measures or that show theoretically interesting differences. For instance, Bodie, Jones, Vickery, Hatcher, and Cannava (2014) found that individual self-reports of the tendency to engage in active listening behaviors, using the Active-Empathic Listening Scale (see Profile 2), were: 1) unassociated with the perception of whether those behaviors occurred in a conversation by a conversational partner, and 2) unassociated with the actual occurrence of those behaviors as coded by trained raters. Had the researchers only used one operationalization of active listening, the results would have told only part of the story.

Of course, sometimes there is no readily available operationalization of the listening construct of interest. For instance, when Imhof and Janusik (2006) sought to study how individual conceptualizations of listening might vary as a function of culture, they were unable to turn to the literature because no such operationalization of listening concepts was available. Instead, they had to develop an initial measure of this construct (see the Listening Concepts Inventory, Profile 32).

Develop the Initial Measures

Here, we focus on a number of important considerations when developing measures, including: generating items, determining scale lengths, formatting the measures, and the importance of time considerations and expert reviews. We also examine the level of specificity of a measure and distinctiveness from other constructs.

Generate an Item Pool

Measurement instruments can contain a variety of types of items, including open-ended questions that might require a few words to answer or extended essays that may require several minutes of talking or paragraphs of text. Most tests of listening comprehension rely exclusively on multiple-choice questions that force respondents to select among a preformulated answer set; however, there are examples of tests that include open-ended questions (see the Communication Competency Assessment Instrument, Profile 10). For the most part, the measures that are profiled in this book include opinion and observation-based questions that comprise self-report scales and behavioral coding rubrics. So, the first decision when creating a new instrument is to choose the types of items that can best operationalize the construct of interest.

After the purpose of a measure is identified, a large pool of possible items should be generated. At this time, redundancy is to be applauded because testing multiple items aids in identifying those that are best suited to measure the construct under study and ideally leads to discarding weak or irrelevant items. These choices should be theory driven. Theory will direct your item generation by suggesting the characteristics of the construct of interest as well as its possible facets. A good example of this principle is found in the work on implicit theories of listening competence, or how people view listening. One strategy for investigating how people define listening is to ask a single item, "What is listening?" to which participants can respond with as few as one word or as many as several hundred words. Once data are gathered, the researcher is then responsible for coding the responses into categories that can be either preselected from existing research (see Haas & Arnold, 1995) or derived systematically from the data themselves (see Bodie *et al.*, 2015; Bodie, St. Cyr, Pence, Rold, & Honeycutt, 2012). A second strategy was taken by Imhof and Janusik (2006), who first generated a set of 65 words that either were synonyms of listening (e.g., attention) or had been used to describe different facets of listening in the research literature (e.g., understanding and comprehending). Respondents were asked to indicate the degree to which they thought each word was similar to listening. The scale that emerged from their subsequent statistical analyses is thought to capture how people view listening (i.e., their implicit theory of listening) (see the Listening Concepts Inventory, Profile 32).

Optimize Scale Length

The Imhof and Janusik (2007) study illustrates an additional principle: In the process of subsequent research, particularly weak or nonrepresentative items will be shed in favor of items that capture more of the variance in the construct of interest. Although they started with 65 items, only 33 items were retained in their final scale. These 33 items were further reduced by Bodie (2010a), who found conceptualizations of listening can be measured adequately by 15 items. In some cases, new items might need to be written to capture additional facets of the construct, and the conceptual definition might even need modification to align with how respondents are interpreting relations among operationalizations of constructs (see, e.g., Bodie, Worthington, & Gearhart, 2013; Gearhart, Denham, & Bodie, 2014; and the Narrative Believability Scale, Profile 44). Item generation is an iterative process that may involve several attempts before a final version of the scale is settled—and even then, perhaps only for a short time before revisions are needed. We encourage researchers to test and retest scales continually, even those (or, maybe, especially those) that are "well established" (Levine, Hullett, Turner, & Lapinski, 2006; see also Kline, 1986; Wengraf, 2001).

Determine Format for Measurement

In this section, we focus on scales of measurement and response categories. Formatting and item generation go hand in hand. The majority of measures profiled in this volume are scales that typically consist of items scored on a continuum, then summed for a total score.

In some cases, researchers decide to assign numbers to observations (e.g., behaviors, answer choices on self-report items, or narratives generated during an interview). The types of numbers that can be applied are often a function of the construct of interest. When measuring listening comprehension, if the test had one item, it could be scored as "correct" or "incorrect" and the numbers 1 and 0 applied to these choices. So, if the correct answer to that single item were *A*, only those participants choosing *A* would receive a correct score (1), but all other choices (*B*, *C*, *D*, etc.) would be scored as 0. Like typical multiple-choice tests, scores are derived by adding the number of correct responses and dividing by the total number of questions (e.g., 38/56 = 67.9%).

This example can help illustrate an important principle highlighted by Stevens (1946). Stevens identified four scale types that vary with respect to "the statistical manipulations that can legitimately be applied" to collected data (1946, p. 677). Those scale types are nominal, ordinal, interval, and ratio.

For *nominal variables,* independent categories are created and numbers assigned arbitrarily to these categories. When entered into a statistical package (e.g., SPSS, SAS, and STATA), each individual or dyad is typically assigned a unique identifier (e.g., 001). Like the number on an athletic jersey, these numbers simply serve to identify a particular player or participant and have no other meaning. An "incorrect" response for a listening comprehension test item would be recorded as 0, whereas a "correct" response would receive a 1. Unfortunately, this approach results in the loss of information (i.e., *A* is the only "correct" response; *B–D* are "incorrect"). This type of binary response allows for little variability and restricts covariation significantly (DeVellis, 2012), as all the participants who selected *B–D* are treated the same. This limitation in variation (and covariation) essentially means that an increased number of items are needed to reach the same level of scale variance that can be achieved with scales using more response categories (e.g., Likert and Semantic Differential scales). The advantage of binary responses, however, is that they are easily and quickly completed by respondents; they also may permit researchers to realize satisfactory scale variance by aggregating information across a larger number of items (DeVellis, 2012).

Ranked data are said to have an *ordinal scale*, just like athletes might be ranked in terms of ability (e.g., a first- vs. second-string quarterback). For example, we could rank participants based on their listening comprehension scores with those ranks corresponding to increasing or decreasing scores on our measure of comprehension. Most measures of listening comprehension contain 10 or more multiple-choice questions, and the number of correct responses across the entire set of items is the measure of comprehension. Thus, ranking participants by listening comprehension scores is analogous to ranking students by exam score, GPA, or another standardized metric of performance.

Of course, if listening comprehension is measured as the amount of information recalled after lecture exposure, the researcher is provided a continuous, not just an ordered, measure of the concept; not only is a higher score greater than a lower score (information that ranking would provide), but also we know how much greater. A score of 20 is five pieces of information more than 15, which is five pieces more than a score of 10. In this case, our scale is at the *interval* level. Many of the measures profiled in this volume use semantic-differential scaling (bipolar adjectives such as *loud/soft*) (e.g., see the Multidimensional Evaluation of Enacted Social Support, Profile 42) or Likert scaling (e.g., 5 = Strongly Agree, 1 = Strongly Disagree) (e.g., see the Communication Competency Assessment Instrument and Interpersonal Reactivity Index, Profiles 10 and 28, respectively); both scaling options approximate interval-level scaling.

A *ratio scale* is a continuous scale that also includes a true zero point. On our listening comprehension measure, because an individual can recall zero pieces of information, our scale is a ratio scale. In a similar manner, data from interview narratives or conversational transcripts can be transformed into one or more ratio-level variables (see Language Style Matching, Profile 29).

Develop a Validity Portfolio (Including Evidence of Score Reliability)

The goal of measurement development is to achieve a measure that accurately captures a construct of interest. Awareness of the issues that may influence the validity of an instrument is needed, such as social desirability, survey fatigue, and inattention. It is possible to measure social desirability (e.g., Strahan & Gerbasi, 1972), design shorter measures, and control study designs. Of course, the primary means of assessing validity is to test the phenomenon under study against other theoretically related constructs. Doing so contributes to the validity portfolio of the new measure or provides evidence that it requires reevaluation.

It is important to note that neither reliability nor validity is a property of measurement instruments. Reliability is a product of scores, and it is more appropriate to talk about the validity portfolio of an instrument or of evidence in support of validity than to claim a measure "is valid." In general, reliability is a necessary but not sufficient condition for validity: when scores are highly variable, there is no single underlying construct (no reliability or validity); even when scores are reliable indicators of a construct, they might not necessarily measure the construct of interest (reliability without validity). When reliability and validity are both high, then researchers can more confidently make claims about populations or processes.

Reliability

Reliability is the degree to which data are consistent: To what degree are an individual test taker's scores consistent over time? To what degree are multiple items consistently measuring the same underlying construct? How consistent is a group of coders or raters assessing the same sample of behavior? More generally, reliability is the degree to which scores are free from *random error*, or error associated with factors that vary from measurement to measurement. Random error can occur because of ambiguity in items or directions, differences in the administration of an instrument (e.g., time of day or temperature of room), and characteristics of the setting, individual test takers, or behavior raters. When random error is high, scores attributed to individuals and their behavior are not likely to be consistent. Researchers are concerned with two types of consistency: homogeneity (internal consistency) and stability (repeatability).

Internal consistency is the degree to which a set of items written to measure a construct correlate with each other. The most commonly reported measure of internal consistency is *Cronbach's alpha,* which is an adjusted measure of the average correlation between each item and all other items. In particular, alpha is calculated as:

$$\alpha = \frac{K * \bar{r}}{1 + ((K-1) * \bar{r})}$$

where K is the number of items, and $\bar{r}$ is the average correlation among all pairs of items. So, the reliability of a set of items increases both as a function of the average correlation and as a function of the number of items. Suppose, for instance, that the average correlation between a set of K items is 0.50. When $K = 2$, alpha will equal 0.667, and this value increases with increasing items as illustrated in Table 2.1.

Table 2.1 Cronbach's alpha values as a function of number of items when $\bar{r} = 0.50$.

No. of items	Alpha value
3	0.750
4	0.800
5	0.833
6	0.857
7	0.875
8	0.889
9	0.900
10	0.909

Stability refers to the "repeatability" of a measure over time. Researchers can calculate a stability coefficient for a self-report measure by administering the same scale to a group of people at two points in time (temporal stability) or by administering what are thought to be equivalent forms of a test at two time points (alternate-forms reliability). The latter of these strategies has been used extensively in research on listening comprehension (see WBLT, Profile 64). Chapter 5 reviews additional information about the internal consistency and stability of self-report measures.

Similar in many ways to the consistency of a set of items, when researchers are measuring listening behaviors, consistency is the degree to which two or more raters agree on a set of observations. Several measures of interrater or intercoder reliability (agreement) exist, including Pearson's *r*, the linear dependence between two variables, the intraclass correlation coefficient, or the proportion of variance attributable to between-group differences. When a researcher has to divide up streams of behavior (e.g., see Profile 47, Nonverbal Immediacy Measures), some measure of unitizing reliability (e.g., Guetzkow's *U*) as well as a measure of categorization reliability (e.g., Cohen's kappa) are necessary. See Chapter 6 for additional information on interrater reliability.

Validity

Validity is an ongoing process involving the accumulation of evidence associated with as many types of validity as necessary, to provide confidence that the measure is performing the way it should. The American Psychological Association's (APA) Committee on Psychological Tests in 1950 divided validity into four types: Below, we examine predictive, concurrent, content, and construct validity, as well as criterion-oriented and responsiveness validity.

Criterion-oriented Validity

According to Cronbach and Meehl (1959):

> The pattern of a criterion-oriented study is familiar. The investigator is primarily interested in some criterion which he wishes to predict. He administers the test, obtains an independent criterion measure on the same subjects, and computes a

> correlation. If the criterion is obtained some time after the test is given, he is studying *predictive validity*. If the test score and criterion score are determined at essentially the same time, he is studying *concurrent validity*. Concurrent validity is studied when one test is proposed as a substitute for another (for example, when a multiple-choice form of spelling test is substituted for taking dictation), or a test is shown to correlate with some contemporary criterion (e.g., psychiatric diagnosis). (p. 175)

We can see examples of each of these subtypes of criterion-oriented validity in the listening research. For example, efforts to build a validity portfolio for the Active-Empathic Listening Scale (see Profile 2) show correlations between it and concurrently administered scales tapping various facets of activity and empathy in sales (Drollinger, Comer, & Warrington, 2006) and general interpersonal contexts (Bodie, 2011a). In terms of predictive validity, research using the Multidimensional Evaluation of Enacted Social Support (Goldsmith, McDermott, & Alexander, 2000) has reported that variations in supportive message content (what people say when trying to be supportive) are predictive of important outcomes of supportive talk (e.g., willingness to discuss bullying incidents with a family member; see Profile 42).

Content Validity

Evidence for content validity can be obtained by "showing that the test items are a sample of a universe in which the investigator is interested" (Cronbach & Meehl, 1959, p. 175). Content validity is similar to face validity insofar as each is concerned with the similarity of the test items to the construct of interest, but content validity goes further by, for instance, surveying an independent sample of content experts for their opinions regarding the representativeness of a set of items created by a research team (see Narrative Believability Scale, Profile 44).

After a set of items is created to operationalize a construct, ideally a panel of experts familiar with the construct will assess the relevancy of the items being considered for inclusion in a measure (DeVellis, 2012). Experts can evaluate the relevancy and clarity of items and make recommendations for additional ones. This process is part of providing evidence of validity for scale items, particularly face validity. Lawshe (1975) recommended asking a sample of experts the following question with respect to each proposed item:

> Is the skill (or knowledge) measured by this item
>
> - Essential
> - Useful, but not essential, or
> - Not necessary
>
> to the performance of the construct? (p. 567)

Higher scores for an item indicate higher levels of agreement (interrater consistency) and suggest the item should be included in the final measure (see Wilson, Pan, & Schumsky, 2012).

Construct Validity

Construct validity is the degree to which one can make valid inferences from the operationalization of a construct to the theoretical construct upon which the measurements are based. As Cronbach and Meehl (1959) defined it, construct validity is the degree to which "a test [can] be interpreted as a measure of some attribute or quality which is not 'operationally defined'" (p. 175); that is, "When an investigator believes that no criterion available to him is fully valid, he perforce becomes interested in construct validity" (p. 176). Whereas the focus of criterion-oriented validity is whether the test operates in line with theoretical predictions, the focus of construct validity is on the "trait or quality underlying the test" (p. 176).

Cronbach and Meehl (1959) proposed a nomological network approach to construct validity. The basic idea is that for a measure to obtain evidence for construct validity, the researcher must first construct a nomological network or "interlocking system of laws … some [of which] involve observables" (p. 187). Subsequent tasks are undertaken in order to "examine the relation between the total network of theory and observations" (Cronbach & Meehl, 1959, p. 188), which might include many of the elements of criterion-oriented validity (e.g., does the measure predict theoretically relevant observables?).

Although the nomological network approach provides a strong philosophical foundation, it was not until the introduction of the multitrait-multimethod (MTMM) matrix by Campbell and Fiske (1959) that researchers had a specific method for assessing two primary forms of construct validity evidence, namely, convergent and divergent validity. In order to show evidence for construct validity by the MTMM approach, you have to demonstrate (a) *convergent validity* by showing theoretically related measures are highly correlated and (b) *divergent validity* by showing measures of supposedly separate constructs are not highly correlated (for an example of this technique, see Bodie *et al.*, 2014).

A final way to provide construct validity evidence for multi-item measures is through confirmatory factor analysis (CFA). CFA can provide a range of useful estimates, including model fit, parameter values (i.e., factor loadings), internal consistency, and three types of error (random, specific factor, and transient). It is most typically used in listening research to provide evidence that items on an existing or newly constructed test are valid indicators of a single latent construct; when multidimensional scales are created, researchers can test whether the multiple dimensions proposed line up with those underlying collected data. Examples of using CFA to show evidence of construct validity can be found in the work on the revised Listening Styles Profile (LSP-R; see Profile 36).

CFA also can be used to provide evidence for (or against) discriminant validity. Like other social sciences, work on listening is plagued by construct proliferation or "the accumulation of ostensibly different but potentially identical constructs representing [listening] phenomena" (Shaffer, DeGeest, & Li, 2016, p. 1) (see, e.g., multiple measures of listening competence and listening style). Each measure seems to represent a different construct, although the similarity among these scales has only recently received empirical attention (Fontana, Cohen, & Wolvin, 2015). CFA can be used to assess the degree to which these different measures do, in fact, represent distinct constructs or whether they might be usefully combined or otherwise replaced.

Assessing discriminant validity begins with identifying a construct of interest, then the set of measures that exist to tap this construct (Shaffer *et al.*, 2016). Data are collected from a sample of participants, and various measures estimating the degree of discriminant validity among each of these measures are generated. In their study, Shaffer *et al.* (2016) asked 220 working adults to answer questions derived from 13 measures of various leadership constructs, all of which share some amount of conceptual similarity (see Leader-Member Exchange [LMX-7], Profile 30). CFA allowed them to estimate correlations between these putatively distinct constructs, correcting for different types of random error. Their results suggested that many of the scales are redundant (for further discussion, see Chapter 5).

Responsiveness Validity

Responsiveness validity is the degree to which a measure, when designed to do so, is able to detect change over time (Beaton, Bombardier, Katz, & Wright, 2001; Husted, Cook, Farewell, & Gladman, 2000). It is particularly relevant for clinical settings or education settings in which participants are predicted to improve on some set of skills (e.g., become better or more aware listeners; Ford, Wolvin, & Chung, 2000). When a measure is administered before and after some treatment that has some known level of effectiveness, the ability of that measure to detect change is often referred to as *internal responsiveness*. For instance, if a university has implemented a listening center and demonstrated that students report being better listeners or being more aware of their listening after visiting it, this treatment could be used to test the representative validity of a new assessment of listening competence. *External responsiveness* is assessed by evaluating, for instance, how changes in the new measure of listening comprehension correlate with other measures of listening improvement. Essentially, to exhibit evidence for representative validity, a measure must be constructed with items that will be responsive to change. Thus, relatively stable listening traits and attitudes are not likely (nor do they need) to exhibit this type of validity; however, facets that should change with practice and training (e.g., listening comprehension) should be evaluated against sensitivity to detecting change. Example measures that should exhibit evidence of responsiveness validity include the WBLT and the listening test of the Internet-based Test of English as a Foreign Language (TOEFL iBT) (see Profiles 64 and 61, respectively).

Administering Items to a Development Sample

Of course, when developing a validity portfolio for some operationalization of a listening construct, you must collect data from actual respondents. Ideally, we collect data from a sample of participants that is representative of some larger population. These data are used to make a case for score reliability and provide evidence for validity (as outlined in this chapter), but these conclusions are only as good as the sample taken.

The only way to achieve a representative sample is to engage in some form of *random sampling*, whereby the probability of selection is equal across persons; that is, no one person has a greater or lesser chance of being a study participant (for a thorough treatment of sampling techniques, see Thompson, 2012). Most listening research falls well short of this criterion (Keaton & Bodie, 2013), although some studies come closer than others. For instance, studies that collect data from subject pools on college campuses that contain students enrolled in general education courses come closer to random sampling of students on that campus than those that draw students from classes associated with a specific major.

Every study seeking validity evidence for a scale is ultimately trying to make some type of inference about the construct of interest. In this section, we discuss the most common types—population and process inferences—and how to best make a case for either when collecting data for random and nonrandom samples of participants.

Population Inferences

Adequate sampling is vitally important when making *population inferences*, which are "the practice of making a statistical statement about a collection of objects from a subset of that collection" (Hayes, 2005, p. 31). Many of the claims we want to make about listening take the form of population inferences. For example:

- What barriers to listening do students experience in the classroom?
- Are better leaders also better listeners?
- Do parents who paraphrase their children have closer relationships with them?

Each of these inferences tries to make a generalization about a particular population. But perhaps the best example of population inference in listening research is the time study (see Profile 60). Time studies try to estimate how much of a person's day is spent listening compared to engaging in other communicative activities. Rankin's original study (1926) asked 21 adults to record their communication activities every 15 minutes over the course of one or more days. Subsequent studies have attempted to replicate the Rankin study using slightly different methods.

Several aspects of these studies are relevant to the population inference they are trying to make. First, none of them utilized standard sampling procedures developed to generate samples that are representative of the overall population of interest. The Rankin (1926) study had merely 21 participants, and although other studies have gathered data from more individuals, they are usually localized to one particular organization or one particular college campus (and often, one particular course taught during one academic semester). If the goal of a time study is to generalize from the data to some population of interest, we want to do our best to make sure that the sample is representative of the population.

A sample is representative if it is similar to the population in all important aspects relevant to the research. Take, for instance, the time studies of Barker, Gladney, Edwards, Holley, and Gaines (1980) and Janusik and Wolvin (2009). Each was interested in the time spent in various communicative activities by college students, and each sampled from their own campus community (Auburn and Maryland, respectively). Because of their interest in college students, their choice to limit data collection to college students is understandable. Limiting the collection to one institution, however, is not. Getting a more representative estimate of the population of college students would require sampling from a number of institutions (for a slightly better strategy, see Emanuel *et al.*, 2008). Auburn and Maryland are large, public universities. Other types of institutions, such as private universities and community colleges, are not represented. If there is reason to believe that the individuals sampled in these two studies are different from individuals who were not sampled, then the estimates provided in these studies do not serve their intended purpose. Of course, there are times when a researcher might be interested in generalizing to one particular student body (e.g., when developing a curriculum for students at a specific institution).

Process Inferences

Without representative samples, population inferences are not valid. Sometimes, however, research is not conducted to make population inferences, but assesses the degree to which data conform to theoretical predictions. Research attempting to make some type of *process inference* is less concerned with estimating the size of an effect (e.g., time spent listening) and more concerned with determining whether a prediction deduced from theory can be supported.

A host of studies interested in population inferences have observed the fact that there are many ways in which one individual can provide coping assistance to another who has experienced some sort of negative event. Research suggests that some ways of providing assistance are quite helpful to our health and well-being, whereas others are not (Barbee & Cunningham, 1995; Goldsmith, 2004; MacGeorge, Feng, & Burleson, 2011). One way of operationalizing quality (i.e., helpful) support is known as person centeredness. *Person centeredness* (PC) is a theoretical explanation for why some messages "work" better than others—they "take into account and adapt to the subjective, emotional, and relational aspects of communicative contexts" (Burleson, 2007, p. 113) (Table 2.2).

Although messages higher in PC tend to lead to a range of positive outcomes (High & Dillard, 2012), the impact of PC messages is moderated by several variables—sometimes we feel better after a high-PC message, but other times these messages don't make us feel quite as good. Bodie and Burleson (2008) have proposed that the impact of supportive behaviors on outcomes is both a function of the content of those behaviors and how they are processed by recipients. In other words, depending on how carefully a listener attends/listens to message content, that content will have more or less of an effect on his or her feeling better. If there is no scrutiny of message content (inattentive listening), then components of the message cannot help. If there is more scrutiny of

Table 2.2 Examples of messages that vary in person centeredness.

Statement/message definition	Sample message
***Highly person-centered message*: Statements that help a person see his or her feelings from a different point of view, and attempt to help him or her understand how these feelings are part of "the big picture."**	"I understand how bummed you must be—to try your best to learn statistics and … you know … to keep struggling, it's very frustrating. You might be thinking that it isn't worth all this aggravation. It certainly does not mean you aren't smart or anything like that. I know it's hard to see things differently, but maybe you have learned something here that can help next time."
***Moderately person-centered message*: Explanation of the event, without focusing on feelings, which tries to lower negative affect and often mentions justifications.**	"Learning statistics is difficult and lots of people don't get it; there are tons of people who cringe at math! I wish you had done better, but I understand how this happened. It's really tough. Maybe you just have trouble learning those formulas. Or maybe you need to work more examples. Your ability doesn't rest just on learning statistics."
***Low person-centered message*: Messages that completely disregard how a person is feeling, and often tell the other how to feel or suggest forgetting about the situation.**	"Forget about learning statistics. There are other more interesting things to learn. Nobody needs math anyway. Just don't think about it and find something else to do."

Note: Conceptual definitions of message types were adapted from Burleson and Samter (1985, p. 45). Example messages reflect those one might receive after failing to learn statistics and were adapted from the same source.
Source: Burleson and Samter (1985). Reproduced with permission of Taylor & Francis.

message content (attentive listening), then messages should have a greater ability to help (see Bodie & MacGeorge, 2015; Bodie & Jones, 2016).

To test the theory, the typical method is to recruit college students and expose them to stressors or have them talk about stressors with some type of listener (e.g., trained in active listening techniques; see Couples Helping Exercise, Profile 14). These studies regularly report robust support for predictions. As a result, we can conclude that the theory can help us understand how and why support works and when perhaps other modes of helping might be more productive. But what sorts of generalizable claims can we make here?

A critique often leveled at this type of work is that the study was conducted using students enrolled in a particular type of college course at one institution who were conveniently accessible to use for this purpose. Some critics might then argue that "the results are not generalizable to 'people in general' or even college students" and that "we still don't know much." To that, the only viable response is "Yes." At the same time, the goal of the study must be considered. In this example, the point is to test theoretical propositions. It does not matter if the participants were not randomly selected from some larger population of interest because the goal of the study was to test a process. We know the process works for at least this sample of data. Future work can now go about testing boundary conditions for the theory. Although it is possible that different types of people are affected differently by supportive messages, the mechanisms underlying effects should be rather consistent. Focusing on surface-level similarities of a sample is too simplistic in this case. The theoretical mechanisms responsible for the association between two variables are universal, and so the representativeness of the population is less of a concern than the internal components of the study (e.g., whether experimental manipulations were carried out in valid ways).

Reporting Numerical Data

When the focus of a published report is on developing a scale for the measurement of some component of listening, scholars often rely on some form of statistical inference to guide decisions. This section explicates two classes of statistics that are often used in research on listening and makes eight recommendations for reporting numerical data (Keaton & Bodie, 2013).

Two Classes of Statistics

Descriptive statistics present summaries about the sample involved in an analysis. Although rarely used to make population inferences, they often tell an interesting story about sample data not available from inferential statistics. Univariate descriptives include information about central tendency (mean, median, and mode), range, variability (variance and standard deviation), and shape (skewness and kurtosis). Bivariate descriptives comprise information in the form of cross-tabs, scatterplots, measures of dependence (e.g., Pearson's r), covariance (which reflects measurement scales for the variables), and slope (a one-unit change in the DV for a one-unit change in the IV). All of these statistics are sample dependent; that is, they are simply descriptions of important characteristics of the collected data.

When the goal is to evaluate hypotheses and draw generalizable conclusions from gathered data to a larger population, scholars will use *inferential statistics.* Statistical inference requires making assumptions that are based upon probability theory, allowing researchers to make predictions. When researchers can assert that their samples are approximately normally distributed, fully parametric statistical tests may be implemented. If, however, a sample is nonparametric (not approaching normal or using only nominal or ordinal data), there are corresponding statistical tests that may be substituted. Using parametric statistics on nonnormal samples can present a variety of issues. Researchers should test for assumptions of normality when using inferential statistics, and any violations of those assumptions should be reported. In addition, procedures concerning the frequency or percentage of missing data should also be discussed. Finally, null hypothesis significance testing (NHST) is only a beginning point, and other reporting elements such as effect sizes and confidence intervals should also be incorporated (Levine, Weber, Hullett, Park, & Lindsey, 2008).

Use of Statistical Data in Listening Research

Keaton and Bodie (2013) found that the use of numerical data to make principled arguments about listening is relatively common, comprising 45.8% of all published material in the first 25 years of the *International Journal of Listening* (*IJL*). They also found that reporting practices are not always followed, with basic sample characteristics like age, ethnicity, and biological sex distribution often unreported. Likewise, contrary to recommendations by the APA style manual (which are followed by most social scientific publications, including the *IJL*), there is an overreliance on NHST with much less reporting of effect sizes and confidence intervals. Other reporting practices inconsistent with APA recommendations included the infrequent reporting of basic descriptive statistics (e.g., measures of central tendency and variability), a complete lack of focus on the shape of sample distributions, a tendency not to report tests relevant to statistical assumptions (e.g., normality), a lack of clarity with regard to missing data, and some noteworthy misappropriations of statistical techniques. Given that these concerns are important for generating valid inferences about listening, we conclude this chapter by presenting a synopsis of guidelines offered by Keaton and Bodie (2013; used with permission). These recommendations provide an essential guide for producing listening research that can sustain valid and generalizable conclusions.

Recommendation 1: Look at Your Data

Following data collection and prior to running any statistics, researchers should assess their data. Visual inspection of data is important for at least two essential reasons. First, it allows researchers to identify trends and patterns in their data (Levine, 2011). Listening scholars can learn a great deal from criminologists and other social scientists who utilize visual displays of data along with descriptive statistics to make principled arguments. The same is true for those employing descriptive or inferential statistics. In both cases, descriptive data can be informative and can provide information beyond the story that inferential statistics can tell. There are excellent examples of communication scholars who utilize descriptives to tell interesting stories and who regularly publish in the top journals (e.g., Levine, 2010).

Second, visual inspection and related descriptive statistics can help researchers spot violations of one or more assumptions underlying an inferential procedure. Choosing the best test for respective research questions or hypotheses benefits the researcher and aids in replicability. The use of most of these statistical methods begins by assuming that their samples are normally distributed. When sample distributions are significantly skewed or suffer from positive or negative kurtosis, assumptions derived from statistics assuming normal distributions can be invalid. Many of these issues can be alleviated with a large enough sample size. For instance, in a large enough sample, skewness does not digress enough from normality to create noteworthy differences in analyses. Furthermore, positive ($n > 100$) and negative ($n > 200$) underestimates of variance associated with kurtosis wane with large enough samples (Tabachnick & Fidell, 2007). Unfortunately, Keaton and Bodie (2013) found that virtually all studies published in *ILJ* between 1987 and 2011 did not report tests for normality (no matter the sample size).

Although many maintain that parametric statistical models may still be used providing the deviations from normality are not acute (Hubbard, 1978), especially given that many nonparametric tests lack versatility in multivariate situations (Nunnally, 1978), serious consequences still can result should the sample exhibit a distribution that is not close to normal. Using parametric statistics based on t, F, or χ^2 to generalize findings from sample distributions not approaching normal can, among other outcomes, compromise the estimation of coefficients and confidence intervals. Therefore, researchers should test for normality (both graphically and with descriptive statistics) and use these tests primarily when the sample distribution is approximately normal.

Recommendation 2: Report Effect Sizes and Confidence Intervals

Very few reputable scholars deny that the social and behavioral sciences are marked by an overreliance on NHST (Cohen, 1994). When engaged in NHST, the researcher is ultimately making a dichotomous judgment. Merely proclaiming statistical significance does not provide a complete picture of the results of a study. Real science is concerned with finding the magnitude of an effect, not with a dichotomous decision rule regarding whether the null (and usually nil) hypothesis is a valid assumption (Ziliak & McCloskey, 2009). Moreover, given a large enough sample, the conclusion drawn from inspecting a p-value is meaningless (Meehl, 1990). Indeed, merely reporting a p-value and claiming a result is "statistically significant" gives readers no indication as to the clinical or practical significance of the results, and failure to discuss the latter limits future attempts to replicate (or refute) results or to conduct meta-analyses.

Although probability values convey the likelihood of a Type I error (incorrectly rejecting the null—i.e., a false positive), they do not reveal the weight of an effect. When a significant effect is detected in a sample, we can, with a degree of certainty, claim that this sample is not derived from a population where this effect is zero. In this case, the null can be rejected, and the researcher can reasonably claim that the sample in question can be generalized as representative of a population where an effect does in fact occur. Because probability values do not give us information concerning the size of an effect, reporting effect sizes and confidence intervals becomes essential to developing a credible science of listening.

Effect sizes and confidence intervals also give consumers of scientific inquiry a basis for deciding if a study is practically significant rather than only statistically significant. Statistical significance does not necessarily imply that findings are of consequence, and nonsignificant results are not necessarily unimportant. Effect sizes and confidence intervals help deflate the overvalued importance of statistical significance and allow for nonsignificant findings that may have practical significance to see the light of day in journal space. Reporting effect sizes along with other descriptive statistics like measures of central tendency (e.g., means) and variability (e.g., standard deviation) as well as the inferential probability that a population mean, for instance, lies between two values (i.e., a confidence interval) also allows researchers to double-check the results of studies or perform meta-analyses of many studies. For results of one study to be compared to those of another and to ultimately build a cumulative body of knowledge, scholars need to know the practical and theoretical importance of findings and how these findings can be interpreted given what we already know. For practitioners to derive any set of best practices from scientific research, they need to appropriately know effect sizes and confidence intervals—that is, they need to be able to discern not the statistical but the practical significance of study results.

Recommendation 3: Psychometrically Validate Scores Derived from the Use of Instruments

Researchers often make the mistake of assuming that existing scales, especially those with a rich history, have been "previously validated" and should thus be treated differently than newer or less established scales. This belief does not, however, reflect sound scientific practice (Levine *et al.*, 2006). Validity is an ongoing process; scales are not "valid" or "invalid," but can be said to have more or less robust validity portfolios. Importantly, scales often exhibit different properties when utilized with different populations (Little, 1997). As such, authors are advised to report the psychometric properties of data derived from the use of instruments, irrespective of the status of the instrument.

This third recommendation is particularly important for instruments that are assumed to have vast validity portfolios. For example, in a recent study, Bodie, Worthington, and Fitch-Hauser (2011) reported data inconsistent with the measurement model of the WBLT-Form C (WBLT-C). In particular, the reported data showed that items on the WBLT-C were largely uncorrelated with each other ($r_{ave} = .03$) and that no pattern of association among items could explain the small amount of shared variance that did exist. Ultimately, the WBLT-C consists of 40 unrelated multiple-choice items (see Profile 64). A similar project recently assessed the LSP-16, with results suggesting major modifications of the scale are needed; when made, these changes resulted in a much more potentially valid scale (Bodie *et al.*, 2013; see Profile 36).

Another noteworthy problem concerns the use of scales exhibiting reliability estimates that do not meet recommended criteria. In their review, Keaton and Bodie (2013) found that the average reported Cronbach's alpha across all *IJL* articles that reported this statistic was .71. Customary evaluative criteria are often in the range of $0.7 \leq \alpha < 0.8$ for acceptable values, with $0.6 \leq \alpha < 0.7$ deemed questionable (Kline, 1999). Higher values of internal consistency are universally recognized as more desirable than lower values. Not only do low levels of internal consistency attenuate relationships between variables and differences between groups, but also they are vitally important when making practical recommendations from studies (Nunnally [1978] is our recommended source).

Most listening scholars ultimately want their research to be useful, to help themselves or others improve the lives of the everyday people about whom we theorize and for whom our work should be targeted. To achieve their goals, listening scholars must create or utilize instruments that produce strong estimates of internal consistency and utilize instruments that have been adequately vetted. Whatever the chosen evaluative method, it is clear that listening researchers need to take greater care in operationalizing listening constructs.

Recommendation 4: Correctly Utilize and Report Factor Analytic Techniques

As reported by Keaton and Bodie (2013), listening scholars often make problematic choices when conducting factor analyses (FAs). The two more commonly utilized procedures (in published articles of the *ILJ* that identified the type of FA) were exploratory factor analysis (EFA) and principal component analysis (PCA). EFA attempts to discern some sort of underlying structure of latent variables by grouping observed variables that are correlated. EFA utilizes only shared variance, whereas PCA uses all of the variance in the data. Consequently, the resultant factors from EFA are thought to be explanatory mechanisms. Components gleaned from PCA, however, are only descriptive—not inferential—groupings of associated items. Using PCA to deduce factors is an inefficient and incorrect use of the method. It is more appropriate to use PCA to reduce the number of items in exploratory scale development (Park, Dailey, & Lemus, 2002).

When the goal is to develop underlying explanatory frameworks of latent, correlated variables, EFA and not PCA is a more fitting method, and reporting the specific procedure is essential in aiding the replicability of the study. When the goal is to report the psychometric properties of scores derived from established measures, then CFA is the preferred technique. Readers interested in more details about EFA and CFA are directed to several sources (Kline, 2005; Levine, 2005; Levine *et al.*, 2006; Raju *et al.*, 2002; Thompson, 2004).

Recommendation 5: Match Sampling to the Population of Interest

The next major point of interest concerns the homogeneity of the overall sample of participants. A typical participant for a study published in *IJL* is a white, 23-year-old college female (usually a freshman or sophomore) from the United States. Although white, middle-class, college-educated, young Americans are a viable population from which to learn about basic structures and functions of listening (Shapiro, 2002), more heterogeneous samples from a variety of cultural contexts are needed to make valid population inferences.

The people, stimuli, and events that a study seeks to illustrate or draw inferences about affect nearly every conclusion of a given investigation; thus, researchers should provide information regarding the individuals included in a given inquiry (e.g., reporting ranges, central tendencies, and variability of samples). Doing so aids in determining homogeneity in samples and replicating research. Reporting means and standard deviations, for instance, helps readers determine exactly what type of sample is being discussed and generalized, and what types of samples need to be tested in future research.

Recommendation 6: Be Clear as to the Implications of Study Results

Here, the issue is our ability to make causal claims about listening (e.g., the antecedents and consequences of listening in particular ways, whether individual differences in listening reliably produce differences in processing, etc.). A review of research published in the *ILJ* reveals an overreliance on cross-sectional research and a striking lack of experimental and longitudinal studies. The overreliance on cross-sectional research is likely one reason that listening research is heavily atheoretical (Bodie, 2009, 2010b, 2011b, 2012; Wolvin, Halone, & Coakley, 1999). At minimum, scholars who pursue cross-sectional research should include a discussion of the limitations of such research and speculate about the theoretical structure among the variables of interest. Ideally, researchers will expand their methods to include more experimental and longitudinal research (for further discussion of these techniques, see Chapter 5).

Recommendation 7: Limit Reliance on Self-Report Measures

In a similar vein, there is an overreliance on self-report measures of listening. Although self-reporting listening is certainly not universally inappropriate—for instance, the Listening Concepts Inventory (see Profile 32) assesses individual conceptualizations of listening akin to the work by O'Keefe (1988) on implicit theories of communication (i.e., message design logics)—most scales are aimed at assessing the general enactment of specific behaviors. For instance, the Self-Perceived Listening Competence Scale (SPLCS; see Profile 57) includes items such as "I can interpret persons' facial expressions correctly." Attempts to assess whether self-reported listening behaviors are associated with the actual enactment of those behaviors are rare (Bodie *et al.*, 2014). Indeed, most studies utilizing self-reports of listening behaviors do not attempt to empirically dismiss other plausible explanations for found associations among measures of listening and important antecedents and consequences, such as common method variance (Podsakoff, MacKenzie, & Podsakoff, 2012). Other research assumes that different perspectives (e.g., direct supervisors versus peers) are driving variability in scores without submitting such speculations to full tests (Cooper & Husband, 1993). As has been pointed out by others, listening is a socially desirable behavior, perhaps even more susceptible to social desirability effects than other communicative actions (Lawson & Winkelman, 2003). Moreover, there are readily available statistical (e.g., structural equation modeling) and methodological (e.g., round-robin designs and multitrait-multimethod studies) techniques that listening scholars can utilize to address these issues.

Recommendation 8: Avoid the Use of Intact Classes or Groups

Listening scholars have a propensity to use intact classes during data collection efforts (Keaton & Bodie, 2013), which has potentially serious ramifications for internal validity. Selecting participants from intact groups can result in selection bias, which can affect the ability of a study to detect a true relation between the IVs and DVs. Because the participants are from an intact group, the effect of X on Y may in fact be due to another variable that every participant is exposed to equally. In experimental work using intact classes, the IV can no longer truly be said to be *independent* because the researcher is not determining the level of the variable that each subject will experience.

If participants were assigned to the intact group using preexisting information about them, then the sample is not random. If comparing two intact groups, the groups may consistently differ because of the nonrandom assignment and not because of the relationship between X and Y. Participants assigned to groups because of ability also are guilty of this problem; however, intact groups may be considered to be experimental if the participants were randomly assigned to the intact groups prior to the experiment. For instance, an introductory general education communication course containing 100 students from all over the university is closer to representing all college students than an upper-level research methods course of 25 communication majors. Even in the former case, however, there are issues. For instance, if a researcher is interested in a particular training protocol on abilities to retain information, using existing classes and providing one class with the training while treating the other as the control group ultimately conflates training with characteristics of the teacher, time of the class, and other potential nuance variables. The better strategy is to randomly assign participants to groups and to control or measure any extraneous variables not of primary interest.

Conclusion

The purpose of this chapter was to provide some guidelines for researchers interested in developing operationalizations of listening constructs. After settling on a conceptual definition, empirically oriented scholars have to move quickly to craft suitable operational definitions, then measure facets of listening and report data to make principled claims. As indicated by the Keaton and Bodie (2013) review, listening scholars can do much to advance listening research and theory by how they design and report their studies. We encourage you to follow the guidelines in this chapter and to conduct the most rigorous of research. Doing so is in our own best interests as researchers and scholars seeking to fully establish listening as its own unique area of communication research.

References

Barbee, A. P., & Cunningham, M. R. (1995). An experimental approach to social support communications: Interaction coping in close relationships. *Communication Yearbook, 18*, 381–413.

Barker, L., Gladney, K., Edwards, R., Holley, F., & Gaines, C. (1980). An investigation of proportional time spent in various communication activities by college students. *Journal of Applied Communication Research, 8*, 101–109. doi:10.1080/00909888009360275

Beaton, D. E., Bombardier, C., Katz, J. N., & Wright, J. G. (2001). A taxonomy for responsivenes. *Journal of Clinical Epidemiology, 54*, 1204–1217. doi:10.1016/S0895-4356(01)00407-3

Bodie, G. D. (2009). Evaluating listening theory: Development and illustration of five criteria. *International Journal of Listening, 23*, 81–103. doi:10.1080/10904010903014434

Bodie, G. D. (2010a). The Revised Listening Concepts Inventory (LCI-R): Assessing individual and situational differences in the conceptualization of listening. *Imagination, Cognition, and Personality, 30*, 301–339. doi:10.2190/IC.30.3.f

Bodie, G. D. (2010b). Treating listening ethically. *International Journal of Listening, 24*, 185–188. doi:10.1080/10904018.2010.513666

Bodie, G. D. (2011a). The Active-Empathic Listening Scale (AELS): Conceptualization and evidence of validity within the interpersonal domain. *Communication Quarterly, 59*, 277–295. doi:10.1080/01463373.2011.583495

Bodie, G. D. (2011b). The understudied nature of listening in interpersonal communication: Introduction to a special issue. *International Journal of Listening, 25*, 1–9. doi:10.1080/10904018.2011.536462

Bodie, G. D. (2012). Listening as positive communication. In T. Socha & M. Pitts (Eds.), *The positive side of interpersonal communication* (pp. 109–125). New York: Peter Lang.

Bodie, G. D., & Burleson, B. R. (2008). Explaining variations in the effects of supportive messages: A dual-process framework. *Communication Yearbook, 32*, 355–398.

Bodie, G. D., & Jones, A. (2016). Dual-process theory of supportive message outcomes. In C. R. Berger & M. E. Roloff (Eds.), *International encyclopedia of interpersonal communication*. [Online First]. Hoboken, NJ: Wiley-Blackwell.

Bodie, G. D., Jones, S. M., Vickery, A. J., Hatcher, L., & Cannava, K. (2014). Examining the construct validity of enacted support: A multitrait-multimethod analysis of three perspectives for judging immediacy and listening behaviors. *Communication Monographs, 81*, 495–523. doi:10.1080/03637751.2014.957223

Bodie, G. D., & MacGeorge, E. L. (2015). Theories of supportive communication. In D. O. Braithwaite & P. Shrodt (Eds.), *Engaging theories of interpersonal communication* (2nd ed.). Thousand Oaks, CA: Sage.

Bodie, G. D., Pence, M. E., Rold, M., Chapman, M. D., Lejune, J., & Anzalone, L. (2015). Listening competency in initial interactions, II: Applying trait centrality to discover the relative placement of listening competence among implicit competency theories. *Communication Studies, 66*, 528–548. doi:10.1080/10510974.2015.1039657

Bodie, G. D., St. Cyr, K., Pence, M., Rold, M., & Honeycutt, J. M. (2012). Listening competence in initial interactions I: Distinguishing between what listening is and what listeners do. *International Journal of Listening, 26*, 1–28. doi:10.1080/10904018.2012.639645

Bodie, G. D., Worthington, D. L., & Fitch-Hauser, M. (2011). A comparison of four measurement models for the Watson-Barker Listening Test (WBLT)-Form C. *Communication Research Reports, 28*, 32–42. doi:10.1080/08824096.2011.540547

Bodie, G. D., Worthington, D. L., & Gearhart, C. C. (2013). The Revised Listening Styles Profile (LSP-R): Development and validation. *Communication Quarterly, 61*, 72–90. doi:10.1080/01463373.2012.720343

Brown, J. I., & Carlsen, G. R. (1955). *Brown-Carlsen Listening Comprehension Test*. New York: Harcourt, Brace and World.

Burleson, B. R. (2007). Constructivism: A general theory of communication skill. In B. B. Whaley & W. Samter (Eds.), *Explaining communication: Contemporary theories and exemplars* (pp. 105–128). Mahwah, NJ: Erlbaum.

Burleson, B. R., & Samter, W. (1985). Individual differences in the perception of comforting messages: An exploratory investigation. *Central States Speech Journal, 36*, 39–50. doi:10.1080/10510978509363198

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multi-trait multi-method matrix. *Psychological Bulletin, 52*, 281–301. doi:10.1037/h0046016

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49*(12), 997–1003. doi:10.1037/0003-066X.49.12.997

Cooper, L., & Husband, R. L. (1993). Developing a model of organizational listening competency. *Journal of the International Listening Association, 7*, 6–34.

Cooper, P. (1998). *The speaking, listening, and media literacy standards and competency statements for K–12 education*. Washington, DC: National Communication Association.

Cronbach, L., & Meehl, P. E. (1959). Construct validity in psychological tests. *Psychological Bulletin, 52*, 174–203. doi:10.1037/h0040957

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Thousand Oaks, CA: Sage.

Drollinger, T., Comer, L. B., & Warrington, P. T. (2006). Development and validation of the Active Empathetic Listening Scale. *Psychology and Marketing, 23*, 161–180. doi:10.1002/mar

Emanuel, R., Adams, J., Baker, K., Daufin, E. K., Ellington, C., Fitts, E.,…Okeowo, D. (2008). How college students spend their time communicating. *International Journal of Listening, 22*, 13–28. doi:10.1080/10904010701802139

Fontana, P. C., Cohen, S. D., & Wolvin, A. D. (2015). Understanding listening competency: A systematic review of research studies. *International Journal of Listening, 29*, 148–176. doi:10.1080/10904018.2015.1015226

Ford, W. S. Z., Wolvin, A. D., & Chung, S. (2000). Students' self-perceived listening competencies in the basic speech communication course. *International Journal of Listening, 14*, 1–13. doi:10.1080/10904018.2000.10499032

Gearhart, C. C., Denham, J. D., & Bodie, G. D. (2014). Listening as a goal directed activity. *Western Journal of Communication, 78*, 668–684. doi:10.1080/10570314.2014.910888

Goldsmith, D. J. (2004). *Communicating social support.* New York: Cambridge University Press.

Goldsmith, D. J., McDermott, V. M., & Alexander, S. C. (2000). Helpful, supportive, and sensitive: Measuring the evaluation of enacted support in personal relationships. *Journal of Social and Personal Relationships, 17*, 369–391. doi:10.1177/0265407500173004

Haas, J. W., & Arnold, C. L. (1995). An examination of the role of listening in judgments of communication competence in co-workers. *Journal of Business Communication, 32*, 123–139. doi:10.1177/002194369503200202

Hayes, A. F. (2005). *Statistical methods for communication science*. Mahwah, NJ: Erlbaum.

High, A. C., & Dillard, J. P. (2012). A review and meta-analysis of person-centered messages and social support outcomes. *Communication Studies, 63*, 99–118. doi:10.7080/10510974.2011.598208

Hubbard, R. (1978). The probable consequences of violating the normality assumption in parametric statistical analysis. *Area, 10*, 393–398.

Husted, J. A., Cook, R. J., Farewell, V. T., & Gladman, D. D. (2000). Methods for assessing responsiveness: A critical review and recommendations. *Journal of Clinical Epidemiology, 53*, 459–468. doi:10.1016/S0895-4356(99)00206-1

Imhof, M., & Janusik, L. A. (2006). Development and validation of the Imhof-Janusik Listening Concepts Inventory to measure listening conceptualization differences between cultures. *Journal of Intercultural Communication Research, 35*, 79–98. doi:10.1080/17475750600909246

Janusik, L. A., & Wolvin, A. D. (2009). 24 hours in a day. A listening update to the time studies. *International Journal of Listening, 23*, 104–120. doi:10.1080/10904010903014442

Keaton, S. A., & Bodie, G. D. (2013). The statistical and methodological acuity of scholarship appearing in the *International Journal of Listening* (1987–2011). *International Journal of Listening*, 115–135. doi:10.1080/10904018.2013.813206

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Fort Worth, TX: Harcourt College.

Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. New York: Methuen.

Kline, P. (1999). *The handbook of psychological testing* (2nd ed.). London: Routledge.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*, 563–575. doi:10.1111/j.1744-6570.1975.tb01393.x

Lawson, M., & Winkelman, C. (2003). The social desirability factor in the measurement of listening skills: A brief report. *Counseling Psychology Quarterly, 16*, 43–45. doi:10.1080/0951507021000050212

Levine, T. R. (2005). Confirmatory factor analysis and scale validation in communication research. *Communication Research Reports, 22*, 335–338. doi:10.1080/00036810500317730

Levine, T. R. (2010). A few transparent liars: Explaining 54% accuracy in deception detection experiments. *Communication Yearbook, 34*, 41–61.

Levine, T. R. (2011). Quantitative social science methods of inquiry. In M. Knapp & J. Daley (Eds.), *Handbook of interpersonal communication* (4th ed., pp. 25–58). Thousand Oaks, CA: Sage.

Levine, T. R., Hullett, C. R., Turner, M. M., & Lapinski, M. K. (2006). The desirability of using confirmatory factor analysis on published scales. *Communication Research Reports, 23*, 309–314. doi:10.1080/08824090600962698

Levine, T. R., Weber, R., Hullett, C. R., Park, H. S., & Lindsey, L. L. M. (2008). A critical assessment of null hypothesis significance testing in quantitative communication research. *Human Communication Research, 34*, 171–187. doi:10.1111/j.1468-2958.2008.00318.x

Little, T. D. (1997). Mean and covariance structures (MACS) analysis of cross-cultural data: Practical and theoretiacl issues. *Multivariate Behavioral Research, 32*, 53–76. doi:10.1207/s15327906mbr3201_3

MacGeorge, E. L., Feng, B., & Burleson, B. R. (2011). Supportive communication. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (4th ed., pp. 317–354). Thousand Oaks, CA: Sage.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports, 66*, 195–244.

Nichols, R. G. (1948). Factors in listening comprehension. *Speech Monographs, 15*, 154–163. doi:10.1080/03637754809374953

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

O'Keefe, B. J. (1988). The logic of message design: Individual differences in reasoning about communication. *Communication Monographs, 55*, 80–103. doi:10.1080/03637758809376159

Park, H. S., Dailey, R., & Lemus, D. (2002). The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research, 27*, 562–577. doi:10.1111/j.1468-2958.2002.tb00824.x

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569. doi:10.1146/annurev-psych-120710-100452

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*, 517–529. doi:10.1037//0021-9010.87.3.517

Rankin, P. T. (1926). *The measurement of the ability to understand spoken language. Dissertation Abstracts, 12*(1952), 847.

Shaffer, J. A., DeGeest, D., & Li, A. (2016). Tackling the problem of construct proliferation: A guide to assessing the discriminant validity of conceptually related constructs. *Organizational Research Methods, 19*, 80–110. doi:10.1177/1094428115598239

Shapiro, M. A. (2002). Generalizability in communication research. *Human Communication Research, 28*, 491–500. doi:10.1111/j.1468-2958.2002.tb00819.x

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677–680. doi:10.1126/science.103.2684.677

Strahan, R., & Gerbasi, K. C. (1972). Short, homogenous versions of the Marlow-Crowne Social Desirability Scale. *Journal of Clinical Psychology, 28*, 191–193. doi:10.1002/1097-4679(197204)28:2<191::AID-JCLP2270280220>3.0.CO;2-G

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: APA.

Thompson, S. K. (2012). *Sampling*. Hoboken, NJ: Wiley.

Wengraf, T. (2001). *Qualitative research interviewing: Biographic narrative and semi-structured methods*. London: Sage.

Wilson, F. R., Pan, W., & Schumsky, D. A. (2012). Recalculation of the critical values for Lawshe's content validity ratio. *Measurement and Evaluation in Counseling and Development, 45*, 197–210. doi:10.1177/0748175612440286

Wolvin, A. D., & Coakley, C. G. (1994). Listening competency. *Journal of the International Listening Association, 8*, 148–160. doi:10.1080/10904018.1994.10499135

Wolvin, A. D., Halone, K. K., & Coakley, C. G. (1999). An assessment of the "intellectual discussion" on listening theory and research. *International Journal of Listening,, 13*, 111–129. doi:10.1080/10904018.1999.10499030

Ziliak, S. T., & McCloskey, D. N. (2009, April 3). *The cult of statistical significance*. Paper presented at the Joint Statistical Meetings (JSM), Washington, DC.

3

Qualitative Approaches and Listening Research

Barbara Cook Overton

One afternoon a few years ago, I sat with my dissertation advisor,[1] and together we discussed methodologies. I was grappling with ways that I might design a study that would allow me to gauge changes in the communication patterns of emergency room workers after the introduction of an electronic medical records (EMRs) system. I considered pre- and posttest designs that might incorporate analysis of variance tests or interrupted time series. Each would have allowed me to make comparisons and measure change. Although these approaches would have produced definitive "before" and "after" portraits of the emergency room that I wanted to study, my advisor encouraged me to consider also capturing many of the moments that would occur *between* those definitive portraits. She explained that the designs I had been considering were useful for taking snapshots—moments frozen in time—but that other approaches, like ethnography, would allow me to string together countless moments, much like a movie. Movies are, after all, an assemblage of thousands of still pictures whizzing past us at 24 frames per second. Being, among other things, a documentary filmmaker, her explanation resonated with me.

Consider, for a moment, your favorite movie. Think of the opening scene. Now think of the opening *shot*. Next, think of the last image on screen before the credits roll. If you strip away everything between the opening and closing shots, can you piece together the story based on those two images? Probably not. Those images *do* tell you a lot, perhaps about the characters (e.g., their gender, ethnicity, and age) or the setting (e.g., space odyssey or historical drama), but they don't tell the whole story. To be fair, researchers interested in change can incorporate more than a single before-and-after measurement profile to capture and document numerical, time-based change (see Chapter 5). Even so, each of these points of measurement produce images—snapshots—that tell us a lot, but do not reveal the entire picture. What many scholars label as *qualitative methods* help fill in these missing blanks.

The goal of this chapter is to showcase the utility and rigor of qualitative methods, methods largely underrepresented in the listening literature. I begin by describing characteristics of qualitative methods, discuss likely reasons why listening scholars do not use qualitative methods regularly, then describe methodological approaches that have appeared in the listening literature. The chapter concludes with a section on "doing" ethnography, including

1 I am grateful to my dissertation advisor, Dr. Loretta Pecchioni, for her insightful analogy.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

descriptions of fieldwork, participant observation, participant listening, in-depth interviews, coding and analyzing data, and suggestions for writing ethnographic reports.

Characteristics of Qualitative Inquiry

Tracy and Muñoz (2011) noted many qualitative methodologies used to "define themselves by what they were *not*" (p. 60)—that is, qualitative methodologies were *not* about numbers or statistics. Definitions of qualitative inquiry centered on the kind of data collected (i.e., narrative- or word-driven data), data collection techniques (e.g., interview or participant observation), and the relative advantage of words over numbers for capturing and explaining communication phenomena. At one time, such definitions made sense because many communication researchers "questioned the value of qualitative approaches" (Tracy & Muñoz, 2011, p. 60), and qualitative researchers often felt compelled to defend or justify their methodological choices. Now, it is widely accepted that qualitative and quantitative inquiry contribute *equally* to communication research—neither approach is inherently better nor more complete than the other. Each approach has "something to offer the other," and the two approaches are, in fact, "complementary" (Purdy, 2010, p. 36). Moreover, "the combination of qualitative and quantitative approaches provides a more complete understanding … than either approach alone" (Creswell, 2014, p. 4); combining methods is, arguably, stronger "than a single method because the supplemental component enhances validity … by enriching or expanding our understanding or by verifying our results from another perspective" (Morse & Niehaus, 2009, p. 14). Despite this advantage, and as Purdy (2000) lamented, "the tendency of most listening researchers is to follow the pattern of … statistical study" (p. 48). In this chapter, I describe four basic characteristics of qualitative methodologies and discuss some misconceptions about qualitative research—in so doing, I hope to dispel doubt about the utility of qualitative inquiry for listening researchers.

I limit my discussion to what I believe are the four most basic components of qualitative work: work that focuses on word-driven data, is subjective, is nonlinear, and takes place in situ. First, qualitative researchers prioritize singular occurrences (Wolcott, 2010) and techniques like participant observation and in-depth interviews to produce word-driven data from which patterns can be extrapolated. Open-ended, probing questions posed to smaller samples generate data that "are meaningless without a researcher who gives meaning" (Purdy, 2011, p. 134). Second, most qualitative data are interpreted subjectively—qualitative investigators will "deliberately use [their] own responses to the phenomena under investigation as one source of data" (Philipsen, 1982, p. 9).

Investigators' experiences "vis-à-vis the subjects of the inquiry" (Philipsen, 1982, p. 10) are important aspects of qualitative work, but "according to the … epistemology that currently holds sway, [personal reflections] are subjective and, therefore, biased" (Kovarsky, 2008, p. 51). The idea of subjectivity is *crucial* in qualitative work. As Purdy (2011) wrote:

> The grounding of a study is not complete for many qualitative researchers unless the researcher has described his or her own life situation, prejudices, and potential biases. The context of the researcher's situation is not neutral or irrelevant but integral to the project of research (p. 134).

According to Miller (2006), *all* research is inherently imbued with "the values of the researcher, the research community, and society" (p. 58).

Third, qualitative inquiry is nonlinear. Linearity in "the design and conduct of research," according to Philipsen (1977), "refers to the specification, in advance of data collection, of both a purpose for research and of the steps to be taken to serve the purpose" (p. 42). In experimental research, an investigator usually predicts what will happen *before* the data are collected, but qualitative work is nonlinear and involves "a blurring of data collection, analysis, and theorizing" (Miller, 2006, p. 63).

Iterative approaches allow researchers to examine data, refine collection techniques, revise research questions, and theorize cause–effect relationships while still in the field. For example, while conducting observational research for my dissertation, I discovered midway through data collection that certain interaction patterns among healthcare providers that I associated with providers' cognitive dissonance were, in fact, indicative of a phenomena called *stucturational divergence* (see Nicotera & Mahon, 2013). Subsequently, I refined the kinds of behaviors I attended to when observing and listening to providers' interactions, and I tweaked the questions I asked informants during interviews. Because I shifted my focus while simultaneously gathering and analyzing data, I uncovered important theoretical linkages that, for the first time, established empirically structurational divergence nexus-cycle escalation (Overton, 2015).

Fourth, qualitative research takes place in situ, meaning that investigators commonly observe "what the subject does when left to his or her own devices" (Philipsen, 1982, p. 7). Philipsen (1982) maintained "the heavy reliance of artificial contexts," such as laboratories, "limits the degree to which findings can be generalized to non-trivial contexts" (p. 4), thereby making many laboratory-based studies less valid ecologically than studies that take place "in the settings and at the times which are the usual contexts for the subject's actions" (p. 6). Carbaugh (1989) agreed, writing, "To listen fully to culture, one must be positioned there, in the meaning-filled world listening for the meanings created within it, from the standpoint of those who create them, rather than standing elsewhere" (p. 279).

In sum, qualitative research is word-driven, subjective, nonlinear, and conducted in situ. Nonlinear approaches allow researchers to refine data collection techniques and theorize relationships while still in the field, which, for many researchers, is an advantage over linear approaches. Additionally, field-based research generates data and findings that are oftentimes more ecologically valid when compared with laboratory-based work. Next, I address issues believed by many researchers to reduce rigor in qualitative studies, and I describe steps that can be taken to enhance rigor.

Rigor and Qualitative Methods

As noted in this chapter, all research *is* subjective to a degree, but approaches like ethnography (despite being a valuable methodology for producing rich data and thick description) have been thought to lack not only objectivity but also methodological rigor and utility. By addressing these misconceptions, I hope to demonstrate the utility and rigor of qualitative approaches for listening researchers.

With respect to subjectivity, Madison (2012) stressed that contextualizing one's views makes the inherent subjective voice in research "accessible, transparent, and vulnerable to judgment" (p. 10), thereby making the research appear more rigorous when compared with work—either qualitative *or* quantitative—that omits or obscures investigators' perspectives. As a researcher, you should state expressly your aims, thus making your own subjectivity apparent—in so doing, you also make your research more methodologically sound. For instance, in my dissertation about ways that EMRs used in emergency departments impact healthcare providers' interaction patterns, their perceptions of communicative acts, and listening,[2] I acknowledged my subjectivity by (a) stating my opinion of EMRs (i.e., I find EMRs are rarely practical and regularly problematic), and (b) admitting that my husband, an emergency medicine physician, complained routinely about EMRs. I acknowledged also that even though my aim was not to write a critical ethnography, my dissertation nonetheless took on a critical voice following 18 months of fieldwork and an additional 6 months of coding, analyzing, and reflecting on data. By contextualizing my subjectivity and critical voice, I avoided deception and achieved a measure of transparency (see Purdy, 2010).

The perceived lack of "objectivity," as Perlmutter (2015) observed, is, in fact, evident in every method. For instance, researchers who have employed methods such as experiments, surveys, and content analysis have also been, as Perlmutter (2015) noted, "concocting data sets, suppressing negative findings, and embellishing positive results" (para. 11); hence, "the idea that [qualitative work] is especially flawed is absurd" (para. 11). Nonetheless, the notion that qualitative work is somehow more subjective and less rigorous has persisted, which may explain, in part, why "there has been a consistent bias against qualitative research in some areas of communication study [and] unfortunately listening is one of those areas" (Purdy, 2010, p. 43).

Purdy (2010, pp. 41–42) offered five suggestions for "maintaining rigor and method," suggesting that qualitative researchers:

- acknowledge their assumptions if "research is to be well-rounded";
- "acquire first-hand experience of the culture being researched";
- develop thorough literature reviews to prompt research questions appropriate for the methodology;
- work with other researchers because "many heads can make better sense of the cultural meaning of listening, of how to structure research and how to make sense of research results"; and
- "use the feedback from others, as well as [their] own intuition, to be as 'rigorous' (as in critical and creative) as possible."

Despite its utility, qualitative research has at times, when compared with experimental research, been thought "theoretically aimless and methodologically shifty" (Philipsen, 1977, p. 42), because it is largely nonlinear, and qualitative researchers seldom predict what will happen during their studies. Many critics have mistakenly assumed that this process is without purpose or plan (Philipsen, 1977, p. 43). Philipsen (1977, p. 48) proposed a solution, recommending investigators "specify a phenomenon of interest, link that phenomenon conceptually to the process of communication, and specify a framework for describing that phenomenon and its particularity in any given social field" *before* data collection. Hence, it is important that you achieve rigor in your qualitative research—however nonlinear your study's design—by stating clearly your aims, theoretical perspective(s), methodology, and methods. Likewise, because your study's design may evolve (e.g., your research aims may narrow or change, or you may employ additional

2 Although I did not set out expressly to study listening, I discovered through a nonlinear, iterative process that providers' definitions of and experiences with listening were impacted by EMRs in meaningful ways.

data collection techniques), you should note major changes in your research report. For example, at the start of my dissertation research, I specified that I was interested in EMR-induced changes to emergency medicine providers' communication patterns; that structuration and adaptive structuration theories would frame my analysis; that ethnography was my guiding methodology; and that I would collect data through participant observation, listening, interview, document analysis, and questionnaires. Later, I incorporated structurational divergence theory into my analysis and data collection, so I explained the rationale behind that decision in my written report.

The notion that qualitative work is less rigorous when compared with experimental research may have been perpetuated, in part, because some qualitative articles, even seminal studies like Carbaugh's (1999) piece on listening practices among Blackfeet, lack a "Methods" section. Although it is apparent that Carbaugh (1999) observed, spoke with, and interviewed his informants, readers may wonder how long he spent in the field (i.e., readers may assign varying degrees of assuredness to conclusions drawn after 2 weeks, 2 months, or 2 years of fieldwork); if he took fieldnotes and/or audio-recorded conversations; how he coded data and subsequently looked for themes; and if he took steps to ensure descriptive validity. On the matter of how much an investigator *should* "say about method," Wolcott (2010) wrote:

> The reader does need to know specifically how you gathered the information that you used in order to assess the extent that what you have to say can be relied on. Did you depend mostly on one or two informants, for example, or did your information come from a broadly distributed group? We must not lose track of the fact that our readers want to evaluate our reporting, just as we have tried to assess the group or individuals about whom we write. (p. 35)

It bears mentioning that Carbaugh's methodological choices, although not appearing in his 1999 report, have been written about elsewhere (e.g., Carbaugh, 2005). In an email interview, Dr. Carbaugh explained to me that he typically cites other pieces "where the theory and methodology has been explicated more fully, rather than summarize as much in an article" (personal communication, November 25, 2015). He acknowledged that "it is crucial that we, as qualitative researchers, use the most exacting procedures possible and make those available to our readers," but emphasized the importance also "of crafting an article that is compelling to read for the widest audience possible" (personal communication, November 26, 2015). He elaborated:

> The objectives can pull in different directions. If I have written in detail about methodology before, relevant to a study, I find it not just easiest but best to cite that, then move on—why else publish the methodology? Those interested readers can, then, find the detailed statement of method and read or study it. There is often another reason, for some studies, where the methodology is so detailed and involved, especially for example in my Blackfeet studies which have gone on for decades, that it is nearly impossible to write the methods in a couple paragraphs. And when I have tried, it did the methodology an injustice! So, we do the best we can, given the circumstances in which we find ourselves. I find it best to allow different strategies for different occasions, or different genres, or different audiences. (Personal communication, November 26, 2015)

Carbaugh has written extensively about qualitative methods generally and ethnographic methods specifically. Of particular interest to students are his descriptions of the three phases of ethnographic inquiry (Carbaugh, 2006). The first phase, *prefieldwork*, includes reading about ethnographic theory and methods. The second, *fieldwork*, involves "distinct phases such as generating data (through interviews, observations, document collection, surveys, and so on), recording data (through transcribing, audio and video recordings, and other field notation systems), analyzing data (through various quantitative and qualitative procedures), and continued reading" (Carbaugh, 2006, p. 158). The third, *postfieldwork*, continues in-field analysis, "sometimes leading back to the field … in order to generate better perspective and new data" (Carbaugh, 2006, p. 158).

I believe that Carbaugh's guidelines are helpful for researchers considering qualitative processes, but I maintain that students and emerging scholars *should* describe their methodologies. No emerging scholar, after all, has a body of work, as does Carbaugh, to refer readers to for additional information. Molina-Markham's (2014) article on listening practices among Quakers is an exemplar. Based on a chapter from her dissertation (directed by Carbaugh), she succinctly described in one paragraph that she conducted observational and interview research over a 2-year period and noted that the primary data analyzed included "audio recordings of two [Quaker] meetings … that occurred during two consecutive months" (Molina-Markham, 2014, p. 160). She also indicated that she reviewed "detailed fieldnotes on 11 other meetings for business and 58 meetings for worship … as well as recorded interviews" (Molina-Markham, 2014, p. 160). By briefly describing data collection methods, Molina-Markham enhanced the perceived methodological rigor of her work. Describing methods allows other scholars the opportunity to assess, critique, and even replicate aspects of a study's design.

The utility of qualitative methods also has been questioned because its findings are rarely replicated (Fiske, 1991). Fiske (1991) pointed to an "epistemological crisis in ethnography … concerning the status of the knowledge it produces" (p. 330) and cited the lack of reproducibility of ethnographers' findings by other scientists, which, as Fiske suggested, made ethnography "a discursive science, not an empirical one" (p. 330). It is important to note, however, that replicability of findings is no more guaranteed in ethnography than it is in any other methodology (Perlmutter, 2015). Although "no two ethnographers can study the same community" (Perlmutter, 2015, para. 14), typically, that is the case for other methodologists as well (e.g., experimental researchers).[3] As Frey, Botan, and Kreps (2000) noted, "there is, of course, no way to ever replicate someone's study exactly, since every investigation involves a different researcher and different research participants" (p. 135).

To review, qualitative research is characterized by the kinds of data collected (i.e., word-based); the data collection techniques used, such as in-depth interviews or observation; data analytic procedures, which include the subjective coding and

3 In an interesting study, researchers attempted to reproduce results from 100 experimental and correlation studies. According to the study's authors, "The mean effect size (*r*) of the replication effects ($M_r = 0.197$, SD = 0.257) was half the magnitude of the mean effect size of the original effects ($M_r = 0.403$, SD = 0.188), representing a substantial decline" (Open Science Collaboration, 2015, para. 4). Results indicated that replications produced significantly smaller results when compared with original studies, thereby suggesting that reproducibility is, in fact, no more guaranteed for experimental researchers than it is for qualitative researchers.

interpretation of data; linearity of research designs; and location of study (i.e., in the field). Issues believed to impact rigor in qualitative research include investigators' subjectivity, nonlinear designs, and reproducibility of findings. Specifying research aims, areas of inquiry, guiding theories and methodologies, data collection techniques, and potential sources of investigators' bias—even in nonlinear designs—can assuage criticisms regarding rigor without sacrificing fluidity and creativity. Lastly, replication of findings is no more guaranteed in qualitative research than it is in quantitative research.

This section has not presented an exhaustive list of differences between qualitative and quantitative approaches, but it has identified important areas of divergence you should consider when deciding which approach is most appropriate for your study, or if, in fact, your study would be best served by using a combination of qualitative and quantitative methods. For example, before conducting research on EMRs and communication in emergency departments, I considered my research aims, the kinds of questions I was interested in asking and answering, and whether to take a qualitative, quantitative, or mixed-method approach. Although a lot of EMR research explored quantifiable changes in emergency departments following EMR adoptions (e.g., increases in medication errors, imaging studies, and laboratory tests), I wanted to understand the kinds of changes that EMRs potentially introduced to providers' communication patterns. I could have administered questionnaires to providers, and I could have conducted interviews with them; however, in the early stages of planning a dissertation project, I did not know what to ask *specifically* about EMRs' effects on providers' communication practices. I decided to observe and listen to the goings-on in an emergency room for a few weeks so that I could collect preliminary data, which, I hoped, would inform how I set about formulating research questions. The resulting qualitatively oriented pilot study produced rich descriptions of providers' experiences that shaped both the interview questions I later posed to providers and the majority of items on a questionnaire I administered to them (e.g., items designed to assess respondents' views on their EMR-specific interactions incorporated phrases such as *order dumping* and *gaming the system* that, as I heard, providers used frequently). Meldrum (2011), who also studies listening in healthcare settings, noted:

> Qualitative methods are often used inductively for exploratory studies to set the groundwork for further theory building and description. In addition to providing a rich description of complex relationships, this method offers a conceptual framework through which to explore the meaning of "listening" in this particular context. (p. 149)

You may find that qualitative approaches will likewise help you narrow your focus or contribute to item and/or scale development (see, e.g., Nicotera & Clinkscales, 2010; Nicotera, Mahon, & Zhao, 2010).

As I contemplated my dissertation project, I took into account several practical considerations, which you should also contemplate when deciding if your study should be qualitatively or quantitatively driven. First, I assessed my own strengths and weaknesses as a researcher—I preferred analyzing words to numbers. Second, I focused on what was doable. Experimental designs or variables were beyond the scope of my study (i.e., without the ability to manipulate variables and a control group for comparison, an experimental design was, effectively, impossible). I could, however, gain entry to a site, recruit and interview participants, and dedicate long hours over many months to observing, listening to, and noting changes in providers' communication patterns.[4]

As you contemplate your own study, ask yourself if you have access to a research site and the cooperation of gatekeepers who will help facilitate your research and connect you with potential subjects (e.g., my husband worked in an emergency room and helped arranged my access). Moreover, do you have the time and resources necessary to invest in participant observation and/or in-depth interviewing, which can go on for weeks or months? If you answered "no" to these questions, then quantitative techniques (e.g., questionnaires) may prove more doable given such techniques are, generally, faster to administer and analyze when compared with observational and/or interview-based approaches. If, however, you wish to utilize a qualitative approach in your study, then you should consider the various methodologies and methods available to you. A brief overview of some of the most popular in listening research is presented next and is followed by a discussion of ethnographic methods.

Qualitative Methodologies

Purdy's (2010) typology of methodologies and methods, although not exhaustive, contains many approaches that listening researchers may find useful. From among the methodologies on Purdy's (2010) list, listening researchers have advanced consistently three methodological approaches: phenomenology, hermeneutics, and ethnography (see, e.g., Carbaugh, 1989, 1991, 2006; Hymes, 1962; Lipari, 2010, 2012; Philipsen, 1977, 1982; Philipsen & Carbaugh, 1986; Purdy, 2000, 2010, 2011). Of these, ethnography has appeared more regularly in the listening literature, so it is given more attention in this chapter. Drawn from the field of anthropology, it is broadly defined as a "written representation of a culture" (Van Maanen, 2011, p. 1). Many ethnographers have focused also on communication practices in general (e.g., Philipsen, 1975) and listening in particular. Examples of listening-oriented ethnographies include Podkalicka's (2009) work on perceptions of listening and being listened to within the context of a radio production workshop for disadvantaged youth; Molina-Markham's (2014) study on silence, listening, and decision making among Quakers; and Carbaugh's (1999) piece on "'listening' as a cultural form of communication" (p. 250) among Blackfeet. Carbaugh's ethnographic approach made the Blackfeet study an exemplar in the listening literature—inasmuch as Carbaugh explored listening as a situated communication activity using qualitative methods, the article was, as Purdy (2000) described it, "one example of ways of studying listening that may open new understanding for research" (p. 47).

Ethnography of Communication

Carbaugh's (1999) work, like Philipsen's (1975), Molina-Markham's (2014), and more than 200 additional works (see, e.g., Philipsen & Carbaugh, 1986; Carbaugh, 2005), falls under the rubric *ethnography of communication* (EOC), which is a methodological and theoretical movement launched by Dell Hymes in 1962. EOC advances "a theory of linguistic communication, which is grounded in the comparative analysis of many

4 I received a dissertation-year fellowship from the Louisiana State University A&M graduate school that released me from teaching responsibilities and freed up ample time for this type of project.

communities and their distinctive ways of speaking" (Philipsen & Carbaugh, 1986). Hymes (1962) advocated EOC as a methodological approach for studying not just culture, which was the impetus driving anthropologic ethnographies, but also communication practices *specifically*. EOC encouraged researchers who were embedded in speech communities to observe and analyze naturally occurring communication events and/or situations. The key components of an EOC study are captured in the acronym SPEAKING (Hymes, 1962):

- **S**etting or scene
- **P**articipants
- **E**nds (i.e., goals)
- **A**ct
- **K**ey (i.e., emotional tone)
- **I**nstrument (i.e., channel used for message delivery, such as mediated or face-to-face)
- **N**orms
- **G**enre

EOC's primary goals include "descriptive representation and theoretical rigor" (Carbaugh, 1989, p. 263). EOC researchers tend to emphasize description over critical evaluation by prioritizing "the participant or native view" (Carbaugh, 1991, p. 336).

The absence of a critical voice in EOC, according to some scholars, is problematic. Critical theorists and feminist scholars have, according to Frey *et al.* (2000), "challenged the 'traditional' view of ethnography as an attempt to be a relatively 'objective' report about another culture" (p. 260). Critical ethnography, in comparison, "begins with an ethical responsibility to address processes of unfairness or injustice *within* a particular domain" (Madison, 2012, p. 5). For Fiske (1991) and other critical theorists, "the point of producing knowledge is not just to understand our social conditions but to work to improve them" (p. 234). Hence, Fiske (1991) argued that EOC's descriptive methods fundamentally, and Carbaugh's work specifically, repressed knowledge that may have shown the various speech communities that were studied under the EOC rubric were "deeply divided by gender, race, class, and other differences" (p. 335). Carbaugh (1991) countered that EOC is simply "one way to integrate cultural interpretation into communication inquiry, toward the goals of understanding communication practices *sui generis*, on their own terms, and as they are variously lived in various places" (p. 341). Elsewhere, Carbaugh (1989) stressed, "One does not necessarily have to evaluate a system in order to describe and theorize about it … just because a critical voice is nonessential, that does not mean it is necessarily excluded from ethnography" (p. 264).

"Doing" Ethnography

Although "observation is the central data collection method in ethnography" (Cooper & Endacott, 2007, p. 816), ethnographers regularly utilize other data collection methods. Tracy and Geist-Martin (2014) compared qualitative researchers, like ethnographers, to *bricoleurs*, who "piece together data they have collected from a broad variety of data sources, which may include, for example, participant observation fieldnotes, interview transcripts, organizational documents, and websites" (p. 247). In this section, I discuss participant observation, participant listening, and interviews—techniques that are used frequently in ethnography and that, I believe, are valuable also in listening research.

I also describe the process of fieldwork, gaining access to research sites, sampling procedures, strategies for overcoming validity threats, and approaches for coding and analyzing data (for more complete descriptions of qualitative research methods, see, e.g., Goodall, 2000; Lindlof & Taylor, 2011).

Fieldwork

Fieldwork is the "investment of a researcher over a lengthy period of time (typically unspecified) and consists mostly of ongoing interaction with the human targets of study on their home ground" (Van Maanen, 2011, p. 2). It is, in many respects, a personal endeavor, and each fieldworker must adapt to the demands of the scene and the participants under investigation. Fieldwork requires effort, astute observation, and active listening as well as patience, steadfastness, decorum, and oftentimes diplomacy. For instance, while conducting fieldwork in an emergency room, I had to balance my interests as an investigator with hospital administrators' desire that I present the emergency department in a positive light. Furthermore, I had to be sensitive to patients' privacy concerns. I also needed confidence to enter a site in which—despite wearing surgical scrubs and posing as an insider—I was clearly an outsider.

For many fieldworkers, the "outsider" status can be difficult to overcome. As Van Maanen (2011) noted, the fieldworker arrives "knowing few people, if any" (p. 2). Successful fieldworkers, according to Van Maanen (2011), must:

> learn to move among strangers while holding themselves in readiness for episodes of embarrassment, affection, misfortune, partial or vague revelation, deceit, confusion, isolation, warmth, adventure, fear, concealment, pleasure, surprise, insult, and always possible deportation. (p. 2)

Although Van Maanen's (2011) description of fieldwork may seem daunting, fieldwork is, in fact, a deeply rewarding experience. After nearly 2 years of fieldwork, I left the scene with rich, nuanced data that captured the goings-on in an emergency room with greater detail than the data I accumulated from questionnaires alone. I also left the scene having made friends with whom I am collaborating in ongoing studies.

Although fieldwork is difficult to teach (Goodall, 2000), mainly because each instance of fieldwork is unique, all fieldwork incorporates certain steps that you should work through, including "what to study, gaining access, what observer role to assume, how long the observational period should last, what to look for while making observations, and how to record observations" (Frey *et al.*, 2000, p. 265).

What to Study

Before gaining access to a site and collecting observational data, researchers must decide on a communication phenomenon and population to investigate. Goodall (2000) advised students to "read widely in your field" with an eye toward identifying gaps in scholarly knowledge, ways to address these gaps, and, ideally, how you might offer solutions to problems (p. 52). For instance, I read extensively on the subject of EMRs in emergency departments and found two gaps, which I addressed in my dissertation. First, like much listening research, most EMR research is carried out using surveys. Although survey research can be useful, it is generally cross-sectional in nature, thus limiting an ability to make claims about change. Likewise, designing questionnaires can often be limiting due to concerns about participant fatigue and clarity of items (see Chapter 2). Second, emergency department–based EMR research is set, almost exclusively, in large academic medical centers. By using an ethnographic approach to study EMR-induced changes in a community hospital's emergency department, I contributed to the scholarly literature in new and meaningful ways. Additionally, I proposed solutions for helping providers overcome EMR-induced communication problems. Similarly, listening researchers can study topics that have been previously examined but investigate them using novel methodological approaches and in new settings.

Gaining Access

If you decide to study listening phenomena and/or people in a public setting, gaining access is rarely problematic; however, if the setting is private, you "must negotiate with 'gatekeepers' who have the power to grant or refuse access" (Frey *et al.*, 2000, p. 266). If you want to conduct your study in a private or restricted-access setting (e.g., someone's home or an emergency room), your first task is to identify gatekeepers and seek their permission. Lindlof and Taylor (2011) recommended submitting a formal research proposal and institutional review board (IRB) approval to gatekeeper(s) for consideration when making your request, especially if you want to conduct research in an organizational setting. Having members who can vouch for you can facilitate access (e.g., my husband and his colleagues in the emergency department). Once on site, you will need a sponsor who "goes around and personally introduces you, vouches for your study, and helps you gain access" to persons of interest (Lindlof & Taylor, 2011, p. 101).

The importance of a good sponsor cannot be overstated. In my case, my sponsor was the emergency department nursing manager, who made certain my presence and behavior in the emergency room complied with hospital policy, announced my visits beforehand to the emergency room staff, and reminded them of my research goals. She helped arrange face-to-face interviews with several physicians, distributed and collected questionnaires on my behalf, and reprimanded administrators who did not respond to my requests for interviews in a timely fashion. My advice is that you seek out a sponsor with whom you can easily communicate and collaborate; otherwise, you may find gaining access to spaces and persons of interest more challenging.

Sampling Strategies

Although you may have decided generally whom you wish to study (e.g., healthcare providers, in my case), once you are on the scene, you must decide specifically whom your sample will include. Most ethnographers engage in purposive sampling, as opposed to random sampling, because they want to "make informed judgments about what to observe or whom to interview" (Lindlof & Taylor, 2011, p. 110). Because one of the goals of ethnographic inquiry is to get as deep an understanding of the scene as possible, it is important to select a sample as representative of the site's population as possible. During my fieldwork, for instance, I observed and/or interviewed technicians, nurses, midlevel providers, physicians, administrators, and EMR trainers. Although the sample ($N = 37$) may appear small when compared with samples recruited for large-scale survey studies, it reflected sufficiently the types of persons working in the emergency department. The sampling process ended "when new data no longer added much of significance to the concepts … developed" (Lindlof & Taylor, 2011, p. 117). In essence, when newly recruited participants to my study no longer contributed novel insights but, instead, simply restated and/or confirmed what existing participants had previously revealed, then I knew I had a large enough sample. Having then achieved *saturation* (i.e., no new information was seen or heard), I stopped enlisting new participants.

You may approach sampling in a similar fashion, or you may consider other strategies. You may select participants based on certain criteria, such as occupation, gender, or age. For instance, I solicited a *criterion sample* while conducting my pilot study: Individuals who self-identified as emergency medicine providers were invited to complete a questionnaire, whereas other emergency room workers who did not participate in direct patient care (i.e., secretaries and janitorial staff) were excluded. Criterion sampling is similar to *maximum variation sampling*, which "taps into a range of qualities, attributes, situations, or incidents of the phenomenon under study" (Lindlof & Taylor, 2011, p. 113). Lindlof and Taylor (2011) described additional methods: *snowball sampling*, which includes recruiting participants who then recruit additional participants from among their acquaintances; *typical-case sampling*, which involves selecting sites, people, or events that are typical of the phenomena being studied; and *atypical-case sampling*, which essentially focuses on the rare or exotic. It is worthwhile to remember that most qualitative researchers "develop a sampling plan along the way, rather than strictly in advance," and oftentimes "justify their research goals based upon the site or sample" (Tracy & Geist-Martin, 2014, p. 247).

Participant Observation

Ethnographers tend to favor observation over other data collection techniques because observation "overcomes the discrepancy between what people say they do and what they actually do" (Cooper & Endacott, 2007, p. 817; see also Chapter 6). Ethnographic observation is achieved through a practice called *participant observation*, which is the "craft of observing and recording events in social settings" and involves "being in the presence of others on an ongoing basis" (Lindlof & Taylor, 2011, p. 135).

There are several degrees of participant observation you should consider. Your level of involvement in the setting and with the persons you study will determine your observer role. There are two prevailing typologies that describe types of participant observation. The first typology, developed by Schwartz and Schwartz (1955), defines observation as either passive or active. *Passive observation* is, more or less, covert, meaning that researchers operate as "anonymously and unobtrusively" as possible (Lindlof & Taylor, 2011, p. 144). *Active observation* is overt, and researchers interact with participants regularly and openly.

The second typology of participant observation was developed by Gold (1958) and includes four roles researchers play, depending on "how much researchers participate in the activities being observed" (Frey *et al.*, 2000, p. 267). The first role is *complete participant*. Complete participants immerse themselves fully in the setting. Their true identities—as embedded researchers—are unknown to participants. Researchers pretend to be members of the milieu (Frey *et al.*, 2000), which can oftentimes involve deception. As Lindlof and Taylor (2011, pp. 145–146) noted, "contemporary fieldworkers do not embrace the role of complete participant" (p. 145) because the covert practice can create "ethical 'problems,'" such as the lack of informed consent.

The second role, *participant-observer*, characterizes a researcher who participates actively in a scene "where people know they are being studied" (Frey *et al.*, 2000, p. 267).

Participant-observers "study a scene from the vantage point of one or more positions within its membership" (Lindlof & Taylor, 2011, p. 146), which can give investigators a fairly complete picture of the scene. For example, in his dissertation about communication among members of an Evangelical church's leadership team, Hartwig (2010) participated as a church member, leadership team member, and organizational consultant—his engagement with the church in several capacities allowed him to describe the goings-on in considerable detail.

Third, the *observer-participant* "is primarily invested in observing" (Lindlof & Taylor, 2011, p. 147) and "participates to a limited extent" (Frey *et al.*, 2000, p. 268). I acted as an observer-participant while collecting data for my dissertation—I spent time with nurses and physicians in emergency rooms, but I did not participate in activities associated with patient care. I watched and listened intently to providers engaged in delivering care to their patients, but I relied on interviews with providers to round out the observational data that I collected. Although very detailed, the observations I recorded in fieldnotes did not capture the "depth of vivid first-hand experience" (Frey *et al.*, 2000, p. 268) associated with the complete participant and participant-observer roles.

The last role is the *complete observer*. Complete observers study "social actors without being present or known to them" (Lindlof & Taylor, 2011, p. 148). According to Frey *et al.* (2000), the complete observer is "concerned with faithfully gathering data about people's behavior without influencing them in any way" (p. 269). Researchers who adopt this role may collect data by eavesdropping or videotaping in public spaces, like "crowd scenes and public websites" (Lindlof & Taylor, 2011, p. 148; see also Chapter 6). Although eavesdropping has yielded interesting insights into dyadic communication (see, e.g., Deakins, Osterinck, & Hoey, 1987), Lindlof and Taylor (2011) noted that this role does not allow researchers "meaningful contact with participants [and] denies them [participants] the opportunity to influence our evolving interpretations" (p. 148).

What to Attend to While in the Field

Lindlof and Taylor (2011) recommended that investigators begin the task of observing by noticing "as many persons, objects, and events that are 'happening' in a site as possible" (p. 151). The most important question is "What is going on?" Lindlof and Taylor (2011) proposed six additional questions that you should consider as you work at unraveling what is going on.

First, "Who are the actors?" If you are conducting research in an organization, it is important to discern participants' roles and their corresponding duties. If your research takes place in a casual setting, determine how participants see themselves in relation to other players on the scene. For example, if you are studying listening behavior in families, you may ask how family members are related to one another. Are the dyads you observe composed of siblings, cousins, or parents and children? You may ask about ways that actors prioritize listening behaviors differently in sibling dyads versus in parent–child dyads. Irrespective of where you situate your study, you should be attentive to ways that participants perform their roles, relate to other actors, and navigate interactions.

Next, you should ask, "How is the scene set up?" Notice how participants use props and if their actions suggest how they want to be regarded by others. In my pilot study, for instance, I observed that nurses used props (e.g., soda cans, perfume bottles, and snacks) as place markers to carve out individual work places along a communal counter in the emergency room, but hierarchically superior providers, like physicians and nurse practitioners, often ignored the place markers. My observation suggested something important about ways that providers interacted with one another and occupied space.

According to Frey *et al.* (2000), the third question—"How do initial interactions occur?"—forces you to attend to early encounters and ways that "new members are socialized into a culture" (p. 270). This includes observing who initiates conversation; the tone, volume, and speed of vocal exchanges; the number of interactants; the length of interaction; the topics discussed; and what the interactions accomplish. Such observations suggest ways that power is distributed among players on the scene. For instance, I observed that physicians initiated discussions with fellow providers more often when compared with nurses—even when nurses had questions regarding physicians' orders, frequently nurses waited for physicians to initiate talk before questions were posed. This observation revealed that many physician–nurse interactions were hierarchically imbalanced.

Fourth, you should ask, "How do actors claim attention?" To answer this question, Lindlof and Taylor (2011) suggested that you turn your attention to "the topics that stimulate them [players] to argue and discuss, while others are ignored and forgotten" (p. 153). Moreover, you should ask how players address the issues that demand their attention (Lindlof & Taylor, 2011). I noted that some nurses refused to use the EMR under certain conditions (e.g., triaging new patients), which stimulated talk among providers about ways that EMRs impeded their productivity and workflow. This, in turn, led to other observations of providers duplicating work (i.e., handwriting triage notes and later typing those same notes into the EMR) and seeking "workarounds" to combat slow-performing EMRs.

Next, you should ask, "Where and when do actors interact?" You should record who interacts with whom, where interactions take place, at what times, and under what conditions (Lindlof & Taylor, 2011). These observations can reveal much about relationships and interaction patterns (see, e.g., Eisenberg, Baglia, & Pynes, 2006; Nugus *et al.*, 2011).

Finally, you should ask, "Which events are significant?" According to Lindlof and Taylor (2011), "recognizing significant events requires that you decide whether and how they *count as* examples of a relevant concept in your study" (p. 155). I decided that negative communication spirals between nurses and physicians, especially negative spirals that were initiated by physicians' requests for nurses to perform data-entry tasks, were significant. I attended to these significant episodes, which led me to link conceptually structurational divergence nexus-cycle escalation with physicians' agency restoration attempts (i.e., physicians' efforts at reclaiming power amid a forced EMR adoption by shifting their data-entry tasks onto nurses).

Another way you may structure observations, according to Goodall (2000), is to focus on "'performances' of everyday life" (p. 166), which manifest as routines, rituals, and rites of passage. *Routines* are what people do every day; *rituals* are acts people perform regularly, which are symbolic or meaningful; and *rites of passage* (e.g., exchanging marriage vows) change people in fundamental ways (Goodall, 2000). Ethnographers also examine isolated practices: surprise and sense-making episodes, risk-taking episodes, face-saving episodes, and crises (Goodall, 2000). Therefore, it is important that you consider the mundane *and* the exceptional, as both potentially can reveal something about your site and your sample.

Additionally, Goodall (2000) emphasized the importance of listening and attending to the *types* of conversations that social actors have. In fact, "much of what is actually recorded as data [during participant observation] is sonic rather than visual" (Forsey, 2010, p. 562), so listening to conversations is an important data collection technique. Moreover, listening as a communicative practice offers an entry point for researchers—like critical ethnographers—to transform oppressive conditions and facilitate social change (see, e.g., Dutta, 2014).

Participant Listening

There is a bias toward vision that obscures how investigators conduct and write about fieldwork (Lindlof & Taylor, 2011). Frequent use of terms like "participation *observation*" to describe fieldwork and/or ethnographic inquiry underscores this bias. Consider, as Forsey (2010) noted, that the terms "ethnography and participant observation [are] whimsically interchanged" (p. 559). Conflating the two—ethnography with observation—implies "ethnography can only be realized by, or with, participant observation" (Forsey, 2010, p. 561) and, thus, precludes data derived from other senses from equal consideration. In actuality, much of what is reported in ethnography as "observation" comes from "people conversing" (Forsey, 2010, p. 563). In reviewing my own dissertation, I concede, like Forsey (2010), that what I recorded and reported was derived overwhelmingly from things I heard (i.e., things told to me and things heard or overheard in conversation). Lindlof and Taylor (2011) cautioned that researchers "should carefully monitor how sensory bias shapes perception and interpretation of events" and, moreover, that researchers "should open their senses to experience and record the aesthetic and nondiscursive textures of their chosen sites" (pp. 138–139). Although ethnography, according to Forsey (2010), is as much about listening as seeing, we should avoid pitting one method/sense against the other because doing so introduces an unnecessary and problematic dichotomy. Our senses *are* integrated—what we hear (and listen to) is a function not only of sounds but also of sights, touch, tastes, and smells.

With this in mind, let us return to Goodall's (2000) assertion that "verbal exchanges are the *organizing focus* of everyday experiences" (p. 98). Goodall (2000) created a typology of verbal exchanges for which researchers should listen, ranging from phatic communication to dialogue. The first, *phatic communication*, refers to social niceties that people exchange, such as "Hello" and "How are you?" Phatic exchanges, although seemingly simplistic, can reveal patterns of hierarchy, status, race, class, gender differences, and rules governing turn-taking sequences (Goodall, 2000). For instance, I noticed that nurses who were female interacted less with physicians of both sexes when compared with nurses who were male. Differential patterning of interaction based on the sex composition of the dyad can have far-reaching implications for nurses and suggests that gender inequality persists in some emergency room environments.

Second is *ordinary conversation*, which is made up of "patterns of questions and responses that provide interactants with data about personal, relational, and information issues and concerns" (Goodall, 2000, p. 103). Listening and attending to the kinds of information that people exchange reveal a lot about hierarchies and power, especially in organizational settings.

The third type of verbal interaction, *skilled conversation*, reflects more nuanced information exchange. Skilled conversation may, at times, incorporate conflicting views. How conflicts are resolved "are important emblems of how a culture structures rules (and violations) for managing or resolving disputes" (Goodall, 2000, p. 104). For instance, Eisenberg *et al.* (2005) found that "professional fault lines" (p. 392) in emergency departments separated physicians from nurses, which hindered communication among them. As Eisenberg *et al.* (2005) noted, "Many nurses simply remained silent when they disagreed [with physicians], following physicians' orders and only later expressing their objections to other nurses" (p. 402). Likewise, I found that nurses rarely addressed grievances with physicians directly, which supported my reading of the emergency room as a hierarchically imbalanced site of struggle.

Fourth and fifth on Goodall's (2000) typology are personal narratives and dialogue. *Personal narratives* are marked by mutual self-disclosure in which participants use disclosure "to situate, coordinate, detail, and explain or retell pivotal events in a personal or organizational life" (Goodall, 2000, p. 104). While conducting fieldwork, I found that personal narratives centered regularly on episodes that took place before EMRs were installed in the emergency room. Mutual disclosures were positioned frequently as *gripe sessions*. As a nurse described them, "Gripe sessions are when we complain about EMRs and remember the good old days *before* there ever were EMRs. Those talks bring us closer together." Although personal narratives can foster feelings of connection for interactants, dialogue can be transformative. In *dialogue*, according to Goodall (2000), "talk moves from exchanges of information and the coordination of new understanding to a higher level" (p. 104).

Goodall's (2000) emphasis on conversations reminds researchers to broaden their focus from what is visual to include what is heard. Like conversations, interviews are an extension of participant listening. Interviews produce rich data that, like conversations, require researchers to *listen*. Types of interviews, interview questions, and logistical concerns are described next.

Interviews

Unlike in survey research, where highly structured interviews unfold in similar or identical ways, in-depth interviews are more often exploratory and "proceed inductively, using an unstructured format of open questions" (Frey *et al.*, 2000, p. 273). According to Lindlof and Taylor (2011), in-depth interviews are useful for understanding social actors' experiences and perspectives; gathering information about things, people, or processes that cannot be directly observed; inquiring about past events; verifying information obtained from other sources; clarifying processes or procedures; and eliciting language forms used by social actors in situ.

Lindlof and Taylor (2011, pp. 176–180) identified five types of in-depth interviews: ethnographic, informant, respondent, narrative, and focus group.

- *Ethnographic interviews* are spontaneous interviews that occur "in a cultural scene, while the investigator is busy hanging out with the people being studied," and they are useful for soliciting participants' immediate reactions to events or conversations (Lindlof & Taylor, 2011, p. 179).
- *Informant interviews* solicit insiders' perspectives on group practices; interviewees are "*informants* because they inform the researcher about the scene—the scene's history, customs, and rituals; the local 'lingo'; the identities and actions of key players; and so forth" (Lindlof & Taylor, 2011, p. 179).
- *Respondent interviews* elicit open-ended responses from cultural members and are used to clarify interviewees' meanings; elucidate their opinions, beliefs, and attitudes; identify sources of influence that are tied to certain beliefs and behaviors of interviewees; classify complex attitude patterns; and understand people's attributions about

what motivates their beliefs and behaviors. They are "stand-alone procedures," meaning they occur outside the bounds of regular fieldwork. Unlike informants, who speak about the scene, respondents "speak only for, and about, themselves" (Lindlof & Taylor, 2011, p. 179).

- *Narrative interviews* solicit entire stories from interviewees, which become the basis for analysis.
- *Focus groups* are interviews conducted with several people at once.

Although in-depth interviews tend to be unstructured, there are certain types of questions that are regularly posed in this type of research. Spradley (1979) identified three types of questions. The first, *descriptive questions*, include tour questions (e.g., "Can you describe a typical day in the emergency room?"), example questions (e.g., "Can you give me an example of order-dumping?"), experience questions (e.g., "In your experience, how do patients react to the computers?"), and native-language questions (e.g., "How did you come up with the phrase 'order-dumping,' and what does it mean?"). Next are *explanation questions*, which explore respondents' site-specific cultural knowledge (Spradley, 1979) (e.g., "Can you help me understand *this* emergency department's EMR adoption?" and "How did the users *here* decide on this particular EMR?"). Finally, *contrast questions* conjure comparisons among dissimilar things to elucidate meaning. For example, researchers may discover a symbol's meaning by asking how it is similar to and different from other symbols (Spradley, 1979) (e.g., "How are Epic EMRs like McKesson EMRs, and in what ways are they different?").

Additionally, Madison (2012) suggested that researchers consider the following:

- *Advice questions* (e.g., "What advice would you give to other emergency medicine providers using EMRs?")
- *Quotation questions* (e.g., "Someone said 'EMRs are inventions of the devil,' but what do you think?")
- "*Once-upon-a-time" questions*, whereby interviewees are invited to share stories that exemplify an occurrence related to the phenomena under study.

In my experience, there is an additional question researchers should pose to participants: "What more would you like to share?" By transferring control of the interview to participants, I effectively allowed them to decide what more I should know about the subject at hand. I found that oftentimes participants offered keen insights on aspects of the topic about which I had not thought to inquire.

The kinds of in-depth interviews you conduct and the types of questions you ask will be determined largely by your topic, site, and sample. You should, however, approach every interview situation in the same way—by giving consideration both to developing rapport with participants and to logistical concerns. Developing rapport, in many ways, depends on researchers' individual traits and mannerisms (i.e., introverted vs. outgoing) and will require building relationships with participants and earning their trust *over time*. Moreover, researchers "must create for the participant the feeling of being respected and of being genuinely heard" (Madison, 2012, p. 39). According to Lindlof and Taylor (2011), active listening is the most crucial way to build rapport during an interview.

Logistical concerns center on when and where to conduct interviews; whether interviews will occur face-to-face, over the phone, or electronically (e.g., via email); and if interviews will be audio-recorded or captured in fieldnotes. First, you must consider where to conduct interviews. Although ethnographic and informant interviews occur regularly during fieldwork, all other in-depth interviews should be "conducted in a convenient and comfortable place for respondents [and] at a time convenient to respondents" (Frey *et al.*, 2000, p. 276).

Next, you should decide how to conduct interviews. Each approach has its advantages and disadvantages, but face-to-face interviews are generally preferred because nonverbal cues provide valuable information missing from other types (e.g., phone or email interviews). Like phone interviews, email interviews lack immediacy but provide a written text, reducing the time and cost associated with transcription (Lindlof & Taylor, 2011). Telephone and email are oftentimes the only available means for interviewing participants for whom face-to-face interviews are impossible given geographical distances (e.g., my interview with Dr. Carbaugh).

Lastly, you should decide if you will record interviews or rely solely on notes. In some instances, particularly when in the field, circumstances and background noise make recording interviews challenging. Generally, though, an audio recording is "the medium of first choice if accuracy and completeness are required" (Lindlof & Taylor, 2011, p. 193). Recorded interviews are typically transcribed, which involves listening to, rewinding, and playing back recorded interviews while typing, word for word, what was said. Although difficult and time-consuming, the transcription process "gives us the first of many opportunities to peruse and reflect upon" the data (Lindlof & Taylor, 2011, p. 212). At the same time, all transcription is theoretical, full of choices that the researcher makes with respect to, for instance, how to represent spoken language in written form (Ochs, 1999).

How to Document Observational Data

According to Lindlof and Taylor (2011), fieldnotes are "concerned with describing and interpreting (i.e., textual) qualities of communication in social action" (p. 159)—they are written accounts of what researchers find significant, namely events and interactions that reveal something about the scene or participants. Fieldnotes can be recorded on the spot (i.e., scratch notes) or can be committed to memory (i.e., headnotes) and written down later. Taking notes in the field, as events and conversations unfold, is preferable to writing notes after leaving the scene for one main reason—more of what is seen and heard makes its way onto the page when written *then and there*. In some instances, however, scratch note taking may not be possible, such as when physical conditions prohibit note taking or when participants are made uncomfortable by it. For example, early in my dissertation fieldwork, providers were suspicious of me because of my note taking, but, eventually, my constant scribbling became the butt of good-natured jokes (e.g., providers likened my notebook to Santa's naughty list). As you navigate your scene and get to know your participants, decide if it makes sense to take notes "then and there" or to write notes after you leave the scene. If you document your experiences afterwards, you should do so as soon as possible.

Fieldnotes serve two major functions. They allow researchers to "develop two important forms of intersubjectivity: (1) empathetic understanding of their participants' experience and (2) successful representation of that understanding for others" (Lindlof & Taylor, 2011, p. 159). To achieve these aims, they should "create a chronological record of your involvement in the scene"; preserve the character of communication observed and/or heard; "contain extensive (if not exhaustive) descriptions of appearances and activities"; provide "rich, specific detail" about what was observed; and "record participants' remarks and conversations as close to verbatim as possible" (Lindlof & Taylor, 2011, pp. 158–159). It is important to note that fieldnotes are "intended to capture and represent the lived experiences of *others*" (Goodall, 2000, p. 90), so consider recording your reactions to what is observed by keeping a separate, personal diary. Taken together with fieldnotes, your diary becomes an important record of your time in the field.

Addressing Validity Threats

In qualitative research, conventional notions of validity derived from postpositivistic research (e.g., predictive validity) do not hold as much relevance (Lindlof & Taylor, 2011), but descriptive validity—"the factual accuracy of the reportage of events" (Lindloff & Taylor, 2011, p. 276)—can be assessed by triangulating multiple sources and methods (Frey *et al.*, 2000; Lindlof & Taylor, 2011). *Triangulation* includes "comparison of two or more forms of evidence [and] if data from two or more methods point toward the same conclusion, then validation is enhanced" (Lindlof & Taylor, 2011, p. 274). "Triangulation can be done with multiple methods," whereby "the researcher looks for convergent data in fieldnotes, interviews, documents, or other qualitative evidence" (Lindlof & Taylor, 2011, p. 274). Additionally, researchers can conduct *member checks*, which involve "taking findings back to the field and determining whether the participants recognize them as true or accurate" (Lindlof & Taylor, 2011, p. 279).

Although you can take steps to ensure the descriptive validity of findings (e.g., triangulation), there is a potential threat to your study's internal validity: the *Hawthorne effect*, whereby people behave differently when they know they are being observed. As Wolcott (2010) noted, however, "No one can keep up appearances forever so the ethnographer eventually sees things as they are" (p. 92). The key to overcoming the Hawthorne effect is to stay in the field long enough for your participants to become comfortable with you, and, ultimately, you will come closer to answering the question "What is going on?"

Deciding When to Leave the Field

Also relevant to time in field is the decision of when to leave. In many cases, the phenomenon under investigation decides the length of observation. For instance, if you are studying a fleeting, one-time occurrence, "only a single observation period is necessary" (Frey *et al.*, 2000, p. 269). In my case, I chose to study an emergency department's EMR adoption from start to finish, which lasted almost 2 years. Long-term research "helps establish quality relationships between researchers and research participants (Frey *et al.*, 2000, p. 270), although "length of stay is no guarantee of better fieldwork" (Wolcott, 2010, p. 101). As Wolcott (2010) elaborated:

> Most of us can act our best selves, at least for a while. But the longer we stay, the less likely we may be able to keep up a front or play a role. Fieldworkers are therefore as apt to overstay their welcome as to leave too soon. In a sense, the longer you stay, the greater your chances of screwing up the relationship, antagonizing someone, or taking a giant misstep. Mistrust is far easier to achieve than trust. (p. 101)

In my experience, when the data collected no longer reveal anything new about the scene or the participants (i.e., saturation is achieved), then it is time to leave the field.

Coding and Analyzing Data

Coding and analyzing narrative-based data, in most cases, begins while researchers are still in the field. The nonlinear, iterative nature of qualitative inquiry allows researchers to collect, sort, analyze, and evaluate assumptions while conducting fieldwork, permitting them to test hunches, fine-tune data collection techniques, and rephrase research questions. After fieldwork has ended, however, researchers move typically into a deeper, more nuanced phase of data analysis. Analyzing fieldnotes and other textual data (e.g., interview transcripts) involves three steps (Lindlof & Taylor, 2011, p. 243):

- *Data management* involves categorizing and sorting.
- *Data reduction* "means that the use value of evidence is prioritized according to emerging schemes of interpretation."
- *Conceptual development* involves recognizing the links between themes, which become "more dense and elaborate" throughout the analysis process.

For example, while working on my dissertation I began the process of managing data by "identifying patterns of behavior," which is the "precursor to finding themes" (Wolcott, 2010, p. 39). I started by *open coding* the data, which is "the initial, unrestricted coding" (Lindlof & Taylor, 2011, p. 250) of data to generate categories. During a second round of open coding, I incorporated *in vivo coding*, which is "coding the terms used by social actors to characterize their own scene" (Lindlof & Taylor, 2011, p. 251). For example, providers routinely mentioned "gaming the system," which included techniques and workarounds to avoid using slow EMRs. Theoretical lenses and sensitizing concepts—interpretative devices or frameworks "through which researchers see, organize, and experience the data" (Tracy & Geist-Martin, 2014, p. 246)—guided my *sorting process* and determined inclusion or exclusion of categories for further analysis. For example, I recorded many instances of food politics playing out in the emergency room, but those instances were excluded from analysis, because, although food politics are an important aspect of emergency room culture, they are unrelated to EMR adoptions. Thus, in sorting data, I also reduced those data to the most salient and applicable categories from which I derived themes.

Owen (1984) established three criteria for identifying and classifying themes. A *theme* emerges when there are (a) recurring descriptions, phrases, or utterances with the same meaning, irrespective of wording; (b) repeated use of the same wording; and (c) forcefulness of expression, as evidenced in vocal pitch and/or volume (Owen, 1984). In reviewing my data, I identified and refined 15 themes connected to the EMR adoption. During conceptual development, I reexamined my themes, reviewed the data again, and revisited the theoretical frameworks. From 15 initial themes, I arrived at 7 major themes and several subthemes that reflected providers' EMR-related talk and actions.

There are many frameworks and approaches for coding data and analyzing themes, but regardless of the approach you take, you need to decide if you will analyze data manually or use computer-assisted qualitative data analysis software (CAQDAS). Manual methods vary among researchers. As Lindlof and Taylor (2011) observed, some researchers cut and paste passages taken from fieldnotes onto notecards, which are grouped around common themes; other researchers sort fieldnotes and place them in piles; still other researchers scribble codes in the margins of their fieldnotes. I opted to color code my fieldnotes based on emerging themes, map out those themes on a whiteboard, and draw lines between similar themes, which helped to highlight conceptual linkages. Because I typed many of my fieldnotes, I was able to use the Search function in Microsoft Word to generate word counts, which helped me also identify themes. CAQDAS can expedite data analysis by providing faster text search and retrieval functions when compared with manual sorting. There are several CAQDAS packages available, including NVivo, MAXQDA, and ATLAS (for a more complete review of qualitative data coding and analysis procedures, see, e.g., Tracy, 2013).

Writing the Report

When writing an ethnographic research report, you can follow "the well-worn formula, I-H-M-R-D" (Wolcott, 2010, p. 139)—introduction, hypothesis, methods, results, and discussion—which we are accustomed to seeing in peer-reviewed academic journals. You may take inspiration from journal articles that include, typically, an introduction; a literature review or theoretical rationale and descriptions of theoretical perspectives; hypotheses or research questions; data collection methods, which incorporate descriptions of participants, setting(s), and sampling strategies; data analytic procedures; results; and discussion. An ethnographic report, however, should also offer interpretations of the scene as seen from the participants' point of view (Carbaugh, 1991).

Although using this formula to guide your writing is acceptable, there are other ways to approach writing an ethnographic study. Consider, for instance, Van Maanen's (2011) typology of ethnographic tales that distinguishes between realist, confessional, and impressionist tales.

- *Realist tales* are narrated in "a dispassionate, third person voice … the result is an author-proclaimed description and something of an explanation for certain specific, bounded, observed (or nearly observed) cultural practices" (Van Maanen, 2011, p. 45). Realist tales typically remove the "I" and, in so doing, assume a "studied neutrality" and prioritize participants' "sayings, doings, and supposed thinking" (Van Maanen, 2011, p. 47).[5]
- *Confessional tales*, by contrast, are highly personalized and center on fieldworkers' experiences and their interpretations of events. For Van Maanen (2011), the fieldworker's point of view is provided "as something of a character-building conversion tale in which the fieldwork, who saw things one way at the outset of the study, comes to see them in an entirely diffident way by the conclusion of the study" (p. 77).
- *Impressionist tales* strike a balance between the native's perspective and the fieldworker's perspective by keeping both "the subject and object in constant view." They focus on the exceptional, rather than the mundane, and resist making claims or offering interpretations (Van Maanen, 2011, p. 102). Additionally, impressionist tales tend to be engaging, dramatic accounts of characters and events.

5 *Realism* is a philosophical stance on the nature and existence of constructs and/or generals. For listening scholars, for instance, there are debates regarding whether something like "listening comprehension" really exists. Clearly, whether a person answered a certain number of multiple-choice questions correctly is not an issue, but the degree to which the answers to those questions represent something bigger is not as clear-cut. For a realist, abstract constructs have an existence apart from the way in which they are operationalized. For antirealists, of which there are many brands, these constructs are often characterized as "convenient fictions," useful for talking in the abstract but having no real existence in any ontological sense.

Your report may combine aspects of all three approaches, but you should take your intended audience into account when deciding how to structure your report for publication.

Lindlof and Taylor (2011) detailed commonly accepted practices for writing qualitative research reports or narratives that warrant your consideration: "Above all else, narratives should not be *boring*" and should follow a coherent structure, with a setup, middle, and ending (Lindlof & Taylor, 2011, p. 313). Reports should be well-written, plausible, and interesting; address multiple audiences; demonstrate authorial awareness; and alternate between the researcher's experience and interpretation. Authors should also address ways they shaped events in the field, explore how their writing may be shaped by and perpetuate hegemonic ideologies, and explain how their work is substantive and scalable (Lindlof & Taylor, 2011).

Conclusion

In this chapter, I have explored some of the characteristics of qualitative methods, while demonstrating the utility of qualitative approaches for listening researchers. I have described ways that qualitative researchers can enhance methodological rigor. I have also provided an overview of steps involved in "doing" ethnography.

Returning to the snapshot–movie analogy referenced in the introduction to this chapter, although I have advocated for qualitative approaches—that is, moviemaking over taking snapshots—undoubtedly, "quantitative" snapshots and "qualitative" movies each have their place in listening research. I ask only that you consider all the available tools when deciding how you will examine listening phenomena.

References

Carbaugh, D. (1989). The critical voice in ethnography of communication research. *Research on Language and Social Interaction, 23*, 261–282. doi:10.1080/08351818909389324

Carbaugh, D. (1991). Communication and cultural interpretation. *Quarterly Journal of Speech, 77*, 336–342. doi:10.1080/00335639109383965

Carbaugh, D. (1999). "Just listen": "Listening" and landscape among the Blackfeet. *Western Journal of Communication, 63*, 1999, 250–270. doi:10.1080/10570319909374641

Carbaugh, D. (2005). *Cultures in conversation*. Mahwah, NJ: Erlbaum.

Carbaugh, D. (2006). A role for communication theory of ethnography and cultural analysis. *Communication Theory, 2*, 156–165. doi:10.1111/j.1468-2885.1992.tb00035.x

Cooper, S., & Endacott, R. (2007). Generic qualitative research: A design for qualitative research in emergency care? *Emergency Medicine Journal, 24*, 816–819. doi:10.1136/emj.2007.050641

Creswell, L. (2014). *Research designs: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.

Deakins, A. H., Osterink, C., & Hoey, T. (1987). Topics in same sex and mixed sex conversations. In L. B. Nadler, M. J. Nadler, & W. R. Todd-Mancillas (Eds.), *Advances in gender and communication research* (pp. 89–108). Lanham, MD: University Press of America.

Dutta, M. J. (2014). A culture-centered approach to listening: Voices of social change. *International Journal of Listening, 28*, 67–81. doi:10.1080/10904018.2014.876266

Eisenberg, E. M., Baglia, J., & Pynes, J. E. (2006). Transforming emergency medicine through narrative: Qualitative action research at a community hospital. *Health Communication*, *19*, 197–208. doi:10.1207/s15327027hc1903_2

Eisenberg, E. M., Murphy, A. G., Sutcliffe, K., Wears, R., Schenkel, S., Perry, S., & Vanderhoef, M. (2005). Communication in emergency medicine: Implications for patient safety. *Communication Monographs, 72*, 390–413. doi:10.1080/03637750500322602

Fiske, J. (1991). Writing ethnographies: Contribution to a dialogue. *Quarterly Journal of Speech, 77*, 330–335. doi:10.1080/00335639109383964

Forsey, M. G. (2010). Ethnography as participant listening. *Ethnography, 11*, 558–572. doi:10.1177/1466138110372587

Frey, L. R., Botan, C. H., & Kreps, G. L. (2000). *Investigating communication: An introduction to research methods* (2nd ed.). Boston: Allyn & Bacon.

Gold, R. L. (1958). Roles in sociological field observations. *Social Forces, 36*, 217–223.

Goodall, H. L. (2000). *Writing the new ethnography*. Lanham, MD: AltaMira Press.

Hartwig, R. (2010). *A facilitative ethnography of the communication of an evangelical church leadership team* (Unpublished doctoral dissertation). University of Colorado at Boulder.

Hymes, D. (1962). The ethnography of speaking. In T. Gladwin & W. C. Sturtevant (Eds.), *Anthropology of human behavior* (pp. 13–53). Washington, DC: Anthropology Society of Washington.

Kovarsky, D. (2008). Representing voices from the life-world in evidence based practice. *International Journal of Language and Communication Disorders, 43*, 47–57. doi:10.1080/13682820701698036

Lindlof, T. R., & Taylor, B. C. (2011). *Qualitative communication research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Lipari, L. (2010). Listening, thinking, being. *Communication Theory, 20*, 348–362. doi:10.1111/j.1468-2885.2010.01366.x

Lipari, L. (2012). Rhetoric's other: Levinas, listening, and the ethical response. *Philosophy and Rhetoric, 45*, 227–243. doi:10.1353/par.2013.0001

Madison, D. S. (2012). *Critical ethnography: Method, ethics, performance* (2nd ed.). Thousand Oaks, CA: Sage.

Meldrum, H. (2011). The listening practices of exemplary physicians. *International Journal Listening, 25*, 145–160. doi:10.1080/10904018.2011.604603

Miller, K. (2006). *Communication theories: Perspectives, processes, and contexts* (2nd ed.). Boston: McGraw-Hill.

Molina-Markham, E. (2014). Finding the "sense of the meeting": Decision making through silence among Quakers. *Western Journal of Communication, 78*, 155–174. doi:10.1080/10570314.2013.809474

Morse, J., & Niehaus, L. (2009). *Mixed method design: Principles and procedures*. Walnut Creek, CA: Left Coast Press.

Nicotera, A. M., & Clinkscales, M. J. (2010). Nurses at the nexus: A case study in structurational divergence. *Health Communication, 25*, 32–49. doi:10.1080/10410230903473516

Nicotera, A. M., & Mahon, M. M. (2013). Between rocks and hard places: Exploring the impact of structurational divergence in the nursing workplace. *Management Communication Quarterly, 37*, 90–120. doi:10.1177/0893318912458214

Nicotera, A. M., Mahon, M. M., & Zhao, X. (2010). Conceptualization and measurement of structurational divergence in the healthcare setting. *Journal of Applied Communication Research, 38*, 362–385. doi:10.1080/00909882.2010.514001

Nugus, P., Holdgate, A., Fry, M., Forero, R., McCarthy, S., & Braithwaite, J. (2011). Work pressure and patient flow management in the emergency department: Findings from an ethnographic study. *Academic Emergency Medicine, 18*, 1045–1052. doi:10.1111/j.1553-2712.2011.01171.x

Ochs, E. (1999). Transcription as theory. In A. Jaworski & N. Coupland (Eds.), *The discourse reader* (pp. 167–182). London: Routledge.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), 943–943. doi:10.1126/science.aac4716

Overton, B. C. (2015). *Examining electronic medical records system adoption and implications for emergency medicine practice and providers* (Unpublished doctoral dissertation). Louisiana State University, Baton Rouge.

Owen, W. F. (1984). Interpretive themes in relational communication. *Quarterly Journal of Speech, 70*, 274–287. doi:10.1080/00335638409383697

Perlmutter, D. D. (2015). In defense of ethnography. *The Chronicle of Higher Education*. Retrieved February 1, 2017, from http://chronicle.com/article/In-Defense-of-Ethnography/231191/

Philipsen, G. (1975). Speaking 'like a man' in Teamsterville: Cultural patterns of role enactment in an urban neighborhood. *Quarterly Journal of Speech, 61*, 13–22. doi:10.1080/00335637509383264

Philipsen, G. (1977). Linearity of research design and ethnographic studies of speaking. *Communication Quarterly, 25*, 42–50. doi:10.1080/01463377709369262

Philipsen, G. (1982). The qualitative case study as a strategy and communication inquiry. *Communicator, 12*, 4–17.

Philipsen, G., & Carbaugh, D. (1986). A bibliography of fieldwork in the ethnography of communication. *Language in Society, 5*, 387–398. doi.org/10.1017/S0047404500011829

Podkalicka, A. (2009). Young listening: An ethnography of youth Worx Media's radio project. *Journal of Media and Cultural Studies, 23*, 561–572. doi:10.1080/10304310903015704

Purdy, M. W. (2000). Listening, culture and structures of consciousness: Ways of studying listening. *International Journal of Listening, 14*(2000), 47–68. doi:10.1080/10904018.2000.10499035

Purdy, M. W. (2010). Qualitative research: Critical for understanding listening. In A. D. Wolvin (Ed.), *Listening and human communication in the 21st century* (pp. 33–45). Oxford: Wiley-Blackwell.

Purdy, M. W. (2011). Grounding listening: Limitations of theory. *International Journal of Listening, 25*, 132–138. doi:10.1080/10904018.2011.537144

Schwartz, M. S., & Schwartz, C. G. (1955). Problems in participant observation. *American Journal of Sociology, 60*, 343–354. doi:10.1086/221566

Spradley, J. P. (1979) *The ethnographic interview*. New York: Holt, Rinehart & Winston.

Tracy, K., & Muñoz, K. (2011). Qualitative methods in interpersonal communication. In M. L. Knapp & J. Daly (Eds.), *The Sage handbook of interpersonal communication* (pp. 59–86). Thousand Oaks, CA: Sage.

Tracy, S. J. (2013). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact*. Oxford: Wiley-Blackwell.

Tracy, S. J., & Geist-Martin, P. (2014). Organizing ethnography and qualitative approaches. In L. L. Putnam, & D. K., Mumby (Eds.), *The Sage handbook of organizational communication: Advances in theory, research, and methods* (pp. 245–269). Thousand Oaks, CA: Sage.

Van Maanen, J. (2011). *Tales of the field: On writing ethnography* (2nd ed.). Chicago, IL: University of Chicago Press.

Wolcott, H, F. (2010). *Ethnography lessons: A primer.* Walnut Creek, CA: Left Coast Press.

4

Modeling and Measuring Cognitive Components of Listening

Debra L. Worthington

Chapter 1 introduced the three primary components of listening: cognitive, affective, and behavioral. *Affective* components of listening address individual perceptions of listening as well as motivation and enjoyment derived from listening activities, and listening *behaviors* include verbal and nonverbal actions that signal attention and interest. The focus of this chapter is on *cognitive* elements of listening, those internal processes utilized by individuals as they attend to, comprehend, interpret, evaluate, and make sense of spoken language.

We placed this chapter before the affective and behavioral components because the study of cognition represents the earliest foray into listening research. As noted in Chapter 1, Nichols's (1948) work, which addressed how students comprehend classroom lectures, set the stage for most of the listening research conducted throughout the 1950s and into the 1980s. During this time, listening scholars concentrated on three primary activities. First, there was extensive focus on developing definitions of listening, trying to solidify exactly those cognitive processes that could collectively explain what listeners do when faced with processing information aurally. As is evident in the definitions reviewed in Chapter 1, attention, comprehension, and retention were particularly salient to early researchers. Second, and not unrelated to defining listening, scholars spent a good deal of time developing models of the listening process, and most prominently featured cognitive elements (for review, see Wolvin, 1989). Finally, scholars developed measures of listening from these models. Similar to definitions and models, listening measures were primarily designed to tap attention, comprehension, and retention.

This chapter has two primary goals. First, it introduces several listening models and how they identify and emphasize four key cognitive features of listening: attention, retention (or memory), comprehension, and inference making. Second, it examines how listening scholars, past and present, have conceptualized and operationalized these four cognitive components.

Modeling Listening

Models are a principle tool for social scientists. Although several different types of models exist, most models of listening are *process models*, attempting to represent the procedural nature of listening as it happens inside the mind of a recipient. In his

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

influential work on communication models in the social sciences, Deutsch (1952) identified a number of important functions served by process-based models. First, process models are *organizational*, providing structure, specifying operating rules, and identifying connections between components. Thus, models bring order to seemingly disjointed data, illuminating previously unknown connections. At the same time, models serve a *heuristic* function, suggesting areas of research and potential hypotheses. Models also can serve a *predictive* function. The predictive utility of a model can range from simple decisions of presence or absence to more complex quantifications of when or how much. When a model is able to make specific, quantifiable predictions, Deutsch claimed it can serve a *measurement* function. In sum, then, a successful model must: (a) accurately reflect the primary elements of the objects under study along with their interrelationships, (b) frame possible hypotheses, and (c) suggest means of measurement (Deutsch, 1952).

From Deutsch's (1952) description of a successful model, many listening models are quite basic. As the examples in this section illustrate, although many listening models may serve a heuristic function by encouraging and directing research, most have little or no predictive value, nor do they tend to propose specific measurement options.

Taylor's Listening Model

Taylor's (1964) listening model (Figure 4.1) was one of the first to present listening as a distinct process. Taylor described listening as "the total act of receiving auditory communication" (p. 5). His emphasis on the "auditory" was clearly reflected in the three sequential stages of hearing, listening, and auding that comprised the model. *Hearing* refers to the reception of speech sounds and addresses elements such as auditory fatigue, volume, and context. *Listening* includes attention, concentration, rate of input, as well as misunderstanding and emotional responses. It is within the listening stage that meaning begins to be assigned. *Auding* describes the internal process by which words gain meaning for the listener, including elements such as evaluation and general impressions. Taylor's model also presents a number of additional variables that may affect the listening process at any point in time (e.g., experience and background).

Taylor's (1964) model meets several of the functions described by Deutsch (1952): It clearly identifies primary and secondary components, illustrates how they are connected, and suggests when and where they come into play (organizational function). In addition, the heuristic function seems to be met as the model proposes testable hypotheses (e.g., the delivery and rate of input should influence listening directly and auding indirectly, through its influence on listening). The model's predictive and measurement functions, however, are constrained. There is no specification as to how much and in what ways elements such as background and experience should influence hearing, listening, and auding (limited predictive utility), and there is no clear specification of a measurement model (e.g., should one test the presence of meaning? Complexity? And how?).

Taylor's (1964) model does, however, illustrate the relation between definitions and models of listening. Clearly, cognitive processes are the focus of this model. Notable in this model is the key role that attention plays to listening processes. Attention becomes the transitional point between hearing and listening. The relation of attention to listening remains a component of listening models today. Interestingly, Taylor separated listening from auding, a practice that has since been discarded in favor of

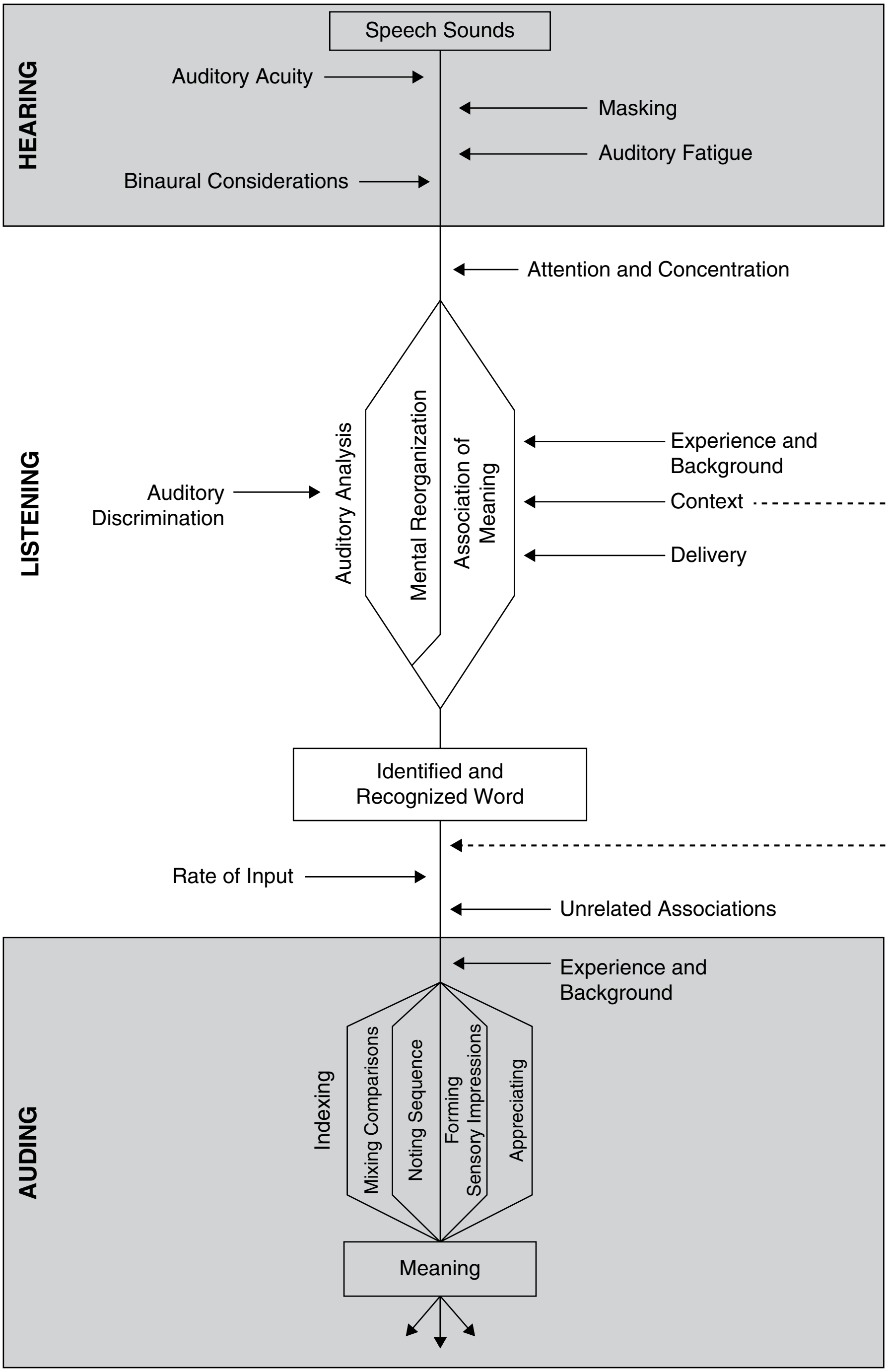


Figure 4.1 Taylor's (1964) listening model.

combining meaning making and auditory processing under the umbrella of listening. Still remaining in contemporary models, however, is a separation of hearing and listening, with the former described as the allocation of attention to available sounds and the latter as purposeful and directed comprehension.

Attention, as the factor distinguishing listening from hearing, continues to be emphasized in contemporary listening research. Rost (2011) argued, "Because of the deliberate nature of attention, we can consider attention to be the beginning of involvement, which is the essential differentiation between hearing and listening" (pp. 19–20). The view of hearing as perception of sound related to the physiological components of sound reception likely explains why some later models of listening place much less emphasis on hearing, focusing instead on listening as an interaction-based and relationally based process.

The MASTER Model

Whereas Taylor's (1964) model is *descriptive* (describing the underlying processes that make listening possible), the MASTER model of listening proposed by Mills (1974; Figure 4.2) is *prescriptive* in nature. Mills's model illustrates what students *should* do after attention is allocated to some orally delivered stimulus (rather than what they *actually do* when listening). The *mental* decision to listen requires an *active* response from listeners that must be *sustained* by a listening *target,* who should work to *eliminate* filters and other barriers to listening. By doing these things, listeners not only are ready to listen, but also should be able to *remember* what they have heard and subsequently be MASTER listeners.

Mental
Active
Sustain
Target
Eliminate
Remember

Figure 4.2 Mills' (1974) model of listening.

Although unique in its prescriptive character, the MASTER model does share with other cognitive models a characterization of listening as an interactive process. Moreover, it is one of the first to introduce memory and barriers to processing as additional cognitive elements—characteristics that remain part of many current models. Whereas the end product of Taylor's (1964) model was making meaning, Mills (1974) viewed memory as listening's end product. Although the MASTER model does not serve advanced predictive or measurement functions, it does suggest that predictions and measurement should focus on information acquisition, a focus that seems to mark most contemporary listening scholarship through the 1980s (Bostrom, 1990). The new emphasis on memory was reflected in later models (e.g., Brownell, 1986; Wolff, Marsnik, Tacey, & Nichols, 1983; Wolvin, 1989) and marked an important addition to the cognitive processes believed to contribute to listening. The Wolff *et al.* model illustrates the rising prominence of memory to listening. As seen in Figure 4.3, retention (memory) is visually depicted as the highpoint of listening.

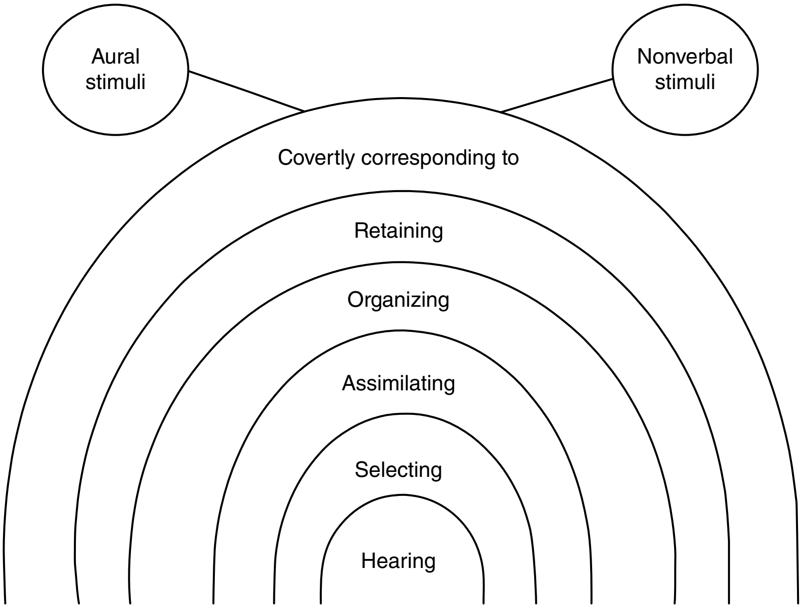


Figure 4.3 Wolff, Marsnik, Tacey, and Nichols' (1983) model of listening. *Source*: Wolff (1983). Reproduced with permission of Nadine Marsnik.

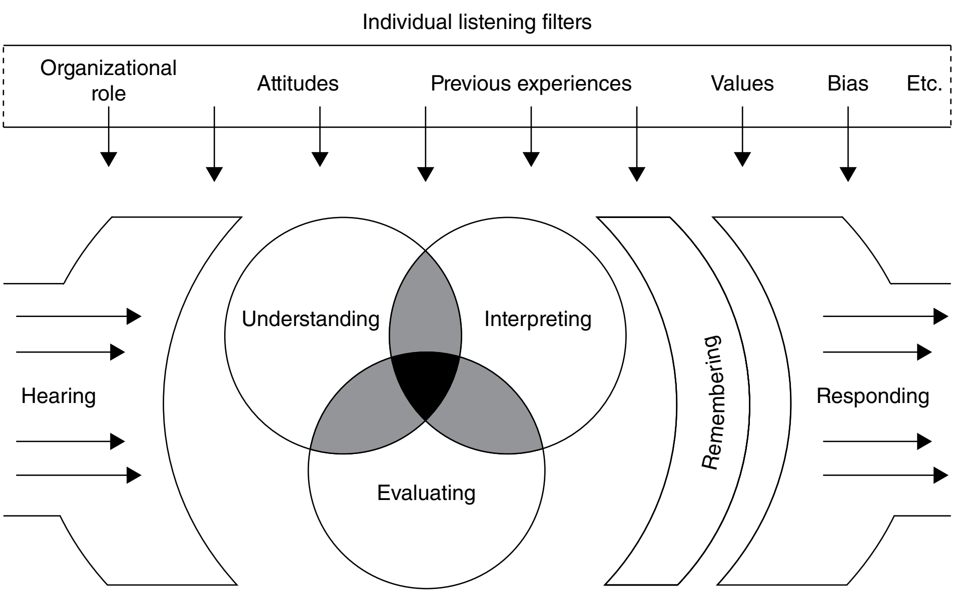


Figure 4.4 Brownell's (1986) model of listening. *Source*: Brownell (1986). Reproduced with permission of Judie Brownell.

HURIER Model of Listening

At about the same time that memory was being stressed as an important component of listening (Bostrom & Bryant, 1980; Bostrom & Waldhart, 1983, 1988), the role of understanding/comprehension and interpretation began to gain interest. Brownell's (1986)

HURIER model (i.e., ***h***earing, ***u***nderstanding, ***r***emembering, ***i***nterpreting, ***e***valuating, and ***r***esponding) prominently featured these factors (see Figure 4.4). More sophisticated than many previous ones, this model incorporated the idea of *filters* that may affect our listening processes as well as interactive elements such as responding. Finally, it suggests several areas of study, including the impact of attitudes and values and one's role and biases on listening. It does not, however, include nonverbal stimuli. Despite this shortcoming, of the models examined thus far, Brownell's comes closest to meeting the first three of Deutsch's (1952) four functions of a good model (i.e., organizational, heuristic, and predictive).

Cognitive Components

The models discussed above reflect several fundamental cognitive components that became and remain key areas of interest to listening scholars: attention, memory, comprehension, and inference making. Each of these areas is reviewed in this section, with a particular focus on how they have been conceptualized and operationalized (see also Chapter 2).

Attention

The role of attention (or attending) has been a constant feature in the listening literature. For listening scholars, what and how we attend to messages affect all other areas traditionally included in the cognitive domain of listening (e.g., retention and understanding). In particular, many scholars have stressed the notion of *selective attention*, or the preferential processing of one signal over competing distractor signals (see Kerlin, Shahin, & Miller, 2010), which is thought to play an essential part of our listening lives. Selective attention allows us to discriminate and process one set of sounds, while seemingly ignoring others. It also prevents us from being overwhelmed by the stimuli we constantly receive.

Conceptualization

Attention has been conceptualized in a number of ways. From an anatomical perspective, attention is an integrated activity of the neocortex, thalamus, and brain stem, which are linked by the reticular activating system. For instance, Bizley and Cohen (2013) charted the process of attention from initial hearing of a sound (i.e., an auditory object) until it is mentally perceived by the individual. Early listening studies approached attention from a framework of limited capacity. From this perspective, selective attention results from our attempts to protect limited resources associated with our sensory systems (Lang & Basil, 1998), particularly those associated with recall (Baddeley & Hitch, 1974). Later research relied on the "human information-processing framework of cognitive psychology" (Buračas, Saenz, & Boynton, 2003, p. 242), a conceptual shift that led many researchers toward a goal-oriented approach (Allport, 1993; Cohen, 1993; Cowan, 1995), where sensory information is filtered based on relevancy to a specific desired end state (e.g., remembering details or enjoying a conversation).

Operationalization

As noted in Chapter 2, *operationalization* is the process of defining the measurement of a construct. Because the conceptualizations of attention outlined in this chapter are abstract, how attention is ultimately operationalized will inevitably be selective, reflecting how the construct is defined as well as any specialized measurement needs (e.g., readability, delivery method, and time). The vast majority of listening measures of attention focus on self-reports of perceptions of attention and attending behaviors. Relatively few measures focus on attention alone—most incorporate select items to assess the construct. The following examples, drawn from measures profiled in this volume, illustrate ways in which attention has been operationally defined.

Daly, Vangelisti, and Daughton (1987) sought to better understand individual differences in *conversational sensitivity* (CS), which is "the propensity of people to attend to and interpret what occurs during conversation" (p. 169). In the CS scale, attention is equated with "identification" (i.e., "… identify power relationships …" and "identify underlying and/or multiple meanings …"). Several similar measures assess individual perceptions of listening in their organization or workplace and tend to assess perceptions of being attended to as a means of gauging a supportive work environment. For instance, Reed, Goolsby, and Johnston (2014) developed the Team Listening Environment (TLE) scale to measure coworkers' perceptions of communication behaviors that exhibit genuine attention and understanding from others. The TLE scale emphasizes the speaker's perception of a listener's behavior (i.e., "The other group members paid attention to me"). Thus, attention is viewed as an element of supportive communication, and one's score is a quantitative indicator of one's affective perception of team listening.

As seen in Figure 4.4, Brownell's HURIER model incorporates six interrelated listening subskills: hearing, understanding, remembering, interpreting, evaluating, and responding (Brownell, 1996, pp. 71ff.). In the hearing component, she essentially combines hearing and attention because, from Brownell's perspective, individuals use hearing to perceive, discriminate, and identify sounds, which are then used to adjust attentional focus. Brownell's self-report measure also includes an affective component of attention. Hearing items include "I overcome distractions such as the conversation of others, background noises, and telephones, when someone is speaking"; "I enter communication situations with a positive attitude"; and "I concentrate on what the speaker is saying, even when the information is complicated." In a similar manner, the Facilitating Listening Scale (FLS) (Bouskila-Yam & Kluger, 2011) measures perceived attention from others as conveyed in the following items: "expresses interest in my stories," "listens to me attentively," and "pays close attention to what I say."

Of course, the only way to measure *actual* attention (and not just a perceived ability to attend or a general tendency of others to attend) is to get into someone's head. Although a number of different techniques have been used to study selective attention, such as electroencephalogram (EEG), magnetoencephalogram (MEG), and functional magnetic resonance imaging (fMRI), a staple in neuroscience is the dichotic listening task (DLT). For almost half a century, DLTs have been used by cognitive psychologists and, more recently, neuropsychologists, particularly those examining asymmetries in auditory speech processing. In a DLT, subjects listen to recordings of co-occurring spoken messages played to each ear and are asked to attend to a specified message (e.g., tone, word, and sound). Scoring typically focuses on the percentage correct for each ear and an assessment of the ear advantage (EA) (Speaks, 1988). Individuals may be categorized as having a right-ear advantage (REA), having a left-ear advantage (LEA), or having no ear advantage (NoEA). Scoring results are believed to reflect cerebral hemisphere dominance for processing the types of signals (e.g., nonsense syllables, digits, and musical tones) presented to a listener (for more on scoring EA, see Speaks, 1988).

Neuroscientists also are interested in mapping the areas of the brain and identifying the neurological systems associated with hearing and listening. As outlined in Profile 21 by Burunat and Brattico, fMRI techniques have been used to map brain responses during

DLTs (Jäncke, Specht, Shah, & Hugdahl, 2003) as well as when exposed to the spoken word (Menenti, Gierhan, Segaert, & Hagoort, 2011). Although a promising technique, neuroimaging studies of listening and listening-related processes are limited, in large part, due to the nature of the technology itself. First, in order to examine the relations between mental processes and brain functions, listening processes must be clearly defined and able to be "validly and selectively manipulated" (Spunt, 2013, p. 63). Unfortunately, as discussed here and elsewhere in this *Sourcebook*, listening processes are complex and currently lack the theoretical clarity to allow for such study. Second, it is notoriously difficult to study naturalistic phenomena, like listening, using neuroimaging techniques because the technology itself limits actions that are a natural part of listening processes (e.g., speech, eye gaze, and means of response). As Spunt (2013) wrote, "[T]he methodological constraints of fMRI research present a situation that is extremely low on what is traditionally termed *mundane realism*" (p. 63, emphasis in original). Thus, the ecological validity of the method must be kept in mind when considering the relevance of this technique. Third, Mather, Cacioppo, and Kanwisher (2013) argued that fMRI is inappropriate for answering questions of causation (i.e., the causal link between a particular brain region and a particular task) (p. 111). Finally, studies requiring precise timing of the human brain should use other techniques that measure at tens or hundreds of milliseconds (e.g., event-related potential [ERP], MEG, intracranial recordings, and transcranial magnetic stimulation [TMS]) (for additional considerations, see Mather *et al.*, 2013).

Memory

Like attention, memory is featured prominently in many listening models. This emphasis likely resulted from Kelly's findings (1965, 1967) that early measures of listening comprehension were more closely related to general measures of intelligence than with each other. The introduction of memory processes into conceptualizations of listening was important, as it allowed researchers to better differentiate listening from other information-processing tasks (Bostrom, 1990).

Short-term recall of information was a particularly important part of listening assessment for several decades. This emphasis was reflected in the development of listening comprehension tests and was based on the belief that individuals exposed to the factual content of a lecture should be able to demonstrate retention when tested (Beighley, 1952; Brown & Carlsen, 1955; Nichols, 1948, 1974). Competent listening was evidenced by the amount of information that could be recalled, typically following the presentation of a 10-minute audiotaped lecture. A number of tests were developed, among them the Brown-Carlsen Listening Comprehension Test (BCLCT; Brown & Carlsen, 1955), the Kentucky Comprehensive Listening Test (KCLT; Bostrom & Waldhart, 1983), the Watson-Barker Listening Test (WBLT; Watson & Barker, 1983), and the Sequential Tests of Educational Progress (STEP; Educational Testing Service, 1957). Of these tests, however, the KCLT was the only one to separate its factors solely on the basis of types of memory utilized while listening: (a) short-term listening, (b) listening with rehearsal, (c) interpretive listening, (d) lecture listening, and (e) short-term listening with distractions (Bostrom & Waldhart, 1983). Brownell (2006) and others (e.g., Bostrom & Waldhart, 1980) have proposed that memory should be further broken down into echoic (i.e., immediate), short-term, and long-term memory.

One criticism leveled against listening research is that it is not grounded in recent attention and memory research, but rather in outdated linearly based research (Janusik, 2010). For instance, Janusik (2010) has argued that listening researchers should look to research in *working memory* (WM; Baddeley & Hitch, 1974), which has been thoroughly tested and whose initial conceptualization has remained largely intact (Baddeley, 1986, 2000, 2003, 2007). WM is comprehensive and based on a four-part structure, accounting for many types of remembering. Listening research has addressed the relations among types of memory (e.g., working memory and long-term memory) and communication context. This research suggests that memory "needs" vary with the context (Bostrom & Bryant, 1980; Waldhart & Bostrom, 1981). Thus, listening to a lecture will carry expectations of long-term memory storage, whereas conversational listening will depend more heavily on short-term listening. Other scholars have challenged the notion that there is a single working memory, arguing that individuals possess several, which may be distinguished by modality (e.g., speech and writing) and by what is being represented (e.g., spatial, serial, and verbal) (see Ronnberg, Rudner, Foo, & Lunner, 2008; Rost, 2011).

Certainly, our understanding of memory has grown as researchers in psychology and the neurosciences seek to discover how the human brain works. For instance, early studies by Teasdale (Teasdale & Fogarty, 1979) and others (e.g., Isen, Shalker, Clark, & Karp, 1978) found that emotional state effects cognitive processes. Their research suggested that information that is closer to our emotional state will be easier to recall. So, for instance, when we are happy, we will more easily recall positive events, essentially leading us to a cognitive loop feeding our memory and our emotional state. Similarly, the affective nature of the words and phrases utilized in memory tests can affect recall (Merluzzi, Rudy, & Glass, 1981).

Whatever the perspective, listening scholars acknowledge that different types of memory are interrelated, that memory use varies with the listening context, and that there are individual differences in memory competence.

Conceptualization

As noted in this chapter, memory, as short-term recall, has long been of interest to listening scholars and was incorporated into most early listening measures. From this perspective, memory was conceptualized by how much information could be retained over a short period of time. Although many of these measures often identified themselves as measures of comprehension, more often than not, they actually emphasized accurate recall (i.e., short-term memory).

It also is important for researchers to keep in mind that different types of memory are used in different ways (Bostrom, 2011). Thus, how memory is conceptualized may differ depending on the focus of study. Memory can be divided into that which can be consciously accessed (i.e., explicit or declarative memory) and that related to conditioned reflexes and motor skills (i.e., implicit or procedural) (Hodges, 1994). Listening scholars, not surprisingly, focus on explicit memory, which can be further subdivided into episodic memory (those related to personal experiences and specific events) and semantic memory (those related to words and their meanings as well as our general world knowledge). Whereas episodic memory is sensitive to time and context, semantic memory is not (Hodges, 1994). Both of these types of memory are elements of our long-term memory. Our short-term or working memory typically refers to the recall of new verbal or spatial information over a short time period (i.e., 5–30 minutes). Examples of implicit (procedural) memory include learned responses and motor skills such as driving and playing the piano.

One context of listening and memory assessment is the area of second language (L2) learning. In order to make sense of the new language, students must retain knowledge of a number of different characteristics, such as vocabulary and language structure. L2 scholars often focus on memory capacity, semantic memory, and concentration when devising comprehension measures (Aryadoust, Goh, & Lee, 2012). Another area of interest is the role of memory in conversational contexts. For example, the Memory for Conversation (MC) instrument (Sillars, Weisberg, Burggraf, & Zietlow, 1990) examines the type and amount of information individuals retain following a conversation (see Profile 38). Similarly, Janusik's (2007) Conversational Listening Span (CLS; see Profile 12) assesses conversational listening capacity (i.e., the number of items that an individual can hold active, paraphrase, and respond to during a conversation). Although somewhat similar in their conceptualizations, they are operationalized quite differently.

Operationalization

Listening scholars have operationalized memory in myriad ways. As noted in this chapter, early scholars focused on developing assessment tests that centered on short-term memory as evidenced by factual recall. These tests employ a multiple-choice format with one correct answer. One notable difference is the attempt to use more naturalistic techniques. For example, several measures of conversational memory incorporate dyadic interaction (e.g., CLS and MC). For the CLS, memory is measured using a holistic rating scale that assesses, in part, an individual's ability to paraphrase presented information; the more information paraphrased, the greater the conversational listening span (and presumably the memory capacity). In contrast, the MC measure has individuals either engage in a conversation or observe one. It incorporates an activity (e.g., an irrelevant task or a video) that is believed to disrupt short-term memory. MC assessment is more flexible, with varying levels of cues (or none) used to direct participants.

The first method of MC assessment, *free recall*, is an open-ended method in which participants recall the conversation they participated in or observed. Recall is made without outside assistance (i.e., no memory cues), and idea units and other aspects of the information remembered are coded using a standardized rubric (e.g., Stafford & Daly, 1984). In contrast, *cued* recall provides participants with some level of direction about the information to be retrieved from memory of the conversation. Assessment rests on the amount of information related to the cue that can be recalled. Finally, the most structured method of assessment is *recognition*, which involves providing participants with specific items, one at a time; individuals indicate if each item occurred during the conversation and, if yes, how often.

The different methods of measuring memory for conversation require different scoring. For instance, studies utilizing free recall may measure memory based on conversational themes, elaborations, and reproductions that are identified, whereas others may use cued recall questions (e.g., sequential cues may address temporal elements of a conversation). When this latter scoring method is utilized, responses that do not correspond with the earlier conversation are recorded as a lack of recall (see Benoit & Benoit, 1988). A fuller discussion of how MC is operationalized can be found in Profile 38.

CLS also relies on conversations but affords less opportunity for a naturalistic conversation than does MC. Instead, when measuring CLS, the researcher engages in interviews with participants using a script. CLS does, however, reflect the ebb and flow of a conversation in that individuals can recall and paraphrase material in any order. In addition, accuracy is based on whether the information is paraphrased and responded to, not whether the participant knows the "correct" answer. Based on lenient rules, an acceptable response can include “I don’t know” (assuming the respondent paraphrases the question asked). Essentially, as the conversation continues, information is added until the participant can no longer paraphrase the material introduced (see Profile 12).

Other measures that tap facets of memory are easier to administer as they rely on self-report data. For example, the Academic Listening Self-Assessment (ALSA) is used by academic English language learners to assess their personal listening skills with the goal of identifying strengths and weaknesses (Aryadoust *et al.*, 2012; see Profile 1). Memory components of the ALSA address memory capacity and ability to concentrate (e.g., “[When I am listening in English, I can …] often remember much of the content of the lecture a day later”). The HURIER model moves beyond self-assessment of personal memory capacity, introducing the self-assessment of an affective component to memory processes as well as addressing multiple contexts (e.g., “I can listen to and accurately remember what my partner says, even when I strongly disagree with her viewpoint,” or “I can remember what the instructor has said in class even when it’s not in the book”; see Profile 23 and also Chapter 5).

Research in cognitive psychology and neuroscience is expanding our understanding of the relations between the brain and memory. For example, using fMRI techniques, researchers are studying the physical changes that occur in the brain when we form memories (Gross *et al.*, 2013). One important finding is that the brain is amazingly efficient at storing information—only a few neurons fire when we consider a particular individual (e.g., your mother vs. your best friend) or place (home vs. Paris). This selective activation even holds true for objects (a tennis racquet vs. a baseball) and, in some instances, the letter strings associated with the person, place, or object under study (i.e., reading your mother’s name activates the same area as viewing a photograph of her). As Gross *et al.* (2013) noted, “These results suggest an invariant, sparse and explicit code, which might be important in the transformation of complex visual percepts into long-term and more abstract memories” (p. 1102). A slight word of caution, however, lest we generalize too broadly: Much of the research in this area has focused on visual perceptions and cues, not aural ones. Less is known about how the brain translates auditory percepts into meaning and memory.

Particularly relevant for processing orally delivered information is the fact that memory is a constructive process (Neath & Surprenant, 2003). As a result, our memories are often inaccurate; and because, at least in real life, we do not have a stable text upon which to refer, there is often no way to reconcile different perceptions of who is right (the old “he said, she said” problem). In contrast, most memory measures rely on recognition and recall. When we complete a multiple-choice exam, we rely on recognition to help us choose from our options. Recall comes into play if the same exam includes fill-in-the blank questions. In most, but not all, instances, individuals perform better on recognition tests than on recall measures (likely explaining why students tend to prefer multiple-choice tests). What leads memories to be inaccurate is that “multiple (incomplete) memory traces are retrieved, combined with current environmental stimuli, and laid over with sense-making cognitive processes to create a ‘recollection’ of what transpired at some point in the past” (Greene & Morgan, 2009, p. 114).

Greene and Morgan (2009) pointed out a number of factors that may affect our memory (and consequently its assessment):

- People find it difficult to accurately remember events.
- Individuals tend to remember pictures better than text.
- Abstract terms (*climate change*) are more difficult to retain than words with visual references (*glacier*).

- People remember the “gist” of a conversation over specific words that occurred.
- Individuals find it easier to remember what someone else said in an interaction in contrast to what they personally said.
- People find it easier to remember when they are in the same state and environment of the event in question (e.g., taking a test in the same room where the information was learned).

Janusik (2007) pointed to an additional problem, noting that listening measures of recall are typically presented in an audio–video format. After being presented with a scene, questions and possible answers are shown on a screen, thus confounding aural and visual cues. Moreover, she argued that recall measures rely too heavily on reading and writing processes. All of these issues should be remembered when attempting to measure and manipulate memory.

Comprehension

A concept closely related to memory but not directly addressed in many listening models is comprehension (i.e., understanding). The lack of explicit recognition in listening models is particularly interesting given the fact that comprehension is often featured in definitions of listening (see Table 1.1 in Chapter 1; for an exception, see the HURIER model). In contrast, as the discussion here will illustrate, the field of L2 learning addresses comprehension, in a variety of forms, quite extensively.

How and when understanding and comprehension occur continue to muddy the research waters for listening scholars. As Bostrom (1990) asked, “What does it mean to ‘understand’ a telephone number?” (p. 6). He also pointed out that we can listen well but not understand immediately—that understanding may actually unfold over time. In addition, it is difficult to disentangle comprehension from inference making, although listening scholars, as illustrated in the Wolff *et al.* (1983; Figure 4.3) and HURIER (Figure 4.4) models, try to do so.

Conceptualization

The emphasis on comprehension was manifest in the majority of measures of listening developed following Nichols’s (1948) original work. As discussed in this chapter, these measures typically focused on short-term recall of information in the classroom context (e.g., Beatty & Payne, 1984; Beighley, 1952; Dow, 1955; McClendon, 1958, Nichols, 1948). The format of these types of measures has remained constant—a multiple-choice test with a single correct answer and multiple incorrect answers.

In a 1966 publication, Brown wrote that some researchers might claim that listening comprehension and reading comprehension are essentially addressing “the ability of comprehension” (p. 416). In many respects, this view reflects the perspective taken by researchers in the field of L2 learning, where comprehension is often related to *structure building* (Sanders & Gernsbacher, 2004). This view focuses on “relating language to concepts in one’s memory and to references in the real world in a way that aims to find coherence and relevance” (Rost, 2011, p. 53). The drive to find coherence also was reported in early listening research by Paris (1975), who found that both comprehension and inference making are part of a constructive process.

It is worth noting that comprehension, from the structure-building perspective, is perceived as an overarching mental process no matter the delivery method (e.g., listening, reading, or observing), whose initial goal is “to build coherent mental representations from concepts" (Rost, 2011, p. 54). This perspective suggests that comprehension requires individuals to build conceptual maps into which incoming information, particularly new information, is placed and connected with information previously integrated into the structure (i.e., held in memory). An important question, however, remains: How do listeners manage their own comprehension?

Measures of metacognitive listening strategies take a different approach toward conceptualizing comprehension. Instead of emphasizing factors affecting comprehension or measuring what individuals can recall, the focus is on listener awareness of, and skill at, regulating their own listening comprehension processes (Goh, 2008). Two metacognitive listening measures are profiled in this book: the Metacognitive Awareness Listening Questionnaire (MALQ; Vandergrift, Goh, Mareschal, & Tafaghodtari, 2006; see Profile 39) and the Metacognitive Listening Strategies Instrument (MLSI; Janusik & Keaton, 2011, 2015; see Profile 40).

Operationalization

Most measures of listening comprehension assess comprehension as a composite of several subskills. For example, the Watson-Barker Listening Test (WBLT; Profile 64) was conceived as a means to measure five facets of adult listening behavior (i.e., interpretation of meaning, interpretation of emotion, understanding, recall, and the ability to follow instructions; Watson & Barker, 1988; Watson, Barker, Roberts, & Roberts, 2001). Some researchers such as Alderson (2000) and Sawaki, Kim, and Gentile (2009) have questioned the viability of separating comprehension into subskills (as done with the BCLCT, KCLT, STEP, and WBLT), particularly because it remains unclear what these skills may be or how to assess them. The focus on subskills also ignores the reality that all of these elements are interrelated and work as part of a greater dynamic process. For instance, recent research has found that watching a speaker's face increases comprehension (Bernstein & Grant, 2009). This argument is especially compelling given that listening scholars have had difficulty replicating the purported factors of many comprehension measures and report that the measures, which purportedly measure the same construct, are essentially unrelated (e.g., the BCLCT, KCLT, STEP, and WBLT; Bodie, Worthington, & Fitch-Hauser, 2011; Fitch-Hauser & Hughes, 1987; Kelly, 1965, 1967; Villaume & Weaver, 1996).

Comprehension has been operationalized in other ways. As noted in this chapter, measures of metacognitive listening strategies address listener awareness of their own listening comprehension processes and their ability to regulate those processes. These processes include self-appraisal and self-regulation (Goh, 2008; Paris & Winograd, 1990). As Janusik describes in Profile 40 of the MSLI, "*Self-appraisal* is recognizing that comprehension is not present, and *self-regulation* is adapting and finding something that will assist with comprehension." Items on the MSLI fall into one of three categories: *problem-solving* (e.g., "I use the words I understand to guess the meaning of the words I don't understand when listening to class lectures and discussions"), *planning-evaluation* (e.g., "As I listen in class, I periodically ask myself if I am satisfied with my level of comprehension"), and *directed attention* (e.g., "I consciously make meaning in my head as I listen to class lectures and discussions").

As noted here, much of the earliest research in listening focused on the academic context, and the interest was mainly on factors affecting lecture comprehension and recall. In addition to assessing memory, the ALSA questionnaire also assesses comprehension. Like metacognitive measures, it addresses self-appraisal items. As

Aryadoust and Goh describe in Profile 1, the ALSA is based on a model of academic listening and "takes into consideration the structure of academic discourse where speaker-related, listener-related, text-related, and situation-related variables play parts in listeners' comprehension processes." ALSA comprehension items address several areas, such as *linguistic components* (e.g., "understand the main ideas and facts of lectures"), *cognitive processing skills* (e.g., "understand simple descriptions given in English about familiar persons, places, and objects by students with the same first language as me"), and *lecture structure* (e.g., "correct my understanding of lectures/tutorials/seminars immediately if my understanding is incorrect").

Interpretation, Inference Making, and Assigning Meaning

Listening scholars often separate comprehension from assigning meaning. In reality, they are entwined; in order to assign meaning, we must first understand the message. Comprehension necessarily involves a variety of types of knowledge: (a) linguistic knowledge (e.g., phonology, lexis, syntax, and semantics) and (b) nonlinguistic knowledge (e.g., general knowledge of the world and its workings). In addition, visual and nonverbal cues may contribute to our interpretation of a message. Listening does, however, require more than prior knowledge. It also requires active inferential processing.

As Merluzzi *et al.* (1981) wrote, our communication does not "contain all the information necessary for effective communication … [people necessarily] fill in the gaps by making inferences and assumptions based on their general knowledge of the world" (pp. 101–102). Several early experiments found that participants could not distinguish between presented stimulus material and the inferences they made when comprehending it (e.g., Baggett, 1975; Bransford, Barclay, & Franks, 1972); moreover, word meaning was found to be affected by linguistic context (Anderson & Ortony, 1975). Thus, speech comprehension is not enough. What we say is affected by how we say it, and listening theory must address this aspect of the listening process (Burleson, 2011).

Brain imaging has identified two systems—mirroring and mentalizing—that may contribute to our understanding of listening (Spunt, 2013). Mirroring appears to be involved in both how we create and how we perceive speech (D'Ausilio *et al.*, 2009; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007), and thus may contribute to successful listening (Berger, 2011). Moreover, when we are asked to attend to how someone is engaging in an action or expressing an emotion, the mirroring system is particularly active (Spunt & Lieberman, 2012) and subsequently may contribute to our ability to be empathic. Of the two systems, mentalizing emphasizes inference making; it addresses "Why?"—why was a message said, and why was it said that way? In other words, it addresses a listener's ability to decode a speaker's mental state and disposition (i.e., goals, motives, beliefs, values, etc.). de Gelder (1987) suggested that this process may be fundamental to conversational interactions. If true, then mentalizing is a key aspect of listening with potential applications in our ability to recognize and understand metaphors and sarcasm, assign meaning to nonverbal communication, and decode ambiguous statements.

Listening processes appear to involve the mentalizing and mirroring systems, with each system operating independently (Spunt, 2013). In most cases, as one system powers up, the other powers down. Because of the energy and concentration needed, in order to mentalize a speaker, we may cut ourselves off from the flood of incoming sensory information. When we do, we run the risk of disrupting our interaction. Spunt

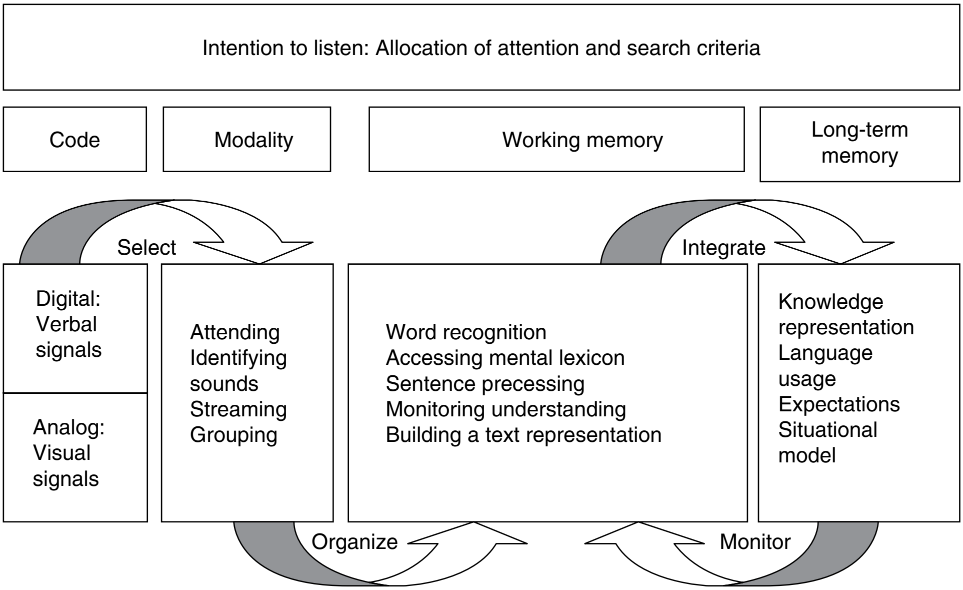


Figure 4.5 Imhof's (2010) model of listening.

and Lieberman (2012) provided evidence for a dual-process model of the two systems, where the mirroring system encodes the "how" of behavior, and mentalizing is activated when we need to understand the "why." These findings provide additional support for proposed dual-process models of listening (see, e.g., Burleson, 2011; Edwards, 2011).

Imhof's (2010) model of listening incorporates elements associated with both of these systems (see Figure 4.5). In her model, she included components associated with decoding, relevance, and the speaker's verbal skills and voice.

Imhof's (2010) model is clearly more comprehensive than the other models we examined thus far (see Figures 4.1–4.4). Unlike previous models, Imhof's is founded in cognitive psychology. Importantly for our discussion, it most closely meets three of Deutsch's (1952) four functions of successful models (organizational, heuristic, predictive, and measurement). As noted at the beginning of the chapter, a successful model must (a) accurately reflect the primary elements of the object under study along with their interrelationships, (b) frame possible hypotheses, and (c) suggest means of measurement. Imhof's model meets these functions in several ways: It clearly identifies and shows the relations across components, suggests areas of study, and suggests how one component might affect other areas (i.e., modality may negatively affect decoding). Like the previous models, however, there is no clear specification of a measurement model.

Another factor that likely contributes to both understanding and memory is schema formation. Studies examining the role of schemata in listening are relatively rare. One exception is a study by Fitch-Hauser (1984), who examined the connection between story recall and individual inference making. Findings from her experiment suggest that listeners use existing schemata to help make sense of what they hear. Importantly, subjects appeared to "fill in the blanks," "remembering" information that was not actually presented to them. Individuals used inference making to reconstruct a story that reflected the schema they held.

Because of the importance of nonverbal cues to inference making or interpretive listening (Bostrom, 2011), measures (and models) of listening should necessarily address nonverbal communication (see, e.g., PONS in Profile 52). This is particularly important given that individuals are less effective at decoding the related vocalic elements of a verbal message (e.g., inflection, tone of voice, and speaking rate) than visual cues (for a review, see Bostrom, 2011). Of course, the role of vocalic cues to listening becomes even more important in an increasingly mediated world, where visual cues may be missing (e.g., in mobile phone calls) or attenuated (e.g., via Skyping or FaceTime).

Conceptualization

Inference and interpretation have been conceptualized in a number of ways. Although it does not identify inferences directly, the Active-Empathic Listening Scale (AELS, Profile 2) (Bodie, 2011; Drollinger, Comer, & Warrington, 2006) does address "sensing." *Sensing* is described as a listener's capacity to understand relational aspects of a message. Likewise, measures designed for comprehension often assess inference making. For example, the WBLT examines individual interpretation of meaning and emotion in contexts such as listening to lectures and interacting in conversations (Watson & Barker, 1988; Watson *et al.*, 2001).

Inference making in conversational contexts is emphasized by a number of researchers. As noted in this chapter, CS (Daly *et al.*, 1987) addresses an individual's level of attention to and understanding of underlying meanings during conversations. More specifically, *interpretation* speaks to an individual's ability to both paraphrase and identify nuances in conversations (e.g., underlying meaning, sarcasm, and irony). Other elements of the CS scale also require the ability to interpret information. For instance, *perceiving affinity* focuses on one's perceived ability to assess the level of liking, attraction, or affiliations between conversational members, whereas *detecting power* addresses the perceived ability to identify power relationships between conversational members.

Interaction involvement (II; Cegala, 1981; Cegala, Savage, Brunner, & Conrad, 1982) addresses inference making somewhat differently, conceptualizing it as three facets of individual conversational involvement: attentiveness, perceptiveness, and responsiveness. *Attentiveness* assesses an individual's perceived awareness of factors affecting an interaction, *perceptiveness* addresses one's perceived understanding of message meanings, and *responsiveness* measures a person's belief that she can respond appropriately to others during an interaction.

Operationalization

Beginning with self-report measures, items comprising the Interaction Involvement Scale (IIS) focus on an individual's perception of his or her own interaction behaviors. For example, Perceptiveness items include "In conversations I am very perceptive to the meaning of my partner's behavior in relation to myself and the situation," as well as the reverse-coded item "In my conversations I often do not accurately perceive others' intentions or motivations" (see IIS, Profile 25). The CS scale has a more expanded view of interpretation and meaning making. Detecting Meaning is operationalized with eight items, among them: "I often hear things in what people are saying that others don't seem to notice" and "I often notice double meanings in conversations." As can be seen in these items, the emphasis is essentially on the ability to "read between the lines" when interacting with others. A similar emphasis is seen in CS Interpretation. This factor consists of three reverse-coded items: "In contrast, I'm usually the last person in a conversation to catch hidden meanings in puns and riddles"; "I often have difficulty paraphrasing what another person said in a conversation"; and "I'm not very good at detecting irony or sarcasm in conversations."

As mentioned, the AELS contains a factor called Sensing, which is measured by the items "I am sensitive to what others are not saying; I am aware of what others imply but do not say," and "I listen for more than just the spoken words." Although it is quite versatile—it can be adapted to self-report, other-report, or observational coding—it has primarily been used in its self-report format. Self-report measures like the AELS come with their own set of problems. As noted in Chapter 2, listening scholars have relied heavily on self-report measures. A review of the measurement profiles in this text will find a variety of cognitive measures relying on self-report. Although many of the common problems associated with this type of measure are identified in Chapters 2 and 5, it is worth noting several factors that should lead us to interpret findings cautiously:

1) People are often unaware of their own cognitive processes because many mental processes operate below their conscious awareness.
2) Social constraints such as social desirability may lead individuals to inaccurately report their activities or motivation.
3) Verbalizing cognitive processes can be difficult, leading subjects to frame responses in a way that they believe is more understandable for the listener.
4) The timing of responses affects findings; retrospective accounts are typically less accurate than reports made in real time.
5) Individuals may not have a clear memory of factors that may have affected their actions and so may make inferences instead of relying on specific memories.

Not surprisingly, self-report measures are not viewed as a particularly reliable means of learning about individual characteristics or behavior, and they are even less effective for learning about cognitive processes, especially when individuals are asked to recall past information.

One means of potentially addressing problems associated with self-report measures is to use a technique introduced by Feinberg and Tanur (1989). They proposed and tested the use of "embedded experiments" within survey designs. In a nutshell, the method allows for randomization, partialing out error variance associated with the subsets of the sample, in what Bostrom (2011) described as an "embedded randomized block design" (p. 23). They noted that a common design is the split-ballot experiment (i.e., split-sample), where alternate questionnaires or procedural variations are randomly assigned to subsets of a sample. This process is qualitatively different from having survey items presented randomly, say, in a computer-administered survey. For example, different scenarios may be used as contextual cues followed by specified scales. Thus, one group may be asked to recall a recent romantic interaction, whereas others are asked to recall an interaction with a family member. Both groups then complete the same survey items. This design allows researchers to test for context-related effects (i.e., priming and schema) and to search for commonalities that extend beyond contexts. Interested readers are directed to Feinberg and Tanur's (1989) article for additional examples and a fuller explanation.

Another way to address issues with self-report measurement is to simply utilize other types of assessment. For example, the WBLT has participants respond to a series of questions following exposure to video-recorded vignettes. Two of the five measured facets are *interpretation of meaning* and *interpretation of emotion* (Watson & Barker, 1988; Watson *et al.*, 2001). Similarly, the TOEFL iBT assesses the ability of L2 learners to infer meaning as they listen. At issue with these types of assessment is that they typically offer multiple options, but score only one option as correct. In reality, our inferences may be partially accurate.

A second means of assessment is the dichotic listening test (DLT). Erhan, Borod, Tenke, and Bruder (1998) utilized DLTs as a means of assessing an individual's ability to identify emotional prosody (i.e., vocal patterns that convey emotions). In their study, participants listened to nonsense syllables, which were presented dichotically, and which reflected multiple emotional categories (e.g., anger, interest, sadness, and happiness). Behavioral assessments included rating participants on accuracy of their interpretation and their response time in identifying the emotion. Electrode montages assessed ERPs (e.g., sustained negativity, late positivity, and slow wave).

A third set of measurement options involves the manipulation of vignettes or some other element of the environment that should cause people to make particular types of inferences. For example, Fitch-Hauser (1984) found that listener inferences could be influenced when different sections of a narrative were manipulated. In some cases, she deleted sections of the narrative associated with the "cause," and in others she deleted the outcome or consequence. In both cases, study participants "filled in" the missing information in order to make a coherent story. Similarly, when she presented listeners with a cause and consequence that were incongruent with one another, participants sought to make sense of the story by supplying a congruent outcome.

New Horizons

The previous review begs the question "Where do we go from here?" The models and measures introduced in this chapter highlight a crucial, underlying issue—how many and what types of subskills should be assessed? As Bostrom (2011) noted, listening scholars continue to disagree on what subskills compose listening comprehension, and that is only one element of a complex, multidimensional communication process. In general, researchers studied the previously discussed areas, often as discrete categories. This approach is problematic as it inhibits integration, synthesis, and theory building.

The call for a greater focus on theory building in the field of listening has grown increasingly forceful in recent years. In a white paper sponsored by the International Listening Association, Bodie, Janusik, and Valikoski (2008) wrote:

> Overall, a theory is a systematic accounting of interrelated phenomena and why their relationships exist. Listening theories are useful to the extent that they aid in understanding the social world (how and why people listen in particular ways and on particular occasions). A primary function of theory, then, is to guide discovery and interpretation of a set of phenomena (Berger & Chaffee, 1987). The more precise and particular a theory, the more precise are the recommendations that the theory advocates (Schutz, 1967). (p. 7)

Studying listening characteristics in isolation is in many ways self-defeating, as the goal is to better understand listening as a holistic process. Thus, the struggle is to continue our study of listening processes, but in a theoretically oriented manner—deriving subskills for testing from strong theories of how people process what they hear. Imhof's

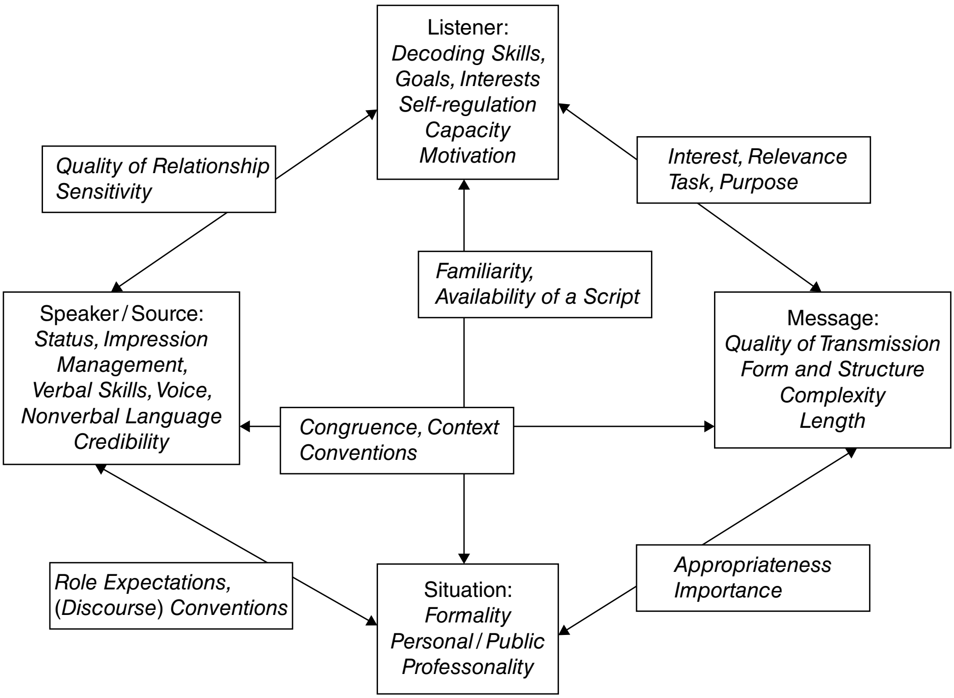


Figure 4.6 Imhof's (2010) mindmap of potential listening variables.

model (Figure 4.5) and mindmap (Figure 4.6) of listening variables, which draw heavily from cognitive psychology, represent a move in the right direction.

Imhof's (2011) mindmap suggests multiple areas of study. Some factors are featured in the models and measures we have discussed in this chapter, whereas others need further study. As she wrote,

> …it needs to be recognized that the process and product of listening depend on the constellation of variables pertaining to the listener, the speaker/the source, the message and the situation, and the mutual interactions. The mindmap can be used both to illustrate effects in listening behavior and to generate hypotheses about causes, effects, and covariation of processes involved in listening. (2011, pp. 109–110)

As seen in Imhof's model and mindmap, cognitive psychology has and continues to contribute to our understanding of listening processes. As noted in this chapter, *metacognition* refers to our ability to monitor our own cognitive processes. It requires us to be aware of our own cognitive resources and our ability to develop a plan based on that awareness or knowledge (e.g., "I won't be able to remember, so I need to take notes"). Individuals who are more metacognitively aware of their listening resources and abilities may be better able to address instances when their listening is challenged. The notion of metacognition has been of particular focus in the field of L2 learning and assessment. Research in this area indicates that students who exhibit greater metacognitive awareness are more successful listeners (Vandergrift, 2003).

Rost (2011) further explored the role of cognitive processing in L2 listening, suggesting two potential areas of study: compensation and transfer. For example, how do individuals address "derailments of attention" (i.e., lack of certainty about what one hears)? Such derailments place stress on the listener and require that the listener, in some way, recover from the attentional lapse that follows (van Herten, Chwilla, & Kolk, 2006). Areas of study include how and when people experience these derailments (certainly not just in L2 situations) as well as how individuals compensate for the inevitable ambiguity and mishearing. Related to inference making, Rost discussed how individuals use their personal schema to guide understanding in L2 contexts (i.e., transfer), noting that the meanings associated with one's native language are another cognitive filter or layer added to listening processes for the nonnative listener. He argued that this transfer process should be considered a strength, not a deficiency, in listening, and that greater metacognitive awareness of their influence can help listeners adjust to the influence of personal schemata on their listening.

Next, fMRI is a promising technique upon which listening theory can be built or at least bolstered. It, and other brain-scanning and brain-mapping techniques, may help reveal the nature of the underlying brain processes responsible for how we interpret and process orally delivered information. Mather and colleagues (2013) outlined four ways fMRI can extend our understanding of cognitive processes:

> First, it can answer questions about which functions can be localized to specific brain regions, questions that are of critical interest for those examining issues related to the modularity of the brain (e.g., Blumstein & Amso, 2013; Cabeza & Moscovitch, 2013; Chiao & Immordino-Yang, 2013).
>
> Second, fMRI data can be used as markers of particular mental processes, allowing insight into what processes are being engaged during different tasks.
>
> Third, fMRI can answer questions about exactly what information is represented in each region of the brain. Such data, for instance, can be used to address theoretical questions about the nature of memory reactivation (e.g., Levy & Wagner, 2013) and working memory (e.g., Reuter-Lorenz, 2013) as well as basic questions about the structure of cognitive processes (e.g., Serences, Ester, Vogel, & Awh, 2009).
>
> Fourth, fMRI can answer questions about whether two tasks engage common or distinct processing mechanisms. This strategy can provide important evidence to address theoretical questions about the nature of tasks (e.g., Rugg & Thompson-Schill, 2013) and how functional circuitry reorganizes with age (e.g., Park & McDonough, 2013). (p. 111)

If you are interested in exploring the contributions of fMRI to your own work, Dimoka (2012) provides an excellent primer on how social science scholars can incorporate fMRI into their research.

In addition to new measurement techniques, researchers assessing listening processes should focus greater attention on meeting statistical considerations. For example, when multidimensional cognitive measures report a single, global score, it may obscure strengths and weakness associated with the construct. Scores also should be discussed in relation to normed data. Unfortunately, global scores are the norm, and normative data are rarely presented. Where appropriate and available, the profiles included in this volume discuss both.

Finally, empirically focused listening scholars may wish to consider emerging areas of statistical analysis. For example, working in the context of L2 listening measures, Aryadoust and Goh (2013) tested fusion modeling (FM) as an option to traditional confirmatory factor analysis (CFA) on tests scored dichotomously (vs. polytomously). When there are fewer than five response categories (e.g., Likert responses), Jöreskog and Sörbom (1981) noted that the fit and function of a polychoic-matrix CFA may be negatively affected. In contrast, the application of FM performed as expected, fitting the data and reflecting the posited four factors. The potential problem of dichotomous scoring has been addressed by other listening researchers. For example, Bodie *et al.* (2011) noted that dichotomous scoring may not fully reflect listening ability and that context may play an important role when choosing to use this type of scoring. Although reliability may improve, validity may be threatened. Thus, the nature and type of statistical analysis should be balanced against the nature and type of the measure utilized.

References

Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Allport, A. (1993). Attention and control: Have we been asking the wrong questions? *Attention and Performance, 14*, 183–218.

Anderson, R. C., & Ortony, A. (1975). On putting apples into bottles: A problem of polysemy. *Cognitive Psychology, 7*, 167–180. doi:10.1016/0010-0285(75)90008-0

Aryadoust, V., & Goh, C. M. (2013). Exploring the relative merits of cognitive diagnostic models and confirmatory factor analysis for assessing listening comprehension. In E. D. Galaczi & C. J. Weir (Eds.), *Exploring language frameworks: Proceedings of the ALTE Kraków conference, July 2011* (pp. 405–426). Cambridge: Cambridge University Press.

Aryadoust, V., Goh, C., & Lee, O. K. (2012). Developing an academic listening self-assessment questionnaire: A study of modeling academic listening. *Psychological Test and Assessment Modeling, 54*, 227–256.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417–423. doi:10.1016/S1364-6613(00)01538-2

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*, 189–208. doi:10.1016/S0021-9924(03)00019-4

Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Baggett, P. (1975). Memory for explicit and implicit information in picture stories. *Journal of Verbal Learning and Verbal Behavior, 14*, 538–548. doi:10.1016/S0022-5371(75)80031-4

Beatty, M. J., & Payne, S. (1984). Loquacity and quantity of constructs as predictors of social perspective-taking. *Communication Quarterly, 32*, 207–210. doi:10.1080/01463378409369553

Beighley, K. (1952). The effect of four speech variables on listener comprehension. *Speech Monographs, 19*, 249–258. doi:10.1080/03637755209375068

Benoit, P. J., & Benoit, W. L. (1988). Conversational memory employing cued and free recall. *Central States Speech Journal, 39*, 18–27. doi:10.1080/10510978809363231

Berger, C. R. (2011). Listening is for acting. *International Journal of Listening, 25*, 104–110. doi:10.1080/10904018.2011.536477

Berger, C. R., & Chaffee, S. H. (1987). The study of communication as a science. In C. R. Berger & S. H. Chaffee (Eds.), *Handbook of communication science* (pp. 15–19). Newbury Park, CA: Sage.

Bernstein, J. G. W., & Grant, K. W. (2009). Audio and audiovisual speech intelligibility in fluctuating maskers by normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 125*, 3358–3372. doi:10.1121/1.3110132

Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews Neuroscience, 14*, 693–707. doi:10.1038/nrn3565

Blumstein, S. E., & Amso, D. (2013). Dynamic functional organization of language: Insights from functional neuroimaging. *Perspectives on Psychological Science, 8*, 44–48. doi:10.1177/1745691612469021

Bodie, G. D. (2011). The Active-Empathic Listening Scale (AELS): Conceptualization and evidence of validity within the interpersonal domain. *Communication Quarterly, 59*, 277–295. doi:10.1080/01463373.2011.583495

Bodie, G. D., Janusik, L. A., & Valikoski, T.-R. (2008). *Priorities of listening research: Four interrelated initiatives*. [A white paper sponsored by the Research Committee of the International Listening Association.] Retrieved from www.listen.org

Bodie, G. D., Worthington, D. L., & Fitch-Hauser, M. (2011). A comparison of four measurement models for the Watson-Barker Listening Test (WBLT)-Form C. *Communication Research Reports, 28*, 32–42. doi:10.1080/08824096.2011.540547

Bostrom, R. N. (1990). *Listening behavior: Measurement and application*. New York: Guilford.

Bostrom, R. N. (2011). Rethinking conceptual approaches to the study of listening. *International Journal of Listening, 25*(1), 10–26. doi:10.1080/10904018.2011.536467

Bostrom, R. N., & Bryant, C. (1980). Factors in the retention of information presented orally: The role of short-term memory. *Western Speech Communication Journal, 44*, 137–145. doi:10.1080/10570318009373999

Bostrom, R. N., & Waldhart, E. S. (1980). Components in listening behavior: The role of short-term memory. *Human Communication Research, 6*, 221–222. doi:10.1111/j.1468-2958.1980.tb00142.x

Bostrom, R. N., & Waldhart, E. S. (1983). *Kentucky Comprehensive Listening Skills Test*. Lexington, KY: Kentucky Listening Research Center.

Bostrom, R. N., & Waldhart, E. S. (1988). Memory models and the measurement of listening. *Communication Education, 37*, 1–18. doi:10.1080/03634528809378699

Bouskila-Yam, O., & Kluger, A. N. (2011, December). The Facilitating Listening Scale (FLS). In A. N. Kluger (Chair), *Listening*. Symposium conducted at the 1st Israel Organizational Behavior Conference, Tel Aviv, Israel.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology, 3*, 193–209. doi:10.1016/0010-0285(72)90003-5

Brown, J. I. (1966). Establishing the validity of a listening tests. In S. Duker (Ed.), *Listening: Readings* (pp. 416–429). New York: Scarecrow Press.

Brown, J. I., & Carlsen, G. R. (1955). *Brown-Carlsen Listening Comprehension Test*. New York: Harcourt, Brace & World.

Brownell, J. (1986). *Building active listening skills*. Englewood Cliffs, NJ: Prentice Hall.

Brownell, J. (1996). *Listening: Attitudes, principles, and skills* (2nd ed.). Boston: Allyn & Bacon.

Brownell, J. (2006). *Listening: Attitudes, principles, and skills* (3rd ed.) Boston: Pearson.
Buračas, G. T., Saenz, M. T., & Boynton, G. M. (2003). Dual nature of visual attention. In G. T. Buračas, O. Rukšėnas, G. M. Boynton, & T. D. Albright (Eds.), *Modulation of neuronal responses: Implications for active vision* (Ch. 11). Amsterdam: IOS Press.
Burleson, B. R. (2011). A constructivist approach to listening. *International Journal of Listening, 25*, 27–46. doi:10.1080/10904018.2011.536470
Cabeza, R., & Moscovitch, M. (2013). Memory systems, processing modes, and components: Functional neuroimaging evidence. *Perspectives on Psychological Science, 8*, 49–55. doi:10.1177/1745691612469033
Cegala, D. J. (1981). Interaction involvement: A cognitive dimension of communicative competence. *Communication Education, 30*, 109–121. doi:10.1080/03634528109378460
Cegala, D. J., Savage, G. T., Brunner, C. C., & Conrad, A. B. (1982). An elaboration of the meaning of interaction involvement: Toward the development of a theoretical concept. *Communication Monographs, 49*, 229–248. doi:10.1080/03637758209376087
Chiao, J. Y., & Immordino-Yang, M. H. (2013). Modularity and the cultural mind: Contributions of cultural neuroscience to cognitive theory. *Perspectives on Psychological Science, 8*, 56–61. doi:10.1177/1745691612469032
Cohen, R. A. (1993). *The neuropsychology of attention*. New York: Plenum Press.
Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford: Oxford University Press.
Daly, J. A., Vangelisti, A. L., & Daughton, S. M. (1987). The nature and correlates of conversational sensitivity. *Human Communication Research, 14*, 167–202. doi:10.1111/j.1468-2958.1987.tb00126.x
D'Ausilio, A., Pulvermuller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, *19*, 381–385.
de Gelder, B. (1987). On not having a theory of mind. *Cognition*, *27*, 285–290. doi:10.1016/S0010-0277(87)80013-6
Deutsch, K. W. (1952). On communication models in the social sciences. *Public Opinion Quarterly, 16*, 356–380. doi:10.1086/266399
Dimoka, A. (2012). How to conduct a functional magnetic resonance (fMRI) study in social science research. *MIS Quarterly, 36*, 811–840. Retrieved from http://ssrn.com/abstract=2428901
Dow, C. W. (1955). Listening instruction at Michigan State: 1954–55. *Journal of Communication, 5*, 110–112. doi:10.1111/j.1460-2466.1955.tb01107.x
Drollinger, T., Comer, L., & Warrington, P. (2006). Development and validation of the active empathetic listening scale. *Psychology & Marketing, 23*, 161–180. doi:10.1002/mar.20105
Educational Testing Services (ETS). (1957). *Sequential Tests of Educational Progress*. Princeton, NJ: Author.
Edwards, R. (2011). Listening and message interpretation. *International Journal of Listening, 25*, 47–65. doi:10.1080/10904018.2011.536471
Erhan, H., Borod, J. C., Tenke, C. E., & Bruder, G. E. (1998). Identification of emotion in a dichotic listening task: Event-related brain potential and behavioral findings. *Brain and Cognition, 37*, 287–307. doi:10.1006/brcg.1998.0984
Feinberg, S. E., & Tanur, J. M. (1989). Combining cognitive and statistical approaches to survey design. *Science, 243*, 1017–1023. doi:10.1126/science.243.4894.1017
Fitch-Hauser, M. (1984). Message structure, inference making, and recall. In R. N. Bostrom (Ed.), *Communication Yearbook 8* (pp. 378–392). Beverly Hills, CA: Sage.

Fitch-Hauser, M., & Hughes, M. A. (1987, March). A factor analytic study of four listening tests. Paper presented at the meeting of the International Listening Association, New Orleans, LA.

Goh, C. M. (2008). Metacognitive instruction for second language listening development: Theory, practice and research implications. *RELC Journal, 39*, 188–213.

Greene, J. O., & Morgan, M. (2009). Cognition and information processing. In W. F. Eadie (Ed.), *21st century communication: A reference handbook* (Vol. 1, pp. 110–118). Thousand Oaks, CA: Sage.

Gross, G. G., Junge, J. A., Mora, R. J., Kwon, H-B., Olson, C. A., Takahashi, T. T., … & Arnold, D. B. (2013). Recombinant probes for visualizing endogenous synaptic proteins in living neurons. *Neuron, 78*, 971–985. doi:10.1016/j.neuron.2013.04.017

Hodges, J. R. (1994). *Cognitive assessment for clinicians*. Oxford: Oxford University Press.

Imhof, M. (2010). What is going on in the mind of a listener? The cognitive psychology of listening. In A. D. Wolvin (Ed.), *Listening and human communication in the 21st century* (pp. 97–126). West Sussex, UK: Wiley.

Isen, A. M., Shalker, T. F., Clark, M., & Karp, N. P. (1978). Affect, accessibility of material in memory, and behavior: A cognitive loop? *Journal of Personality and Social Psychology, 36*, 1–12. doi:10.1037/0022-3514.36.1.1

Jäncke, L., Specht, K., Shah, J. N., & Hugdahl, K. (2003). Focused attention in a simple dichotic listening task: An fMRI experiment. *Cognitive Brain Research, 16*, 257–266. doi:10.1016/S0926-6410(02)00281-1

Janusik, L. A. (2007). Building listening theory: The validation of the Conversational Listening Span. *Communication Studies, 58*, 139–156. doi:10.1080/10510970701341089

Janusik, L. A. (2010). Listening pedagogy. Where do we go from here? In A. D. Wolvin (Ed.), *Listening and human communication: 21st century perspectives*. Oxford: Blackwell.

Janusik. L. A., & Keaton, S. (2011). Listening metacognitions: Another key to teaching listening? *Listening Education*, *2*, 35–45.

Janusik, L. A., & Keaton, S. (2015). Toward developing a cross-cultural metacognition instrument for listening in first language (L1) contexts: The (Janusik-Keaton) Metacognitive Listening Instrument. *Journal of Intercultural Communication Research, 44*, 288–306. doi:10.1080/17475759.2015.1073169

Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V: Analysis of linear structural relationships by the method of maximum likelihood*. Chicago: National Educational Resources.

Kelly, C. M. (1965). An investigation of the construct validity of two commercially published listening tests. *Speech Monographs, 32*, 139–143. doi:10.1080/03637756509375443

Kelly, C. M. (1967). Listening: Complex of activities—and a unitary skill? *Speech Monographs, 34*, 455–465. doi:10.1080/03637756709375556

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party." *Journal of Neuroscience, 30*, 620–628. doi:10.1523/jneurosci.3631-09.2010

Lang, A., & Basil, M. D. (1998). Attention, resource allocation, and communication research: What do secondary task reaction times measure anyway? In *Communication yearbook* (Vol. 21, pp. 443–470). Beverly Hills, CA: Sage.

Levy, B. J., & Wagner, A. D. (2013). Measuring memory reactivation with functional MRI: Implications for psychological theory. *Perspectives on Psychological Science, 8*, 72–78. doi:10.1177/1745691612469031

Mather, M., Cacioppo, J. T., & Kanwisher, N. (2013). How fMRI can inform cognitive theories. *Perspectives on Psychological Science, 8*, 108–113. doi:10.1177/1745691612469037

McClendon, P. I. (1958). An experimental study of the relationship between the note-taking practices and listening comprehension of college freshmen during expository lectures. *Speech Monographs, 25*, 222–228. doi:10.1080/03637755809375236

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology, 17*, 1692–1696. doi:10.1016/j.cub.2007.08.064

Menenti, L., Gierhan, S. M. E., Segaert, K., & Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science: A Journal of the American Psychological Society/APS, 22*, 1173–1182. doi:10.1177/0956797611418347

Merluzzi, T. V., Rudy, T. E., & Glass, C. R. (1981). The information-processing paradigm: Implications for clinical science. In T. V. Merluzzi, C. R. Glass, & M. Genest (Eds.), *Cognitive Assessment* (pp. 77–124). New York: Guilford.

Mills, E. P. (1974). *Listening: Key to communication*. New York: Petrocelli.

Neath, I., & Surprenant, A. M. (2003). *Human memory: An introduction to research, data, and theory* (2nd ed.). Belmont, CA: Wadsworth.

Nichols, R. (1948). Factors in listening comprehension. *Speech Monographs, 15*, 154–163. doi:10.1080/03637754809374953

Nichols, R. G. (1974). Listening: Questions and problems. *Quarterly Journal of Speech, 33*, 83–86. doi:10.1080/00335634709381268

Paris, S. G. (1975). Integration and inference in children's comprehension and memory. In F. Restle, R. Shiffrin, J. Castellan, J. Lindman, & D. Pisoni (Eds.), *Cognitive theory* (Vol 1, pp. 223–246). Hillsdale, NJ: Erlbaum.

Paris, S. G., & Winograd, P. W. (1990). How metacognition can promote academic learning and instruction. In B. J. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 15–51). Hillsdale, NJ: Erlbaum.

Park, D. C., & McDonough, I. M. (2013). The dynamic aging mind: Revelations from functional neuroimaging research. *Perspectives on Psychological Science, 8*, 62–67. doi:10.1177/1745691612469034

Reed, K., Goolsby, J. R., & Johnston, M. K. (2014). Extracting meaning and relevance from work: The potential connection between the listening environment and employee's organizational identification and commitment. *International Journal of Business Communication, 53*, 1–17. doi:10.1177/2329488414525465

Reuter-Lorenz, P. A. (2013). Aging and cognitive neuroimaging: A fertile union. *Perspectives on Psychological Science, 8*, 68–71. doi:10.1177/1745691612469023

Ronnberg, J., Rudner, M., Foo, C., & Lunner T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology, 47*, S171–S177. doi:10.1080/14992020802301167

Rost, M. (2011). *Teaching and researching listening* (2nd ed.). Harlow: Longman.

Rugg, M. D., & Thompson-Schill, S. L. (2013). Moving forward with fMRI data. *Perspectives on Psychological Science, 8*, 84–87.

Sanders, T. J. M., & Gernsbacher, M. A. (2004). Accessibility in text and discourse processing. *Discourse Processes, 37*, 79–89. doi:10.1207/s15326950dp3702_1

Sawaki, Y., Kim, H-J., & Gentile, C. (2009). Q-matrix construction: Defining the link between constructs and test items in large-scale reading and listening comprehension assessment. *Language Assessment Quarterly, 6*, 190–209. doi:10.1080/15434300902801917

Schutz, A. (1967). Concept and theory formation in the social sciences. In M. Natanson (Ed.), *Collected papers of Alfred Schutz: Vol. 1. The problem of social reality* (pp. 48–66). The Hague: Martinus Nijoff.

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science, 20*, 20–214.

Sillars, A. L., Weisberg, J., Burggraf, C. S., & Zietlow, P. H. (1990). Communication and understanding revisited: Married couples' understanding and recall of conversations. *Communication Research, 17*, 500–522. doi:10.1177/009365090017004006

Speaks, C. E. (1988). Statistical properties of dichotic listening scores. In K. Hugdahl (Ed.), *Handbook of dichotic listening: Theory, methods and research* (pp. 185–213). New York: Wiley.

Spunt, R. P. (2013). Mirroring, mentalizing, and the social neuroscience of listening. *International Journal of Listening, 27*, 61–72. doi:10.1080/10904018.2012.756331

Spunt, R. P., & Lieberman, M. D. (2012). Dissociating modality-specific and supramodal neural systems for action understanding. *Journal of Neuroscience, 32*, 3575–3583. doi:10.1523/JNEUROSCI.5715-11.2012

Stafford, L., & Daly, J. A. (1984). Conversational memory. *Human Communication Research, 10*, 379–402. doi:10.1111/j.1468-2958.1984.tb00024.x

Taylor, S. E. (1964). *Listening: What research says to the teacher (No. 29)*. Washington, DC: National Education Association. Retrieved from http://eric.ed.gov/?id=ED026120

Teasdale, J. D., & Fogarty, S. J. (1979). Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology, 88*, 248–257. doi:http://dx.doi.org/10.1037/0021--843X.88.3.248

Vandergrift, L. (2003). Orchestrating strategy use: Toward a model of the skilled language listener. *Language Learning, 53*, 463–496. doi:10.1111/1467-9922.00232

Vandergrift, L., Goh, C. C., Mareschal, C. J., & Tafaghodtari, M. H. (2006). The metacognitive awareness listening questionnaire: Development and validation. *Language Learning, 56*, 431–462. doi:10.1111/j.1467-9922.2006.00373.x

Van Herten, M., Chwilla, D.J., & Kolk, H.H. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience, 18*, 1181–1197. doi:10.1162/jocn.2006.18.7.1181

Villaume, W. A., & Weaver, J. B., III (1996). A factorial approach to establishing reliable listening measures from the WBLT and the DCLT: Full information factor analysis of dichotomous data. *International Journal of Listening, 10*, 1–20. doi:10.1207/s1932586xijl1001_1

Waldhart, E., & Bostrom, R. (1981, March). Note-taking and listening skills. Paper presented at the annual meeting of the International Listening Association, Washington, DC.

Watson, K. W., & Barker, L. L. (1983). *Watson-Barker listening test*. Tega Cay, SC: SPECTRA.

Watson, K. W., & Barker, L. L. (1988). Listening assessment: The Watson-Barker listening test. *International Journal of Listening, 2*, 20–31. doi:10.1080/10904018.1988.10499095

Watson, K. W., Barker, L. L, Roberts, C. V., & Roberts, J. D. (2001). *Watson-Barker listening test: Video version: Facilitator's guide*. Sautee, GA: SPECTRA.

Wolff, F., Marsnik, N., Tacey, W. S., & Nichols, R. G. (1983). *Perceptive listening.* New York: Holt, Rinehart & Winston.

Wolvin, A. D. (1989). Models of the listening process. In C. W. Roberts & K. W. Watson (Eds.), *Intrapersonal communication processes: Original essays* (pp. 508–527). New Orleans, LA: SPECTRA.

5

Measuring Affective Components of Listening

Graham D. Bodie and Susanne M. Jones

Listening competence is a function of three interrelated components: knowledge of listening, the motivation to listen, and appropriate and/or effective performance when listening (Coakley, Halone, & Wolvin, 1996; Halone, Cunconan, Coakley, & Wolvin, 1998; Halone, Wolvin, & Coakley, 1997). The second of these components, *motivation*, is most often defined as an "attitudinal component—the willingness to engage as a communicating listener" (Wolvin & Coakley, 1994, p. 151). Motivation characterizes the affective relationship between speaker and listener (Cronen & Price, 1976), regulating whether people approach or avoid one another (Elliot, Eder, & Harmon-Jones, 2013; Gollwitzer, Fujita, & Oettingen, 2004; Weiner, 1992).

The need for listeners to develop a "positive listening attitude" is pervasive in current definitions of effective listening (see Chapter 1), and the importance of such an attitude is stressed in listening instruction across K–12 (Cooper, 1998) and higher education (Wolvin & Coakley, 2000) alike. Most textbooks stress the importance of "knowing why you are listening" and being aware of listening-related goals and priorities (Brownell, 2013; Worthington & Fitch-Hauser, 2012). In particular, students are taught to take responsibility as a listener, to "attend to others with an open mind," and to recognize "that listening is an active not passive activity" (Wolvin, 2009, p. 137). Each piece of advice taps some facet of the affective dimension of listening with the general assumption that sheer knowledge is not enough to listen well: The genuine desire to listen effectively is equally important.

In addition to being a popular component of teaching listening, affective components of listening also are popular in the academic literature. Keaton and Bodie (2013) reported that 80 out of 110 studies published in the *International Journal of Listening* (*IJL*) between 1987 and 2011 (nearly 75%) examined one or more facets of motivations to or tendencies toward listening. Mirroring work in the social sciences more generally, most of these studies ($n = 67$, 61%) asked participants to report on their own attitudes, motivations, or perceived tendencies; the remaining ($n = 13$, 39%) asked participants to report on another person (e.g., a friend, coworker, or spouse). As a result, much of what we know about listening is limited to what people report about their own listening (retrospective self-report) and how this self-knowledge aligns with (or diverges from) what other people report (retrospective other-report).

This chapter begins by discussing the strengths and limitations of self-reporting methods. After reviewing how self-reporting methods have been used (and abused) in the listening literature, we outline a set of recommendations for proper use.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

A subsequent section discusses several popular measures of affective components to illustrate how research employing them can best be interpreted. The final section, "New Horizons," details how listening scholars can measure affective elements of listening in more nuanced and potentially powerful ways.

Self-Report Measurement

Listening researchers have two primary sources of information at their disposal from which they can make claims about listening: observation and reporting. Methods appropriate for capturing listening behaviors are covered in Chapters 3 and 6. This chapter deals with affective data that capture internal states reported by research participants, which is referred to as *retrospective self-report* or *self-reporting*. Although self-reports also are used when asking people to judge others (e.g., "How well do you think your friend listens?"), we focus here on utilizing self-reports to capture one's own perceptions.

Participants can report on any number of listening variables, including "recollected behavior, experience, events, and affect, as well as global assessments of affective/psychological states and typical behavior based on accumulations of previous experience and knowledge" (Metts, Sprecher, & Cupach, 1991, p. 162). Each type of reporting method has been used to make claims about listening, although a significant portion of this work has focused on discovering trait-like dispositions thought to affect how people behave as listeners. Traits or dispositions are "characteristics of people that are relatively stable across time and situations" (Hoyle & Leary, 2009, p. 12); situational fluctuations or more transitory reactions are captured with state measures. In many cases, trait and state measures differ only on the basis of the specified time frame. Stable (trait) tendencies are captured by asking participants to evaluate "in general" (e.g., "In general, how well do you listen?"). Situational (state) listening measures refer participants to specific situations and times (e.g., "How well did you listen to your partner in the last conversation?").

The earliest self-report measure of a listening trait was the Receiver Apprehension Test (RAT). Wheeless (1975) originally defined Receiver Apprehension (RA) as "fear of misinterpreting, inadequately processing, and/or not being able to adjust psychologically to messages sent by others" (p. 263). RA has since evolved into a construct called Informational Reception Apprehension (IRA), a three-dimensional construct related to an individual's anxiety regarding: (a) listening, (b) reading, and (c) thinking about abstract concepts (Wheeless, Preiss, & Gayle, 1997). We use the Informational Reception Apprehension Test (IRAT) in this chapter to illustrate the advantages and disadvantages of self-reports of listening (see IRAT, Profile 27).

Advantages and Disadvantages of Using Self-Reports

As seen in Table 5.1, each item of the listening subscale was written to reflect how the respondent typically feels while listening. Only the last item ("I have avoided listening to abstract ideas because I was afraid I could not make sense of what was said") requires the respondent to explicitly think about a past behavior (something he or she did); all other items reference experiences or feelings during the interaction. These items do *not*

Table 5.1 Items comprising the IRAT-Listening Subscale.

While listening, I get nervous when a lot of information is given at once.
I get impatient and anxious when listening to someone discuss theoretical, intellectual issues.
I feel agitated or uneasy when someone tells me there is not necessarily a clear, concrete way to deal with an important problem.
While listening, I feel tense when I have to analyze feelings carefully.
When I hear abstract material, I am afraid I will be unable to remember it very well.
It is frustrating to listen to people discuss practical problems in philosophical and abstract ways.
Many classes are annoying and uncomfortable because the teacher floods you with detailed information in the lectures.
I experience anxiety when listening to complex ideas others tell me.
When I listen to complicated information, I often fear that I will misinterpret it.
I feel relaxed and confident while listening, even when a lot of information is given at once. (R)
Listening to complex ideas is a pleasant, enjoyable experience for me. (R)
When listening, I feel relaxed and confident that I can remember abstract ideas that are being explained. (R)
I have avoided listening to abstract ideas because I was afraid I could not make sense of what was said.

Notes: Scale items reprinted with permission. Items marked with (R) are reverse-coded prior to computing an individual's listening anxiety score. For more information about how to score listening anxiety, see IRAT Profile, 27.
Source: Wheeless *et al.* (1997).

capture observable behavior that can be easily witnessed and reported on by another person (i.e., a trained rater or a participant reporting on listening behavior). These items *do* capture affective components, elements of listening that are likely best reported by the individual listener. Of course, quite a few of the IRAT items capture emotions that may be "visible" in a person's behaviors, such as blushing or shaking, which would then make them observable. Emotional experiences are not, however, isomorphic with emotional expressions (Ekman & Friesen, 1975).

The actual experience of nervousness is an internal state known only to the person who experiences it. As McCroskey (1997) argued, "self-report measures … are most appropriate when they are directed toward matters of affect and/or perceptions in circumstances … [and] least useful when they are directed toward matters of fact … unknown or unknowable by the respondent" (p. 196). We recommend that research questions measuring some form of listening motivation utilize self-report scales. Motivational states are affective in nature, and internal states are best assessed by those who experience these states.

In addition to capturing information that may not be readily available from other methods, there are several other reasons why a listening researcher may want to utilize a self-report measure. First, self-report measures are easy to administer and generally inexpensive, although some scales are not free of charge (e.g., the Doctors' Interpersonal Skills Questionnaire, Profile 15). Researchers who have access to online software such as Qualtrics or SurveyMonkey can populate a survey within minutes of Institutional

Review Board (IRB) approval. Moreover, because most self-report research is exempt from full review by university IRBs, they reduce time-consuming paperwork and meetings in comparison with designs that are more intrusive (e.g., deception studies). Second, when administered online, researchers are able to capture data from captive (e.g., college students and paid survey panelists) and noncaptive (e.g., genuinely interested constituents) samples by posting the URL on bulletin boards, listservs, or mobile data collection labs. When utilizing college students, online administration can be advantageous because students may be more likely to participate if they are not required to report to a laboratory at a predetermined time of day. If your research requires non–college student participants, online administration may be the only logically feasible way to collect data. Third, when properly developed, self-report scales often produce scores with high levels of reliability.

Although advantageous for many reasons, there are limitations to the use of self-report measures. First, self-reporting relies on respondent honesty, an issue that varies as a function of the level of stigma, taboo, or sensitivity of the topic. The desire to create a favorable impression might be more pronounced for some participants or in some contexts compared to others (e.g., when listening assessments are part of performance appraisals). Social desirability of a socially positive and acceptable behavior like listening can cause people to overreport tendencies or motivations (Crowne & Marlowe, 1960; Lawson & Winkelman, 2003). For example, students might be honest about their willingness to listen in a classroom setting when asked by a researcher who is distributing an anonymous and confidential survey outside of that setting. If asked these same questions by their instructor, however, students may feel pressure to answer in a way that gives the impression they are attentive to class material even if the survey guarantees complete anonymity. If social desirability varies significantly as a function of group membership or the situation, then the researcher is faced with a systematic source of error (response bias) that is difficult to manage. Even in research contexts that do not seem prone to social desirability effects, respondents may be over- or underreporting; that is, their perceptions and thus their judgments of their own internal states are influenced by desires (e.g., "I *want* to be more sensitive to what others say") and culturally scripted norms and expectations (e.g., "I *should* be more sensitive to what others say"), rather than their "true" internal tendencies (e.g., "I *am* more sensitive to what others say"). Of course, it is always the case that other biases or moods cloud judgments of one's own internal states. In fact, people cannot introspectively assess themselves in a completely accurate manner; that is, participants may be less self-aware than researchers assume (Nisbett & Wilson, 1977). But the social desirability bias adds just another source of systematic error.

Second, even when social desirability is not a major concern, participants may not have the ability to provide accurate judgments about their internal states. Measures that assess facets of attitudes or motivation that participants have not previously considered are particularly prone to this problem, as are measures designed to assess retrospectively recalled attitudes (e.g., "How nervous were you last week?"). Unfortunately, participants readily provide judgments for attitudes, feelings, and past experiences they have never experienced or pondered, a problem further compounded by online surveys that often force participants to respond to all scale items. Most surveys do not include a scaling option for "I have never considered that" or "I do not know." Even if these options were available, some may be reticent to choose them for fear of looking incompetent or otherwise unable to make a decision.

When asking people to report on behaviors, they may not be aware of how they typically act or how they did act, "especially if they are asked about experiences from the distant past or things that are not very salient" (Metts *et al.*, 1991, p. 168). The ability to remember specific details of conversations or events pales in comparison with our tendency to remember the gist of situations or people (Stafford, Burggraf, & Sharkey, 1987; Stafford & Daly, 1984; see Profile 38, Memory for Conversation, as well as Chapter 4). Indeed, researchers tend to overestimate participants' abilities to report accurately on enacted behavior (e.g., the amount of eye contact, or the number of open-ended questions asked).

In our view, it is valid to use self-report scales that capture some perspectives or subjective recall of experiences ("impressionistic") (Metts *et al.*, 1991). When, however, participants are asked to retrospectively recall behaviors, chances are that their recollections will be biased by, for instance, current mood or unique experiences (Miell, 1987). Work showing that reports of behavior are not significantly correlated with actual enactments of listening (Bodie, Jones, Vickery, Hatcher, & Cannava, 2014) suggests that when respondents are asked to report on their own listening competence, scores should be interpreted as (a) *motivations* for or *attitudes* toward taking an active view of listening or (b) an individual's *beliefs* about his or her competence. A researcher using self-report scales of listening should refrain from labeling these kinds of data as reflecting actual behavior (see Chapter 6 and below for further discussion of this issue).

Third, even when respondents are able to report behavior accurately (e.g., they truly can remember, recall, and/or introspect appropriately), they might be unable to understand or interpret particular items. The ability to comprehend items is more likely to affect the measurement of abstract concepts, such as "feeling insecure while listening" as opposed to "my heart beat fast when I had to listen to that lecture" (see Table 5.1). Because it is abstract and uniquely tied to personal experience and personality, concepts like "insecurity" conjure up multiple meanings. When constructing survey items, be cognizant of writing clear directions and using concrete, familiar, and unambiguous language (see Chapter 2).

It is nearly impossible to ensure that all respondents interpret all questions in the exact same way, which is one reason to develop scales cumulatively and continue to assess their psychometric profile. In addition, most scales that assess affective listening components use ordinal-level scales (e.g., from *strongly agree* to *strongly disagree*). Just as respondents interpret items differently, they also interpret and use scales differently; what Person A rates as 4 (*agree*) might be rated as 5 (*strongly agree*) by another person, not because they have different opinions but because they interpret the meanings of scale points differently. Differences in scale point interpretations produce different scores that reflect a source of systematic error and call for questioning the validity potential of the scale. Some work suggests that scale use is a function of one or more personality traits (e.g., Austin, Deary, Gibson, McGregor, & Dent, 1998).

Finally, with the proliferation of online survey software, researchers are faced with choices regarding ease of data collection and a need to balance convenience with control over the survey environment. Although online technologies allow research teams to collect larger, more diverse samples in an efficient manner, these technologies also introduce concerns about the validity and representativeness of data. For instance, researchers have to assume participants are paying adequate attention when completing an online survey. One way to test whether participants are paying adequate attention is to introduce questions that are randomly distributed throughout core items of a specific scale. Including an item such as "Please click the radio button under the number 4 for this item" in the middle of a survey allows you to remove respondents who clicked any other button.

Conclusion

Numerous self-report measures of listening are profiled in this book, and there is good reason to include one or more of them in your research. Self-reporting is viable whenever you would like to assess one or more affective components of listening, such as motivation, attitudes toward listening, and beliefs about listening. When there is a need to design an instrument to measure a currently unavailable component of listening, we direct you to the general guidelines presented in Chapter 2. We will return to an overview of specific measures toward the end of this chapter. For now, we turn our attention to common uses and abuses of self-report methods in listening research.

The Uses and Abuses of Self-Report Methods in Listening Research

The purpose of this section is to provide an overview of how self-report methods have been used in listening research. We first detail some general tendencies of self-report listening research. A second section then offers a few suggestions for future improvement.

Some General Patterns in Self-Report Listening Research

Our first observation about self-report listening research is that it tends to be cross-sectional in nature; that is, participants are asked to answer several self-report items during a single data collection session. Although cross-sectional, self-report research is useful for many purposes. We, however, must recognize several methodological issues. First, the only conclusions you can draw from cross-sectional research are correlational, rather than causal, in nature. If you want to test cause–effect relationships, you need to show that (a) the cause temporally precedes the effect, (b) both cause and effect are related with one another in meaningful ways, and (c) there are no other variables that could have plausibly caused the effect (Shadish, Cook, & Campbell, 2002). Causal relationships are most difficult to establish in the social sciences, including communication sciences, because there usually are many causes that influence effects. Work on listening anxiety has found that this trait correlates with reports of verbal aggression (Schrodt & Wheeless, 2001) and a variety of other trait-like personality and communication variables (e.g., Ledbetter & Schrodt, 2008; Schrodt, Wheeless, & Ptacek, 2000). What we are left to speculate about, however, is whether listening anxiety is *caused* by these traits, whether these traits are caused by anxious dispositions, or whether some third variable explains the association between anxiety scores and other trait-like variables. Discovering what causes an individual to experience listening anxiety would have great theoretical and practical payoff, yet a definitive answer is impossible by relying on cross-sectional data. Questions of causality are best answered with experimental research or longitudinal studies that track a set of individuals over time.

Our second observation is that self-report scales tend to generate quite a bit of error. Measurement error can be divided into two types: random error and systematic error. *Random error* contains any factor that randomly affects the measurement of a variable across a sample. For instance, some participants in a study on listening anxiety might experience brief lapses of attention or fatigue that interferes with responding. Fatigue artificially inflates or deflates participant scores but does not affect all participants in the same way (i.e., it is not systematic). To put it in statistical terms, random error affects variability around the average of a sample, but it does not influence the average itself.

Systematic error, on the other hand, *does* affect the sample average; as a result, differences among groups can be a hidden function of systematic error rather than of true group differences. *Systematic error* consists of any factor that systematically affects the measurement of a variable across a sample. For instance, a researcher may administer a survey when classes end and when there is considerable noise, whereas other administrations might be done when there is no external noise that interferes with survey participants. Because entire sets (i.e., an entire class) of participants are affected by noise in a similar manner, in this example external noise is a source of systematic error.

Whereas random error, by nature, is impossible to predict and control, there are several strategies one can take to limit systematic error. Consistency in testing environment and instructions is crucial, especially as the number of people collecting data increases; proper training is a must. If there are known influences on a key variable, one way to limit systematic error is to either control those influences (e.g., collect data at multiple times of day) or measure those influences (e.g., track the time the survey is taken); statistical procedures can then be used to test for any systematic error from those influences. One potentially dangerous and yet not highly recognized source of systematic error is known as *common method variance* (CMV). CMV occurs when shared variance among scores is a function of the method used to collect data rather than the constructs the measures represent (Podsakoff & Organ, 1986; Richardson, Simmering, & Sturman, 2009). In the work on listening anxiety, for instance, because respondents self-report both their listening anxiety and their aggressive tendencies, it is possible that reported correlations between these constructs are a function of a common reporting format for these variables and not a function of an actual relationship between them (see Kotowski, Levine, Baker, & Bolt, 2009). Ways to handle CMV are covered in the next section.

A third issue with self-report listening research is that most studies do not present any evidence for construct validity. *Construct validity* is the extent to which a scale measures the theoretical construct of interest (e.g., listening anxiety). Typical in the listening literature are studies that report evidence for convergent and discriminant validity by showing correlations between a newly constructed scale and measures of theoretically relevant constructs, or *nomological network validity* (Cronbach & Meehl, 1959). Nomological networks are representations of connections between concepts, their observable manifestations, and how these concepts are related. What studies often fail to do, however, is provide adequate empirical evidence that justifies the creation of a new construct and its observable manifestation—the actual scale. Questions that remain unanswered include whether we need the new scale (e.g., whether it replicates an existing construct) and whether the items that comprise that new scale actually "represent" the construct of interest. Perhaps most emblematic of this problem are studies that create scales to measure self-reported listening competence. As reported by Fontana, Cohen, and Wolvin (2015), there are dozens of scales that allegedly measure perceptions of listening competence. Some of these scales have some items in common, but most items are usually not isomorphic. Thus, the researcher who constructed the scale may lay claim to a supposedly "new" kind of listening competence construct, even though there are scales that already measure exactly that, albeit with different items. To date, perceived listening competence scales have not been administered simultaneously to a sample of participants to determine scale overlap. We are left to wonder whether these scales actually do measure different facets of listening competence or whether they are simply iterations of the same construct. The literature can thus become littered with conceptually indistinct and hence unnecessary constructs, a problem known as *construct proliferation*: "the accumulation of ostensibly different but potentially identical constructs representing [listening] phenomena" (Shaffer, DeGeest, & Li, 2015, p. 1). Ways of ascertaining evidence for discriminant validity of a particular scale are discussed in the next section.

Of course, discriminant validity is only one part of the overall psychometric portfolio of a scale. Part of the validity portfolio of any scale is evidence that the scale factors as expected across a range of populations. In other words, if a listening scale contains various subscales, then evidence is needed that items making up each subscale "hang together" in expected ways. A prime example of this problem comes from the Listening Styles Profile (LSP-16; see Profile 36). The LSP-16 has been used widely by researchers, educators, and practitioners, but the factor structure has been assumed as stable in these studies rather than systematically tested and retested. Instead, the scale had only been submitted to principal component analysis, and in those analyses the four listening styles accounted for just over 50% of the variance among the 16 items (Bodie & Worthington, 2010). The remainder of the unexplained variance is explained either by another set of factors or by measurement error. When submitted to confirmatory factor analysis (CFA) using data from more than 700 participants, Bodie and Worthington found that the scale was not psychometrically sound. A large culprit for this poor model fit was random error, suggesting that individual items were poor indicators of their supposed constructs (i.e., the listening styles). New items needed to be written, and a case for construct validity needed to be made anew (see Profile 36). The assumption of much listening research is that if a scale has been published, it is "valid"—a dangerous assumption indeed. Many "established" scales lack validity evidence, but even in the face of no evidence (or counterevidence), people still use them (Levine, 2005). Scales need to stand the test of time and be submitted to rigorous tests of validity. Notably, validity is an ongoing process. Scales are not valid or invalid; instead, scales can have larger or smaller validity portfolios.

A final observation about self-reported listening research is that it is heavily biased toward assessing self-reports of behaviors or behavioral tendencies rather than self-reports of internal tendencies, an issue we discussed in this chapter. Although there are examples of self-report listening scales directed toward internal cognitive states (e.g., IRAT; see Table 5.1), beliefs about listening (e.g., the Listening Concepts Inventory), motivations (e.g., Willingness-to-Listen), and situationally derived goals (e.g., LSP-R), there are many more examples of scales attempting to gauge behaviors. As described in Chapter 6, *behavior* is defined as something a listener *does*—specific actions a listener enacts while interacting with another person.

Examples of scales that attempt to measure behavior through self-report (all are profiled in Section Three) include the Organizational Listening Survey (OLS), the Self-Reported Listening Competence (SRLC) scale, the Active-Empathic Listening Scale

(AELS), and the Academic Listening Self-rating Questionnaire (ALSQ). Although each scale attempts to measure slightly different facets of listening competence, the scales are similar insofar as they assume people are able to accurately report on what they typically do while listening. Work by Bodie *et al.* (2014) and others (Cooper & Husband, 1993; Ford, Wolvin, & Chung, 2000; Lawson & Winkelman, 2003) questions this assumption; participants tend to overestimate their own listening abilities. In general, the reader is urged not to rely on self-report methods when interested in studying listening behavior.

How to Improve Self-Report Listening Research

In this chapter, we referenced and briefly outlined four abuses of self-reported listening research, namely, that it tends to: (a) Be cross-sectional, (b) contain measurement error, (c) rely on scales that lack validity evidence, and (d) measure behavior rather than internal states. In this section, we discuss some solutions to these abuses.

Make Cross-Sectional Work Part of a Larger Program of Research

A cross-sectional study involves measuring a set of variables at one point in time only. A typical cross-sectional study involves creating a questionnaire by combining multiple scales and demographic information such as age, biological sex, and occupation and distributing that questionnaire online or in person to a sample of individuals drawn from a selected population of interest. Such a design allows researchers to compare many different variables at the same time at a relatively low cost. But because a cross-sectional survey captures only a single moment in time, we are left only with correlational data and must speculate about which variables are antecedent and which ones are consequent. For instance, if you distributed a questionnaire that included the IRAT (see Table 5.1) and a measure of typical communication patterns in families, you might find that individuals scoring higher on listening anxiety also report coming from families that stress a climate of homogeneity of attitudes, values, and beliefs (i.e., conformity orientation; see Ledbetter & Schrodt, 2008). Although it might be tempting to conclude that the ways that families communicate *cause* children to develop a tendency to experience listening anxiety, such a conclusion is beyond the available data.

For questions that require causal answers, longitudinal designs can be useful. Like cross-sectional studies, longitudinal studies involve asking participants to report on their dispositions, typical activities, preferences, and/or perceptions. In a longitudinal study, however, researchers repeatedly measure the same participants over a period of time, sometimes lasting several months or even years. This kind of longitudinal design is called a *cohort* or *panel study*. Another type of longitudinal study covered in depth in Chapter 3 is *ethnographic fieldwork*. Neither panel studies nor ethnographic work allows the researcher to establish causality. Regardless, longitudinal designs allow researchers to explore how key variables change over time and to discover evidence that suggests one causal order might be more plausible than another (e.g., listening anxiety and family communication patterns).

A study design that can more clearly enable causal claims is the experimental study, more specifically, the *randomized controlled trial* (RCT), which is the “gold standard” of evidence-based empirical research. The RCT requires that samples be randomly drawn and that participants be randomly assigned to conditions. Whereas cross-sectional and longitudinal designs allow researchers to investigate issues without directly influencing or manipulating the environment, experiments involve manipulating the environment and then observing and measuring reactions to those manipulations. The variable that is manipulated is the *independent variable*, and the variable that is thought to change as a function of the manipulation is the *dependent variable*. If researchers are concerned with causal relationships (as we often are in listening), the appropriate design is the experiment, offering full control over the independent variables under question. To study listening anxiety and family communication patterns, for instance, we might train some families to exhibit behaviors more indicative of a conversation orientation, one in which all family members are encouraged to participate in unrestrained exploration of all kinds of topics. Then we would document the degree to which listening anxiety changed as a function of this intervention. If anxiety levels vary as a function of the intervention relative to control groups that did not receive the intervention, we have more definitive evidence of a causal relationship between these two variables. But we still must rule out other plausible variables that could confound our results, and we must be cautious because this particular example did not allow for random assignment, which might, in fact, introduce error (a concern we have addressed in this chapter).

In general, the best method for your own research depends on the questions you are asking. Moreover, your choice of research method is at least partially a function of the research that was previously done on your research question. You must assume researchers before posed precisely the same question and tested it. All scientific research is cumulative, and we encourage you to thoroughly research your question to determine (a) the existing knowledge base and (b) existing scales. Depending on your specific area of listening research, you might be able to conduct a cross-sectional study. Returning to the Ledbetter and Schrodt (2008) report on family communication patterns and listening anxiety, the study was the first to establish whether there are links or associations between these constructs. A cross-sectional study design was thus appropriate because the purpose of their study was descriptive: to discover the prevalence of one or more variables in a particular population or subgroup of that population. When conducting work in an area that has quite a bit of formative research already, however, longitudinal or experimental designs are more appropriate, yet also more costly.

Create New Constructs (and Measures) Only When Necessary

As interest in listening has increased, so too have the number of constructs used to describe its operation. At first blush, this does not seem overly problematic. Indeed, to understand something as abstract and complex as listening, one would assume the need for myriad constructs and operative mechanisms. As the Greek poet Hesiod famously wrote, however, moderation is best in all things. According to Occam's razor, which sets the standard for good scientific theory, when possible, the most parsimonious explanation is preferred. For listening scholars, that means generating the fewest number of constructs possible to explain how and in what ways individuals behave as listeners.

Similar to our discussion in Chapter 1 regarding the proliferation of listening definitions, there also is a tendency to add adjective descriptors to listening in hopes to set these listening constructs apart from already existing constructs. Terms like *active listening* (Rogers, 1955; Weger, Castle, & Emmett, 2010), *therapeutic listening* (Wolvin & Coakley, 1993), *supportive listening* (Bodie, Vickery, & Gearhart, 2013; Jones, 2011), and *empathic listening* (Myers, 2000) all describe a mindful approach to listening, one that truly seeks to understand what others are really saying. Adding to this proliferation are closely related constructs, such as *responsiveness*, *understanding*, *attentiveness*, *conversational sensitivity*, *interaction involvement*, and *affectionate communication*, that share a great deal of conceptual space with listening and may not be empirically distinct from it. Then, there are broader terms such as *empathy*, *cognitive complexity*, and *listening anxiety* that likely contribute to an individual's ability to listen in a competent manner, as well as terms such as *listening goals* and *listening concepts* that describe cognitive schemata related to listening.

The basic point is that authors of supposedly new constructs should, at minimum, advance a strong theoretical argument for conceptual distinctions between their "new" construct and related constructs. Advancing an argument forces scholars to examine existing knowledge bases and may encourage them to abandon their constructs, which might ultimately strengthen a research field. Pragmatically, most journal editors and reviewers require authors to advance an argument for conceptual distinctiveness. As explained by Harter and Schmidt (2008), however, the "implicit assumption … [that if] they and other researchers can make a logical or conceptual distinction between constructs or measures, then this distinction will exist in the minds of … respondents to surveys … may not hold true" (p. 36). In other words, constructs that are conceptually distinct may not be empirically discrete when data are collected from surveys created to operationalize these constructs.

Measures that are similar lack empirical distinctiveness and are correlated at or near 1.0 after correcting for measurement error. Evidence from other literatures (e.g., organizational science) suggests that the empirical distinctiveness between related constructs is rarely investigated (Shaffer *et al.*, 2015). Broadly speaking, we are in the realm of *discriminant validity*, or the extent to which measures of theoretically distinct constructs are unrelated empirically to one another. Gathering evidence of discriminant validity is part of generating strong evidence for construct validity.

Generate Strong Evidence for Construct Validity

As mentioned, current listening measures often exhibit a good deal of measurement error and lack sufficient evidence for construct validity. Measurement error is unavoidable, of course, because the constructs we wish to study are abstract idealizations of a concrete empirical reality. And, of course, no measure ever fully captures the construct that it operationally defines—a frequent and misguided assumption known as *definitional operationalism* (Campbell, 1969). But sound research design and strict adherence to data collection procedures can eliminate a lot of error. Procedures outlined in Chapter 2 can assist in the development of measures that include lower levels of error.

A very basic way of modeling measurement error is to look at Cronbach's alpha as an estimate of the internal consistency of a set of scores. Chapter 2 presented the formula for alpha, and this statistic can be used to explore how much measurement error has attenuated the relation between two constructs. Say, for instance, we have collected responses from both the Self-Perceived Listening Competence Scale (SPLCS; see Profile 57) and the listening anxiety (LA) items from Table 5.1. Furthermore, suppose we find a correlation of .60. Suppose, in addition, that the reliability estimates for the SPLCS and LA are .76 and .68, respectively. The formula used to correct the observed correlation for measurement error is:

$$r_{corrected} = \frac{r_{xy}}{\sqrt{(r_{xx})(r_{yy})}}$$

where r_{xy} is the Pearson product moment correlation coefficient, estimating the SPLCS–LA relationship; r_{xx} is the internal consistency estimate of SPLCS; and r_{yy} is the internal consistency estimate of LA. Plugging in our values, we get a value of $r_{corrected} = .83$. This value is not 1.0, but it is rather high, suggesting quite a bit of overlap between the SPLCS and LA. This correlation does not suggest that LA and listening competence are the same construct, but it does suggest that the scales used to measure these constructs are not as distinct as the developers originally thought.

Measurement error can also be accounted for and modeled through CFA, which also allows for testing a central facet of construct validity, the degree to which a scale factors appropriately across a range of populations. Unlike Pearson's coefficient, the bivariate relation generated by CFA will be corrected for measurement error, making the extra step of correcting for attenuation unnecessary. In addition, the CFA output will include factor loading information, as well as information regarding the extent to which the data conform to the theoretical measurement model—that is, the model presented by the test developers (Hoyle, 2000; Levine, Hullett, Turner, & Lapinski, 2006; Raju, Laffitte, & Byrne, 2002; Thompson, 2004).

Building a case for construct validity also requires evidence that the scale is related with theoretically similar measures (convergent validity) and lacks associations with theoretically dissimilar measures (discriminant validity). The former evidence is more common in the listening literature and generally consists of authors developing a nomological network for the construct under question and then including measures of the associated constructs along with the new measure. For instance, if we were developing a scale of LA, we might also ask participants to complete scales or tasks that tap listening comprehension and listening competence, reasoning that people with higher LA should have lower comprehension scores and lower levels of listening competence. If correlational data indeed were to show positive associations between LA, competence, and comprehension, our LA scale would possess some convergent validity.

The case for discriminant validity is not as readily made in the listening literature. To establish a case for discriminant validity, a first step is to show that our newly created measure does not duplicate existing measures. As seen here, we can do this by correcting the observed correlation for measurement error. We also can use CFA to model items in various configurations to generate a best fitting model that best captures the factor structure of our scale items. In our SPLCS–LA example, we would first specify that the SPLCS items all loaded on that factor and the LA items all loaded on a separate factor. We would compare this *unconstrained model*, the model that allows these constructs to freely covary, to a *constrained model* that specifies that the items are better represented by a single factor (a unidimensional construct). If the unconstrained model produced better fit than the constrained model, we would argue that there is evidence of discriminant validity (for details on how to conduct such an analysis, see Byrne, 2010).

Another method for gathering discriminant validity evidence is known as the *multitrait-multimethod matrix* (MTMM). The MTMM approach was introduced by Campbell and Fiske (1959) as an empirical solution to construct validity. In this approach, researchers employ measures of at least two traits (constructs) using at least two methods. For instance, in their MTMM study, Bodie *et al.* (2014) were interested in two constructs, active-empathic listening (AEL) and nonverbal immediacy (NVI). They measured these two constructs in three ways: self-reports, partner-reports, and behavioral observations. Participants reported on their tendencies to enact both AEL and

Table 5.2 Multitrait-Multimethod Correlation Matrix

	Active-empathic listening				Nonverbal immediacy			
	Self-report	Other-report	Partner-report	Behavior	Self-report	Other-report	Partner-report	Behavior
AEL self-report	**.87**	.23	−.14	−.07	*.50*	.08	.02	.13
AEL other-report	.20	**.89**	.06	.00	.24	*.48*	.01	.35
AEL partner-report	−.12	.05	**.90**	.35	−.01	.15	*.75*	.09
AEL behavior	−.06	.00	.30	**.80**	.00	.17	.28	*.42*
NVI self-report	*.43*	.21	−.01	.00	**.84**	.37	.11	.19
NVI other-report	.07	*.42*	.13	.14	.32	**.87**	.19	.23
NVI partner-report	.02	.01	*.63*	.22	.09	.16	**.78**	.16
NVI behavior	.11	.30	.08	*.34*	.16	.19	.13	**.81**

Note: AEL = active-empathic listening; NVI = nonverbal immediacy. Correlations above the MM diagonal are corrected for measurement error, whereas those below the MM diagonal are not.
Color coding key:

Monotrait-monomethod correlation	Monotrait-heteromethod correlation	*Heterotrait-monomethod correlation*	Heterotrait-heteromethod correlation

Source: From Bodie *et al.* (2014). Table reproduced with permission.

NVI (self-report) and were then paired with another individual who disclosed a stressful event. The two people talked for 5 minutes, after which the disclosers reported on the listeners' AEL and NVI (partner-report). Because the conversations were video-recorded, the conversations were additionally coded for actual AEL and NVI (behavioral observations). In the end, each participant had three scores for AEL and three scores for NVI, which generated a matrix of correlation coefficients (see Table 5.2).

The first type of correlation generated by an MTMM analysis is the *monotrait-monomethod* (MM) correlation. This correlation is synonymous with the reliability coefficient of a scale. When these values are high, the measures are said to demonstrate a high degree of internal consistency. To generate MM correlations, our research team calculated Cronbach's alpha values for each scale (see Chapter 2); we also corrected subsequent correlations using these estimates (as discussed here). The second type of correlation, the *monotrait-heteromethod* (MH) correlation, represents the association between different measurement methodologies used to measure the same construct. In the AEL–NVI study, the two constructs were measured as self-report, partner-report, and observed behavior. The set of MH correlations for AEL showed the extent to which self-reported, partner-reported, and observed AEL correlated with each other (and similarly for NVI). When MH correlations are sufficiently large, researchers are provided direct evidence of convergent validity. Often referred to as *validity coefficients*, these values should also be sufficiently larger than the heterotrait correlations to demonstrate evidence for divergent or discriminant validity. In the AEL–NVI study, none of the MH correlations were higher than .30, and some were negative, suggesting little evidence for convergent validity. For instance, the correlation between self-reported AEL and partner-reported AEL was not statistically different from zero, suggesting that how the participant saw him or herself was not related to how the partner perceived his or her listening.

The last two correlations produced by an MTMM analysis are both heterotrait correlations, in our case correlations between measures of AEL and NVI. *Heterotrait-heteromethod* (HH) correlations are associations between different measurement methodologies used to measure different constructs, such as correlations between self-reported AEL and partner-reported NVI or observed AEL and self-reported NVI. Second, *heterotrait-monomethod* (HM) correlations are associations between different constructs measured by a common methodology (e.g., self-reported AEL and NVI, or partner-reported AEL and NVI). In the AEL–NVI study, all HM correlations were quite high, suggesting that the method used to measure these constructs explains a good deal of the shared variance among the constructs.

In general (and back to why we started talking about MTMM in the first place), comparisons between the two heterotrait correlations (HH and HM) are used to assess CMV. When the HM correlations are larger than the HH correlations, measurement bias is a concern. Basically, if most of the variance in the dataset is attributable to method, then researchers should be wary of using cross-sectional, self-report data to make definitive conclusions about relations among facets of listening and other trait-like variables. As we showed with the AEL–NVI study, CMV is likely a concern of self-reported listening research using the AELS and NVI measures that should not be ignored. The extent to which other listening constructs might suffer from CMV needs to be examined.

Acknowledge What Self-report Scales can Measure

The discussion of CMV moves us to our final recommendation regarding self-reported listening research: Do not use these measures as proxies for behavior. Appropriately, this discussion also brings us full circle to a key position of this chapter: Self-report measures of listening are most appropriately described as tapping affective components of listening, such as motivations and attitudes toward listening. A listening behavior is something listeners do, not something they think they do, something they remember doing, or even something they think they tend to do in general. Measuring behaviors requires that people observe listeners in action.

In the AEL–NVI study, the data revealed a large degree of CMV. What this means is that individual reports of tendencies to listen in active-empathic ways are moderately correlated to their reports of tendencies to behave in more or less nonverbally immediate ways; the same is true for actual AEL and NVI behavior (as rated by a trained judge), and perceptions of AEL and NVI by a conversational partner. More important, the correlations between methods were much higher than correlations between self- and partner-reported and rated AEL and NVI behaviors. In a larger sense, these results suggest that studies investigating validity issues with single-method measures of listening (e.g., using all self-report measures to demonstrate convergent validity) may be reporting spurious associations rather than real relations among latent constructs (Podsakoff & Organ, 1986). As a result, previously reported associations may need to be adjusted downward to account for this bias.

No matter its degree, the consequences of CMV are considerable. In egregious cases, estimates may be within interpretable bounds yet be entirely a function of shared methods across constructs that inflate Type I error (i.e., the "glop" problem; Bank, Dishion, Skinner, & Patterson, 1990). The question then becomes how one can minimize (rather than completely eradicate) this systematic error variance. The reader is referred to Richardson *et al.* (2009) to determine appropriate solutions for CMV. At the very least, the degree to which CMV changes the conclusions regarding past work should be addressed (e.g., the degree to which relations among listening and relationship outcomes change as a function of decoder perspective used to operationalize listening).

Self-Reports of Listening: What Are They Good For?

Unlike Frankie Goes to Hollywood's answer to "War and what is it good for?" ("Absolutely nothing!"), we think that self-reports of listening are good for many things. Indeed, we started this chapter by discussing some of the advantages of self-reporting of listening, and we retain that attitude—that, when used appropriately, we can learn much from self-reports of listening. In particular, researchers are encouraged to use self-report measures to investigate internal states, beliefs about listening, motivations to listen in particular ways, and situationally influenced listening goals. We touch on each of these in this section, using scales profiled in Section Three as examples.

Measuring Internal States

Recall that listening scholars have two sources of information available when measuring listening facets: things participants report and things that can be observed. Each source of information can provide data about internal states. Observed physiological markers can be used to measure stress or anxiety, for instance, but for the most part, listening scholars have relied on self-report methods to ascertain how listeners think and feel about listening. Moreover, as we noted in this chapter, for some constructs, self-report may be the only feasible way to collect data.

The IRAT (see Profile 24) measures one internal state related to listening, the degree to which a listener experiences anxiety (see Table 5.1). Examples of other profiled measures that capture internal listener states include the Affectionate Communication Index (ACI), Attributional Complexity (AC), the Interpersonal Reactivity Index (IRI), and the Rational-Experiential Inventory (REI). Each of these measures can be used to tap trait-level characteristics. Most measures also can be modified to tap state-level characteristics, such as the degree of listening anxiety experienced before, during, or after a particular listening event.

Measuring Beliefs About Listening

What people believe about listening can have powerful effects on how they enact (or fail to enact) behaviors in the service of attending to others. Our beliefs about listening also likely influence how we judge others as they listen to us. When asked, participants readily list a consistent set of behaviors associated with good listening (Bodie, St. Cyr, Pence, Rold, & Honeycutt, 2012), most of which are represented on scales that tap self-perceived listening competence.

Two examples of listening competence scales profiled in Section Three are the Self-Perceived Listening Competence Scale (SPLCS) and the Organizational Listening Survey (OLS). These scales are intended as self-reports of how well one listens in general; or, if instructions are modified, in a context of interest (e.g., in the classroom or at work). As defined in this book, however, behaviors are things people do, not things people think they do; they are concrete actions that are displayed in the moment. As memory research teaches us, we do not retain much specific information after an interaction, especially information about the exact proportion of time we spent engaging in eye contact or the exact number of open-ended questions we asked. Others have used these competency scales to measure perceptions of interlocutors regarding a listener; so, for example, your coworkers could fill out one of these scales with respect to how they think you listen. This strategy, known as *other-reporting*, is valid insofar as you are interested in what other people think about you. But other-report measures do not tell you what another person did, only what the perceiver thinks that individual did. Of course, perception is one form of reality, so there can be power in studying what other people think about certain listeners. One measure designed to assess what others think about a listener's degree of responsiveness is the Perceived Partner Responsiveness Scale (PPRS).

Another strategy for exploring what people think about listening is to ask them to define the term, a strategy used to develop the Listening Concepts Inventory (LCI). The first step in developing the LCI involved exploring the lay and scholarly literature related to listening and gathering several dozen terms considered synonymous with or closely related to listening (e.g., *understanding*, *attention*, and *learning*). Then, college students in the United States and Germany rated the degree to which each of these terms is identical to or not at all similar to listening (see Profile 32 for exact scale points). Results suggested four broad ways in which people think about listening: as organizing information, as relationship building, as learning and integrating information, and as a critical activity. The work on the LCI is similar in many respects to work on implicit theories of relationships (Knee, 1998), personality (Krzystofiak, Cardy, & Newman, 1988), communication (O'Keefe, 1988; also see the Communication Functions Questionnaire, Profile 11), and other facets of human life that influence how we behave in the presence of others.

Measuring Motivations to Listen

The motivation to listen is an integral part of listening competence—in order to behave in effective and appropriate ways, the listener must not only know how to behave but also have the motivation to behave in that way. One set of measures profiled in Section Three was explicitly designed to tap motivations to listen in particular ways. The Willingness to Listen (WTL) scales were designed to directly measure individual motivations to listen to others in various settings; they have not, however, been used extensively in the literature and do not have powerful validity portfolios.

Other scales also tap listening motivation and have slightly more robust validity portfolios, although they are most often positioned (much like measures of listening competence) as measures of dispositions or tendencies to listen in particular ways. Those include the Active-Empathic Listening Scale (AELS), the Attitude Toward Active Listening Scale (ATALS), the Conversational Sensitivity Scale (CSS), the Interaction Involvement Scale (IIS), the Talkaholic Scale (TAS), and several measures of NVI. The MTMM study detailed in this chapter, which explored reports and observations of AEL and NVI, has already clued you in to our opinion regarding the use of these scales as proxies for behavior. They are, however, useful to the extent that they might tap the motivation to be a particular type of listener. In the case of the AELS, for instance, perhaps this scale taps the degree to which a listener wants to consciously understand another individual from that individual's perspective. Interpreting the scale in this way is supported by high associations between the AELS and measures of empathy (see Profile 2). Likewise, conversational sensitivity might tap the degree to which a listener wants to be sensitive to both content and relational aspects of speech.

Measuring Situationally Derived Listening Goals

A final affective category to which self-reports seem appropriate is the measure of situationally derived listening goals. The first conceptualization of listening-related goals was developed by Watson, Barker, and Weaver (1995), who proposed the construct of *listening style* as the variability in how people attend to and process information. In particular, Watson *et al.* identified four listening orientations—people, action, content, and time—that individuals habitually use, especially in novel situations (Imhof, 2004). Problems encountered in studies utilizing the LSP-16 (Bodie & Worthington, 2010) led Bodie, Worthington, and Gearhart (2013) to revise and frame this typology as representing four distinct "goals that listeners have when engaged in situations that call them to be a particular kind of listener" (p. 17; see LSP, Profile 36). In a similar manner, the typology of listening competencies outlined by Wolvin and Coakley (1993) that directed the development of the SPLCS also can be interpreted as identifying different goals that listeners might seek to accomplish in interaction. Thus, the SPLCS, instead of being framed as a measure of listening beliefs, might best be framed as a measure of the goals available while listening.

New Horizons: Daily Diaries and Experience Sampling Methods

So far, we have covered advantages of self-report measures, as well as proper uses and common abuses of this method. We also detailed some ways to improve self-reported listening research. One key to improvement is to begin shifting our scholarship from an overreliance on cross-sectional self-reports and to incorporate measures with strong validity portfolios into studies that will allow us to explore the ways in which affective components of listening vary over time. One of the best ways to do so is with diary studies.

What Is a Diary Study?

A *diary study* involves asking participants to repeatedly submit self-reports of "events, reflections, moods, pains, or interactions near the time they occur" (Iida, Shrout, Laurenceau, & Bolger, 2012, p. 277). Learning about others from diaries is, of course, nothing new. Historians and literary scholars utilized diary records long before communication studies emerged as a discipline, and the everyday experiences of people have long been an interest of psychologists, anthropologists, linguists, and many other social scientists. Although not classified as a diary method, the ethnographic methods covered in Chapter 3 have much in common with what we are discussing here—a desire to capture life-as-lived as close to its occurrence as possible and within the frame-of-reference of participants. One main difference between diary studies and methods such as open-ended journaling and participant observation, however, is that diary studies employ standard self-report instruments to maintain a degree of standardization. When participants are asked to report using a standardized form, the method is often referred to as *experience sampling* (Larson & Csikszentmihalyi, 1983).

Of course, standardization does not mean that all diary studies employ a standard set of methods. Indeed, diary studies have employed a range of methods and procedures. The larger point is that the same self-report scales used to capture snapshots of participant internal states, motivations, habits, beliefs, and other affective components of listening can be used in a continuous format that allows a broader (and approximating a motion) picture of the lives of listeners (see Chapter 3). As stated by Shiffman, Stone, and Hufford (2008), "global, retrospective reports … [miss] the dynamics of life as it is lived, day-to-day, hour by hour" (p. 3).

Experience sampling is actually one of the earliest methods used in listening research (Rankin, 1926; see Time Studies, Profile 60). Time-use studies attempt to estimate the amount of time people spend doing various activities, usually breaking down estimates for various 24-hour periods and reporting percentages of time spent on specific tasks. Rankin's (1926) study asked a convenience sample of 21 people to keep a log of their communication activities for one or more days in 15-minute increments from 6:00 a.m. to midnight. Several other studies have been published that sought to replicate and extend these results for specific populations (e.g., college students, scientists, and engineers). Moreover, studies exploring time spent listening highlight the variability in methods that can be employed in diary studies. For instance, studies prior to 1980 primarily utilized one or more forms of *time-sampling* procedures, asking respondents to report at various times of the day what communication activities they were engaged in. Rankin asked respondents to report every 15 minutes, and these logs of time spent were recorded for between 1 and 18 days. Perras and Weitzel (1981) used a similar method with reports every 30 minutes of waking time. Bird (1953) reported having students keep "a running record of minutes spent" in the four modes, whereas Weinrauch and Swanda (1975) asked respondents "to keep a careful record of their time spent in communication" (p. 27). Hinrichs (1964) used a primitive form of *signal-contingent recording*, asking participants to set a wristwatch alarm at five random times during the day and report on their communication up to that point in time. Since the publication of the time study by Barker, Gladney, Edwards, Holley, and Gaines (1980), participants are most often asked "to think back over the last 24 hours and answer the questions based on [this] reflection" (p. 103).

Why Conduct a Diary Study?

There are both methodological and theoretical factors that researchers should consider when conducting a diary study. Although retrospective self-reports provide information regarding reconstructed experience or perceptions, summarizing the past from a respondent's current point of view, daily diaries provide information regarding experience as it is lived. Diary data allow researchers to explore presently felt emotions and close-to-real-time reports of ongoing experiences as well as potentially more accurate retrospective accounts of behavior. Conversely, traditional self-report scales better reflect internal, motivational, affective, and perceptual processes that play a role in how people behave, but they fall short of capturing behavior per se. It is important to note that "both types of data are relevant to understanding human behavior" (Reis, 2012, p. 5). Listening scholars want to know both "what actually happened … [and] how people experience or understand events in their lives, given time to reflect on them" (Reis, 2012, p. 5). Thus, we are not arguing to throw out self-report measures completely and replace them with diary studies. Instead, researchers should use global retrospective reports alongside more frequent reports of daily experience (as well as experimental and observational methods).

Theoretically, diary methods "make available a different kind of information than traditional methods do, information that provides a novel and increasingly valuable perspective on behavior" (Reis, 2012, p. 4). In particular, the type of information provided by daily reports of experience and behavior is more ecologically valid than retrospectively recalled behavior. For one, the collection of data is closer to the occurrence of the behavior—asking participants to reflect on their listening anxiety immediately after a conversation will produce data that better describe anxiety felt in that conversation compared to asking them to reflect days or weeks after that interaction. As a type of external validity, ecological validity is important for the representativeness of data—the closer to the experience we can get, the more we can generalize to similar experiences. Moreover, daily reports of experiences and behavior are always contextual, compared to the acontextual nature of most global reports or the contrived nature of many experimental studies; how and why people listen are at least partially a function of the situations they find themselves in. Daily diary methods allow researchers to take seriously calls to contextualize listening research.

Daily Sampling of Listening Experiences

Ideally, if researchers want to know about behavior—what a person does when listening—they should observe the listener in action. At the same time, some listening behaviors are not easily observed by researchers. For instance, discovering how dual-earner couples (i.e., both individuals work full-time outside of the home) listen to each other as they talk about the events of their day poses several logistical puzzles, not the least of which is recruitment. Although you might want to observe couples as they engage in this type of talk, how? Do you get them to come to your lab space? If so, when they get there, how do you introduce their task, to talk about the hassles that made up their day? Don't get us wrong: Having couples talk about stressful events in the lab is possible and oftentimes a desirable design choice (see Bodie, Cannava, Vickery, & Jones, 2016; Bodie, Vickery, Cannava, & Jones, 2015; see Couples Helping Exercise, Profile 14). The point is that recruitment for laboratory studies can pose serious logistical inconveniences, costs that have to be weighted in terms of the benefits that experimental methods provide.

Even if you are able to recruit a sample of couples who agree to come to the lab after work, collecting those data is extraordinarily labor-intensive. Not only are you limited to when you can collect these data (i.e., during the time of day the couples usually talk about their days), but also dual-earner couples may be reluctant to spend one or more hours in the lab after a long day of work. Moreover, those with children will require extra compensation or may be particularly reluctant to give you perhaps the only part of their day they get to see their kids awake. As a result, your sample will be limited to those participants willing and able to come to the lab; these people may or may not adequately represent the larger population to which you want to generalize.

A general rule to research by is to match your methods with your questions. In this case, our question involves how dual-earner couples listen during conversations about daily hassles—what Jefferson (1980) called "troubles talk." If there are little empirical data on this topic, we are warranted to ask participants to answer self-report questionnaires about this type of talk, perhaps by distributing a survey to a sample of dual-earner couples that asks them to retrospectively report how they act in and feel about these conversations. We could ask about their general tendencies in these types of conversations, or we could ask them to remember the most recent troubles talk conversation they had and report thoughts and feelings about how they and their partner listened. Alternatively, we could ask respondents to answer survey items several times—perhaps once a week or more. The point of designing a diary study to explore troubles talk is to get as close to the event as possible and to mitigate recall bias and other problems associated with memory for events. If we decided to go the daily diary route, we are then faced with several subsequent decisions.

Choose a Reporting Method

Much like sampling decisions for participants, researchers using diary methods have to choose how to sample events. In our dual-earner example, the event seems rather straightforward. We could define troubles talk for our research participants as "any conversation in which you and your partner talk about one or more troubles or hassles that happened to you over the course of the day." Our instructions could further specify that each couple is to report on each of these conversations that happened face-to-face for at least 5 minutes, a decision that would remove from consideration texts and email that couples might send throughout their day. This method is known as *event-contingent* sampling, reporting for events predefined by the researcher. We could ask participants to fill out our survey as close to the end of each of these conversations as possible, whenever and wherever they happened during the day. This design choice seems preferable to asking participants to report once per day and to recall all such conversations, unless of course we know from past work that such conversations typically happen only once per day (e.g., at the end of the work day as couples are destressing).

Other sampling choices include *time-contingent* sampling, which is sampling at regular, predetermined intervals, and *signal-contingent* sampling that asks participants to record responses when a signal is sent, for example, through an app loaded on participants' mobile devices or through an email or other messaging platform. Time-use studies tend to use a variation of time-contingent sampling, asking participants to report on every 15-minute interval throughout the day. More specific guidelines and typical schedules for time-contingent sampling are found in Larson and Csikszentmihalyi (1983).

According to Reis and Gable (2000), distinctions between event-, time-, and signal-contingent sampling methods "are not merely procedural details; each protocol is tailored to fit particular operational circumstances and theoretical goals, and findings depend to some extent on the choice of method" (p. 198). The reader is encouraged to consult sources already cited in this chapter, as well as additional sources (e.g., Bolger, Davis, & Rafaeli, 2003; Bolger & Laurenceau, 2013; Reis, Gable, & Maniachi, 2014; Wheeler & Reis, 1991) for more information on sampling-related decisions.

How much do Participants Report?

When events are rare, event-contingent reporting methods are likely the best choice. There is no need to ask couples to report continuously on troubles talk, for instance, if that type of talk occurs only once per day or a few times per week. For more common events, like listening to music, choices have to be made about how often participants should report. With time-contingent sampling, it is rather standard to have participants report once per day or to report, for instance, once in the morning and once at night. For signal-contingent sampling, the researcher has to choose how many times to signal participants without making the study an inconvenience. For some participants, work or other daily events make it impossible for them to respond more than once or twice per day (e.g., when they are not at work).

Depending on the research question, researchers also have to consider the length of the reporting instrument. Clearly, if participants are reporting two or more times per day, you cannot expect them to fill out several hundred scale items. Many researchers who employ diary methods often employ single-item measures of constructs, opting for efficiency over standard concerns about reliability. For instance, a study investigating the experience of listening anxiety might ask, "How apprehensive were you during this listening event?", rather than employ the entire listening subscale of the IRAT; this question seems to get at the general construct under consideration and thus can be argued to be a valid indicator of listening anxiety.

How do Participants Report?

Regardless of the amount of reporting, researchers using diary methods also need to think about *how* data are collected. Some work employs paper-and-pencil diaries, whereas other work employs online or mobile technology. For time- and signal-contingent sampling, software is available (e.g., Qualtrics or Snap Survey) that allows researchers to preprogram emails, having them sent at regularly timed or random intervals, respectively. If you are using online software, you might want to explore how your survey will appear on tablets, e-readers, phones, and standard computers. Most systems now allow you to program surveys to account for differences in survey appearance, but a good rule is to pilot-test your survey before you begin your main data collection. Finally, in our example study of dual-earner couples, an additional consideration is involved, namely, making sure instructions specify that each individual is to report privately so responses are not contaminated by the other's opinions.

Conclusion

Listening is a common and consequential human activity and a fertile area of research. Understanding how and in what ways to attend to and understand others is imperative for building adequate theories of human interaction and behavior, as well as for teaching others to be productive members of society. The ability to conduct research on listening is, however, contingent on the ability to use methods that produce valid data. For most listening research, scholars rely on self-report survey instruments administered at one point in time. When surveys are used to ask participants about the frequency or duration of listening behaviors, the validity of the results are questionable. Instead, self-report methods are best thought of as ways to discover internal states, motivations, beliefs about listening, and situationally derived listening goals. Alternatively, participants can be asked to report on listening-related variables as close to their occurrence as possible, which is an especially useful method when behavioral observations are either logistically difficult or impossible. Such diary methods were once a staple of listening research and should be used more regularly to understand more than just time spent listening.

Most readers can take solace in the fact that there are always problems with methods, no matter the method chosen. Experimental studies, for all the control and internal validity they offer, often lack external validity (and especially ecological validity). Longitudinal studies, for all the rich data they offer with respect to how experiences change over time, are time-consuming and still rely on participant self-reports. Diary data do not ameliorate concerns about participants misrepresenting themselves. One thing is certain: We need much more research on the reliability of various data collection methods and the degree to which these methods are associated with each other. That much work remains to be done should be comforting, especially to graduate students and early-career scholars. The study of affective components of listening is vital to understanding how and why we listen to others in particular contexts and with particular results. We hope this chapter adds to conducting methodologically rigorous and theoretically sophisticated work.

References

Austin, E. J., Deary, I. J., Gibson, G. J., McGregor, M. J., & Dent, J. B. (1998). Individual response spread in self-report scales: Personality correlations and consequences. *Personality and Individual Differences, 24*, 421–438. doi:10.1016/S0191-8869(97)00175-X

Bank, L., Dishion, T. J., Skinner, M., & Patterson, G. R. (1990). Method variance in structural equation modeling: Living with "glop." In G. R. Patterson (Ed.), *Depression and aggression in family interaction* (pp. 247–279). Mahwah, NJ: Erlbaum.

Barker, L., Gladney, K., Edwards, R., Holley, F., & Gaines, C. (1980). An investigation of proportional time spent in various communication activities by college students. *Journal of Applied Communication Research, 8*, 101–109. doi:10.1080/00909888009360275

Bird, D. E. (1953). Teaching listening comprehension. *Journal of Communication, 3*, 127–130.

Bodie, G. D., Cannava, K., Vickery, A. J., & Jones, S. M. (2016). Patterns of nonverbal adaptation in supportive interactions. *Communication Studies, 67*, 3–19. doi:10.1080/10510974.2015.1036168

Bodie, G. D., Jones, S. M., Vickery, A. J., Hatcher, L., & Cannava, K. (2014). Examining the construct validity of enacted support: A multitrait-multimethod analysis of three perspectives for judging immediacy and listening behaviors. *Communication Monographs, 81*, 495–523. doi:10.1080/03637751.2014.957223

Bodie, G. D., St. Cyr, K., Pence, M., Rold, M., & Honeycutt, J. M. (2012). Listening competence in initial interactions I: Distinguishing between what listening is and what listeners do. *International Journal of Listening, 26*, 1–28. doi:10.1080/10904018.2012.639645

Bodie, G. D., Vickery, A. J., Cannava, K., & Jones, S. M. (2015). The role of "active listening" in informal helping conversations: Impact on perceptions of listener helpfulness, sensitivity, and supportiveness and discloser emotional improvement. *Western Journal of Communication, 79*, 151–173. doi:10.1080/10570314.2014.943429

Bodie, G. D., Vickery, A. J., & Gearhart, C. C. (2013). The nature of supportive listening I: Exploring the relation between supportive listeners and supportive people. *International Journal of Listening, 27*, 39–49. doi:10.1080/10904018.2013.732408

Bodie, G. D., & Worthington, D. L. (2010). Revisiting the Listening Styles Profile (LSP-16): A confirmatory factor analytic approach to scale validation and reliability estimation. *International Journal of Listening, 24*, 69–88. doi:10.1080/10904011003744516

Bodie, G. D., Worthington, D. L., & Gearhart, C. C. (2013). The Listening Styles Profile (LSP-R): A scale revision and evidence for validity. *Communication Quarterly, 61*, 72–90. doi:10.1080/01463373.2012.720343

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579–616. doi:10.1146/annurev.psych.54.101601.145030

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York: Guilford.

Brownell, J. (2013). *Listening: Attitudes, principles, and skills* (5th ed.). Upper Saddle Ridge, NJ: Pearson.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Mahwah, NJ: Erlbaum.

Campbell, D. T. (1969). Definitional versus multiple operationalism. *Et Al., 2*(1), 14–17.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105. doi:10.1037/h0046016

Coakley, C. G., Halone, K. K., & Wolvin, A. D. (1996). Perceptions of listening ability across the life-span: Implications for understanding listening competence. *International Journal of Listening, 10*, 21–48. doi:10.1207/s1932586xijl1001_2

Cooper, L., & Husband, R. L. (1993). Developing a model of organizational listening competency. *Journal of the International Listening Association, 7*, 6–34. doi:10.1080/10904018.1993.10499112

Cooper, P. (1998). *The speaking, listening, and media literacy standards and competency statements for K–12 education*. Washington, DC: National Communication Association.

Cronbach, L., & Meehl, P. E. (1959). Construct validity in psychological tests. *Psychological Bulletin, 52*, 174–203. doi:10.1037/h0040957

Cronen, V. E., & Price, W. K. (1976). Affective relationship between the speaker and listener: An alternative to the approach-avoidance model. *Speech Monographs, 43*, 51–59. doi:10.1080/03637757609375915

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349–354. doi:10.1037/h0047358

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial clues*. Englewood Cliffs, NJ: Prentice-Hall.

Elliot, A. J., Eder, A. B., & Harmon-Jones, E. (2013). Approach-avoidance motivation and emotion: Convergence and divergence. *Emotion Review, 5*, 308–311. doi:10.1177/1754073913477517

Fontana, P. C., Cohen, S. D., & Wolvin, A. D. (2015). Understanding listening competency: A systematic review of research studies. *International Journal of Listening, 29*, 148–176. doi:10.1080/10904018.2015.1015226

Ford, W. S. Z., Wolvin, A. D., & Chung, S. (2000). Students' self-perceived listening competencies in the basic speech communication course. *International Journal of Listening, 14*, 1–13. doi:10.1080/10904018.2000.10499032

Gollwitzer, P. M., Fujita, K., & Oettingen, G. (2004). Planning and the implementation of goals. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of self-regulation: Research, theory, and applications* (pp. 211–228). New York: Guilford.

Halone, K. K., Cunconan, T. M., Coakley, C. G., & Wolvin, A. D. (1998). Toward the establishment of general dimensions underlying the listening process. *International Journal of Listening, 12*, 12–28. doi:10.1080/10904018.1998.10499016

Halone, K. K., Wolvin, A. D., & Coakley, C. G. (1997). Accounts of effective listening across the lifespan: Expectations and experiences associated with competent listening practices. *International Journal of Listening, 11*, 15–38. doi:10.1207/s1932586xijl1101_2

Harter, J. K., & Schmidt, F. L. (2008). Conceptual versus empirical distinctions among constructs: Implications for discriminant validity. *Industrial and Organizational Psychology, 1*, 36–39. doi:10.1111/j.1754-9434.2007.00004.x

Hinrichs, J. R. (1964). Communication activity of industrial research personnel. *Personnel Psychology, 17*, 193–204.

Hoyle, R. H. (2000). Confirmatory factor analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 465–497). San Diego, CA: Academic Press.

Hoyle, R. H., & Leary, M. R. (2009). Methods of the study of individual differences in social behavior. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 12–23). New York: Guilford.

Iida, M., Shrout, P. E., Laurenceau, J.-P., & Bolger, N. (2012). Using diary methods in psychological research. In H. Cooper (Ed.), *APA handbook of research methods in psychology: Volume I. Foundations, planning, measures, and psychometrics* (pp. 277–305). Washington, DC: APA.

Imhof, M. (2004). Who are we as we listen? Individual listening profiles in varying contexts. *International Journal of Listening, 18*, 36–46. doi:10.1080/10904018.2004.10499061

Jefferson, G. (1980). On "trouble-premonitory" response to inquiry. *Sociological Inquiry, 50*, 153–185. doi:10.1111/j.1475-682X.1980.tb00019.x

Jones, S. M. (2011). Supportive listening. *International Journal of Listening, 25*, 85–103. doi :10.1080/10904018.2011.536475

Keaton, S. A., & Bodie, G. D. (2013). The statistical and methodological acuity of scholarship appearing in the *International Journal of Listening* (1987–2011). *International Journal of Listening*, 115–135. doi:10.1080/10904018.2013.813206

Knee, C. R. (1998). Implicit theories of relationships: Assessment and prediction of romantic relationship initiation, coping, and longevity. *Journal of Personality and Social Psychology, 74*, 360–370. doi:10.1037/0022-3514.74.2.360

Kotowski, M. R., Levine, T. R., Baker, C. R., & Bolt, J. M. (2009). A multitrait-multimethod assessment of the Verbal Aggressiveness and Argumentativeness Scales. *Communication Monographs, 76*, 443–462. doi:10.1080/03637750903300247

Krzystofiak, F., Cardy, R., & Newman, J. (1988). Implicit personality and performance appraisal: The influence of trait inferences on evaluations of behavior. *Journal of Applied Psychology, 73*, 515–521. doi:10.1037/0021-9010.73.3.515

Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. In H. Reis (Ed.), *New directions for methodology of social and behavioral sciences* (Vol. 15, pp. 41–56). San Francisco: Jossey-Bass.

Lawson, M., & Winkelman, C. (2003). The social desirability factor in the measurement of listening skills: A brief report. *Counseling Psychology Quarterly, 16*, 43–45. doi:10.1080/0951507021000050212

Ledbetter, A. M., & Schrodt, P. (2008). Family communication patterns and cognitive processing: Conversation and conformity orientations as predictors of informational reception apprehension. *Communication Studies, 59*, 388–401. doi:10.1080/10510970802467429

Levine, T. R. (2005). Confirmatory factor analysis and scale validation in communication research. *Communication Research Reports, 22*, 335–338. doi:10.1080/00036810500317730

Levine, T. R., Hullett, C. R., Turner, M. M., & Lapinski, M. K. (2006). The desirability of using confirmatory factor analysis on published scales. *Communication Research Reports, 23*, 309–314. doi:10.1080/08824090600962698

McCroskey, J. C. (1997). Self-report measurement. In J. A. Daly, J. C. McCroskey, J. Ayres, T. Hopf, & D. M. Ayres (Eds.), *Avoiding communication: Shyness, reticence, and communication apprehension* (2nd ed., pp. 191–216). Cresskill, NJ: Hampton.

Metts, S., Sprecher, S., & Cupach, W. R. (1991). Retrospective self-reports. In B. M. Montgomery & S. Duck (Eds.), *Studying interpersonal interaction* (pp. 162–178). New York: Guilford.

Miell, D. E. (1987). Remembering relationship development: Constructing a context for interaction. In R. Burnett, P. McGee, & D. Clarke (Eds.), *Accounting for relationships* (pp. 60–73). London: Methuen.

Myers, S. (2000). Empathic listening: Reports on the experience of being heard. *Journal of Humanistic Psychology, 40*, 148–173. doi:10.1177/0022167800402004

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259. doi:hdl.handle.net/2027.42/92167

O'Keefe, B. J. (1988). The logic of message design: Individual differences in reasoning about communication. *Communication Monographs, 55*, 80–103. doi:10.1080/03637758809376159

Perras, M. T., & Weitzel, A. R. (1981). Measuring daily communication activities. *The Florida Speech Communication Journal, 9*, 19–23.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12*, 531–544. doi:10.1177/014920638601200408

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*, 517–529. doi:10.1037//0021-9010.87.3.517

Rankin, P. T. (1926). The measurement of the ability to understand spoken language. *Dissertation Abstracts, 12*(1952), 847.

Reis, H. T. (2012). Why researchers should think "real-world": A conceptual rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 3–21). New York: Guilford Press.

Reis, H. T., & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In H. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190–222). Cambridge: Cambridge University Press.

Reis, H. T., Gable, S. L., & Maniachi, M. R. (2014). Methods for studying everyday experiences in its natural context. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 373–403). Cambridge: Cambridge University Press.

Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods, 12*, 762–800. doi:10.1177/1094428109332834

Rogers, C. R. (1955). *Active listening.* Chicago: The Industrial Relations Center of the University of Chicago.

Schrodt, P., & Wheeless, L. R. (2001). Aggressive communication and informational reception apprehension: The influence of listening anxiety and intellectual inflexibility on trait argumentativeness and verbal aggressiveness. *Communication Quarterly, 49*, 53–69. doi:10.1080/01463370109385614

Schrodt, P., Wheeless, L. R., & Ptacek, K. M. (2000). Informational reception apprehension, educational motivation, and achievement. *Communication Quarterly, 48*, 60–73. doi:10.1080/01463370009385580

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Shaffer, J. A., DeGeest, D., & Li, A. (2016). Tackling the problem of construct proliferation: A guide to assessing the discriminant validity of conceptually related constructs. *Organizational Research Methods, 19*, 80–110. doi:10.1177/1094428115598239

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32. doi:10.1146/annurev.clinpsy.3.022806.091415

Stafford, L., Burggraf, C. S., & Sharkey, W. F. (1987). Conversational memory: The effects of time, recall, mode and memory expectancies on remembrances of natural conversations. *Human Communication Research, 14*, 203–229. doi:10.1111/j.1468-2958.1987.tb00127.x

Stafford, L., & Daly, J. A. (1984). Conversational memory: The effects of recall mode and memory expectancies on remembrances of natural conversations. *Human Communication Research, 10*, 379–402. doi:10.1111/j.1468-2958.1987.tb00127.x

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: APA.

Watson, K. W., Barker, L. L., & Weaver, J. B., III. (1995). The listening styles profile (LSP-16): Development and validation of an instrument to assess four listening styles. *International Journal of Listening, 9*, 1–13. doi:10.1080/10904018.1995.10499138

Weger, H. W., Jr., Castle, G. R., & Emmett, M. C. (2010). Active listening in peer interviews: The influence of message paraphrasing on perceptions of listening skill. *International Journal of Listening, 24*, 34–49. doi:10.1080/10904010903466311

Weiner, B. (1992). *Human motivation: Metaphors, theories, research*. Newbury Park, CA: Sage.

Weinrauch, J. D., & Swanda, J. R., Jr. (1975). Examining the significance of listening: An exploratory study of contemporary management. *Journal of Business Communication, 13*, 25–32. doi:10.1177/002194367501300103

Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality, 59*, 339–354. doi:10.1111/j.1467-6494.1991.tb00252.x

Wheeless, L. R. (1975). An investigation of receiver apprehension and social context dimensions of communication apprehension. *The Speech Teacher, 24*, 261–268.

Wheeless, L. R., Preiss, R. W., & Gayle, B. M. (1997). Receiver apprehension, informational receptivity, and cognitive processing. In J. A. Daly, J. C. McCroskey, J. Ayres, T. Hopf, & D. M. Ayres (Eds.), *Avoiding communication: Shyness, reticence, and communication apprehension* (pp. 151–187). Cresskill, NJ: Hampton.

Wolvin, A. D. (2009). Listening, understanding, and misunderstanding. In W. F. Eadie (Ed.), *21st century communication: A reference handbook* (Vol. 1, pp. 137–146). Thousand Oaks, CA: Sage.

Wolvin, A. D., & Coakley, C. G. (1993). A listening taxonomy. In A. D. Wolvin & C. G. Coakley (Eds.), *Perspectives on listening* (pp. 15–22). Norwood, NJ: Ablex.

Wolvin, A. D., & Coakley, C. G. (1994). Listening competency. *Journal of the International Listening Association, 8*, 148–160. doi:10.1080/10904018.1994.10499135

Wolvin, A. D., & Coakley, C. G. (2000). Listening education in the 21st century. *International Journal of Listening, 14*, 143–152. doi:10.1080/10904018.2000.10499040

Worthington, D. L., & Fitch-Hauser, M. E. (2012). *Listening: Processes, functions, and competency*. New York: Routledge.

6

Measuring Behavioral Components of Listening

Graham D. Bodie

> *[T]he difference between merely receiving an oral message and listening actively is similar to the difference between scanning a textbook and reading it for comprehension and retention.… In oral communication settings there must be* involved listeners *attempting to internalize and evaluate the message in order for a speaker to achieve his communication objective.*
>
> Barker (1971, pp. 2–3; emphasis added)

The above quote comes from Barker's *Listening Behavior,* one of the earliest listening textbooks. A major goal of Barker's text was to outline actions that listeners can take to become more active participants in (versus passive recipients of) a communication exchange. Barker's definition will likely seem incontrovertible to most readers, but it is important to note that it was not mainstream at the time. Indeed, for much of the history of communication studies, *listening* has been defined as a more passive and uninvolved act than an active and involved attempt to assist others in achieving communicative goals (Beard & Bodie, 2014). During the 1960s and 1970s, the broader communication studies discipline was heavily influenced by the work of, to name but a few, Charles Cooley, John Dewey, Wilhelm Dilthey, George Herbert Mead, Heinz Werner, and Ludwig Wittgenstein, each of whom fundamentally questioned the nature of human knowledge. Ideas borrowed from symbolic interactionism (Blumer, 1969; Littlejohn, 1977) and social constructivism (Pearce, 2009), among other interpretive philosophies, helped to transition the study of listening from an individual to an interpersonal act.

Viewing listening as an interpersonal act is vitally important for behavioral researchers interested in listening. But, closer inspection of the types of behavior covered in Barker's (1971) book reveals each is an internal element of the listening process—the cognitive mechanisms that make listening possible, the barriers to effective listening, the attitudes that define a willingness to listen, the implementation of desirable listening skills, and the individual and situational variables thought to predict variability in listening test scores. Chapter 4 focuses on internal processes enabling comprehension of orally delivered information, and Chapter 5 focuses on motivational forces that increase listening capacity. This chapter proposes a different conceptualization of listening behavior. When people listen, they are not only working on information cognitively but also acting toward another. Listeners are performing a vital role in conversations by enacting various behaviors to convey specific meaning to their interlocutor, behaviors that can be observed and systematically studied. How to study these overt behavioral acts is the focus of the present chapter. Before diving into the issues and design options in

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

behavioral listening research, however, a brief historical sketch of the emergence of research focusing on the behavioral components of listening is provided.

From Listening as Internal to Listening as Overt Behavior

The overt behavioral manifestations of listening were called *feedback* in the 1970s (e.g., Barker, 1971; Lundsteen, 1971; Weaver, 1972). One unintended consequence of using the *feedback* label was that listening scholars largely ignored overt responding in favor of studying the cognitive elements of the listening process. What listeners did after attempting to select, understand, and evaluate was considered "part of a new communication cycle, with the response constituting an initiative of the sender" (Ridge, 1993, p. 7). Weaver (1972), in his own listening textbook, put the case most forcefully: "The listening process concerns *only* the selecting of … stimulus data in order to 'receive it' and the cognitive structuring of it" (p. 6; emphasis added). As a result of conceptualizing listening as solely an internal phenomenon, the literature contains significantly fewer documentations of listening behavior compared to modeling internal elements of the process (Keaton & Bodie, 2013).

A shift from viewing listening behavior as an internal act to an external one was made possible during the 1980s, when K–12 and higher education began stressing teaching and assessing core competencies, skills that students should master prior to graduation. By 1998, these skills were solidified in recommendations by the National Communication Association (NCA; see Table 6.1). Defining *listening* as "the process of receiving, constructing meaning from, and responding to spoken and or nonverbal messages," the NCA was interested in outlining the various skills necessary for competence in listening, organized around several themes: adequately comprehending information, critiquing and evaluating a message, showing empathy for the feelings expressed by others, and appreciating a performance.

Although many of these skills involved internal processes (e.g., remembering details), several involved overt behaviors (e.g., the ability to formulate questions and paraphrase a speaker's message). Moreover, in order to provide assessment of listening competence (even those internal to a listener), it was necessary to observe what listeners did. The assessment of learning objectives became a hot topic in the 1980s, a trend that can be seen in more contemporary movements to train students in "21st century competencies" driven in part by federal funding initiatives (Beard & Bodie, 2014). Today, it is nearly impossible to find a definition of listening that does not include some reference to behavior (e.g., responding; see Table 1.1 in Chapter 1).

Interestingly, contemporary research on what listeners do in interactions provides support to Barker's (1971) claim that listeners are active agents in the communication process, thus bringing the discussion full circle. Listeners have vast influence on the trajectory of conversations and on the outcomes of those conversations, not only because of how they process information but also because of how they act, that is, how they behave as listeners (Bavelas & Gerwing, 2011; Pasupathi & Billitteri, 2015).

From a behavioral perspective, listening is indeed not a passive act, and listeners are not mere receptors of information; they are full collaborators or co-narrators in a storytelling, partners in the meaning-making process (Bavelas, Coates, & Johnson, 2000). Just as Barker (1971), Weaver (1972), and others attempted to stress in the 1970s that listeners wielded great influence over the ways that messages are interpreted and meaning is construed, scholars since the 1980s have been attempting to document how

Table 6.1 List of competencies associated with listening as outlined by the National Communication Association.

In order to be a COMPETENT LISTENER, a person must be able to listen with literal comprehension. Specifically, the competent listener should be able to exhibit the following competencies by demonstrating the abilities included under each statement.

A. RECOGNIZE MAIN IDEAS.

1) Distinguish ideas fundamental to the thesis from material that supports those ideas.
2) Identify transitional, organizational, and nonverbal cues that direct the listener to the main ideas.
3) Identify the main ideas in structured and unstructured discourse.

B. IDENTIFY SUPPORTING DETAILS.

1) Identify supporting details in spoken messages.
2) Distinguish between those ideas that support the main ideas and those that do not.
3) Determine whether the number of supporting details adequately develops each main idea.

C. RECOGNIZE EXPLICIT RELATIONSHIPS AMONG IDEAS.

1) Demonstrate an understanding of the types of organizational or logical relationships.
2) Identify transitions that suggest relationships.
3) Determine whether the asserted relationship exists.

D. RECALL BASIC IDEAS AND DETAILS.

1) Determine the goal for listening.
2) State the basic cognitive and affective contents, after listening.

The COMPETENT LISTENER must also listen with critical comprehension. Specifically, the competent listener should exhibit the following competencies by demonstrating the abilities included under each statement.

A. ATTEND WITH AN OPEN MIND.

1) Demonstrate an awareness of personal, ideological, and emotional biases.
2) Demonstrate awareness that each person has a unique perspective.
3) Demonstrate awareness that one's knowledge, experience, and emotions affect listening.
4) Use verbal and nonverbal behaviors that demonstrate willingness to listen to messages when variables such as setting, speaker, or topic may not be conducive to listening.

B. PERCEIVE THE SPEAKER'S PURPOSE AND ORGANIZATION OF IDEAS AND INFORMATION.

1) Identify the speaker's purpose.
2) Identify the organization of the speaker's ideas and information.

C. DISCRIMINATE BETWEEN STATEMENTS OF FACT AND STATEMENTS OF OPINION.

1) Distinguish between assertions that are verifiable and those that are not.

D. DISTINGUISH BETWEEN EMOTIONAL AND LOGICAL ARGUMENTS.

1) Demonstrate an understanding that arguments have both emotional and logical dimensions.
2) Identify the logical characteristics of an argument.
3) Identify the emotional characteristics of an argument.
4) Identify whether the argument is predominantly emotional or logical.

E. DETECT BIAS AND PREJUDICE.

1) Identify instances of bias and prejudice in a spoken message.
2) Specify how bias and prejudice may affect the impact of a spoken message.

F. RECOGNIZE THE SPEAKER'S ATTITUDE.

1) Identify the direction, intensity, and salience of the speaker's attitude as reflected by the verbal messages.
2) Identify the direction, intensity, and salience of the speaker's attitude as reflected by the nonverbal messages.

(*Continued*)

Table 6.1 (Continued)

G. SYNTHESIZE AND EVALUATE BY DRAWING LOGICAL INFERENCES AND CONCLUSIONS.

1) Draw relationships between prior knowledge and the information provided by the speaker.
2) Demonstrate an understanding of the nature of inference.
3) Identify the types of verbal and nonverbal information.
4) Draw valid inferences from the information.
5) Identify the information as evidence to support views.
6) Assess the acceptability of evidence.
7) Identify patterns of reasoning and judge the validity of arguments.
8) Analyze the information and inferences in order to draw conclusions.

H. RECALL THE IMPLICATIONS AND ARGUMENTS.

1) Identify the arguments used to justify the speaker's position.
2) State both the overt and implied arguments.
3) Specify the implications of these arguments for the speaker, audience, and society at large.

I. RECOGNIZE DISCREPANCIES BETWEEN THE SPEAKER'S VERBAL AND NONVERBAL MESSAGES.

1) Identify when the nonverbal signals contradict the verbal message.
2) Identify when the nonverbal signals understate or exaggerate the verbal message.
3) Identify when the nonverbal message is irrelevant to the verbal message.

J. EMPLOY ACTIVE LISTENING TECHNIQUES WHEN APPROPRIATE.

1) Identify the cognitive and affective dimensions of a message.
2) Demonstrate comprehension by formulating questions that clarify or qualify the speaker's content and affective intent.
3) Demonstrate comprehension by paraphrasing the speaker's message.

Source: National Communication Association.

observable behaviors influence the way that conversations are structured and whether they have positive or negative consequences for interlocutors.

As stated previously, behavioral listening research is not as common as self-report methodology or assessments such as the Watson-Barker Listening Test (WBLT; see Profile 64). Perhaps one reason for this state of affairs is the costs of behavioral research compared to other methods. It is far less time and labor intensive to collect a battery of self-report scales or to administer tests of listening comprehension than it is to video-record conversations or group discussions. Indeed, behavioral listening research raises extensive logistical issues. Not only do researchers have to choose the context of listening (e.g., conflict, support, or initial interaction), but also they must consider whether these interactions will involve strangers, acquaintances, friends, or romantic partners; how long the conversation will last; whether to assign participants to roles or let the conversation unfold in a more naturalistic manner; whether to observe interactions in a structured or relatively unstructured environment; and the list goes on. Before these decisions, the researcher has to have the capacity for data collection—minimum requirements include audio–video equipment and recording software (or numerous trained observers), laboratory space (or the ability to capture dialogue as it happens outside the lab), and research assistants; many studies further necessitate monetary compensation of participants. In addition, although it may take only a few weeks to adequately sample for a self-report or assessment study, collecting behavioral data takes several months or years depending on the scope of the project. Likewise, although self-report and assessment data are easily analyzed using readily available statistical packages, behavioral data have to be coded, transformed, or otherwise handled in line with specific theoretical and practical purposes. Decisions relevant to this latter issue are not easy to make, especially when research interests go beyond readily available coding rubrics or established rating scales.

Even so, behavioral data are rich and can offer insights not afforded by other methods. Presently, our knowledge of listening comes primarily from work interested in aural information reception and from work that asks people what they think they do as they listen, leaving us with a simplified and perhaps erroneous view of how listening actually works in our everyday conversations and the impact it likely has on important outcomes. Attending to how the listener contributes to dialogue shifts the notion of listener as a passive recipient and retainer of information to an active constructer of meaning, much like early work on cognitive models of the listening process (see Wolvin, 1989; also, Chapter 4). In general, opening the field of listening research to include overt responding allows scholars and practitioners greater insight into the importance of listening to everyday life. Just as methods aimed to study the cognitive and affective domains of listening, however, behavioral methods involve a specific set of tradeoffs and choices and pose concerns about validity, reliability, and feasibility.

Defining Behavior

For purposes of this chapter, listening *behavior* is defined as something that individuals (or dyads or larger groups) *do* that can be observed by others. Although many different "skills" are listed as important for developing competence in listening, those that qualify as behavior are normally organized under the label of *response* or *responding*. For instance, Ridge (1993) listed the following as "listening skills of responding (R):

- Asking questions
- Giving appropriate feedback commensurate with purpose of speaker
- Responding in consonance with speaker/situation/mood
- Withholding preparation of response until speaker has finished
- Paraphrasing or checking back for understanding" (p. 7).

Similarly, Brownell (2013) listed several verbal and nonverbal components of a skillful listening response, including perception checking, avoiding "you language," expressing feelings using nonconfrontational language, appropriate eye contact, vocal pleasantness, and using gestures to add emphasis to particularly important words.

As you listen to another person, you do a variety of things like nod in agreement (or disagreement), say "Mhm" and "Yeah" to encourage continued disclosure, ask questions, and tell reciprocal stories. As you listen to the radio, you move your body to the music or shout out loud "No way!" when you hear something surprising. When relaxing to the sounds of a sunset, you might close your eyes and lie on your back to take in all the available soundscape. All these are listening behaviors because they are outwardly observable by others.

An important implication of this definition is that listening behaviors cannot be captured by asking respondents to self-report their tendencies. Although self-reporting of listening can provide insight into what people think they do (or perhaps their motivation to act in particular ways; see Chapter 5), the only way to validly capture what people do is to observe them acting. Such observation can be "live" or recorded for later analysis, and recording can be more or less obtrusive. And whether behavior is observed live or recorded for later use, the researcher must make several choices with respect to how to reflect those behaviors (e.g., whether they are measured or discussed in a more narrative manner). In service of arming you with the necessary tools to design a behavioral listening study, we cover these issues in depth in the following sections of this chapter.

Designing a Behavioral Listening Study

When designing a behavioral study, first consider "how much and what kind of structure to impose" (Sillars, 1991, p. 199). In an ideal world, we would be able to unobtrusively observe people as they listened in all domains of life (like a real-life *Truman Show*) and draw conclusions that have a high degree of realism and direct applicability to those domains. Ethnographic research is especially well suited to observing listening behavior in situ (see Chapter 3), although behavioral listening scholars more commonly audio- and/or video-record participant actors in either their natural environment (e.g., Imhof, 2008) or a staged environment (e.g., Bodie, Jones, Vickery, Hatcher, & Cannava, 2014; Bodie, Vickery, Cannava, & Jones, 2015).

Natural and Laboratory Observation

For many reasons, studying listening behavior in situ is the ideal. Take, for example, the claim that listening is the primary way that children acquire language: that by listening to expert language users, children learn the meaning, pronunciation, and usage of words. One way to substantiate this claim is anecdotally; that is, you can think about the caregiver–child relationships in your own life and draw the conclusion that the children must learn the language by listening because (a) they cannot read at such an early age, and (b) they use words in ways that very much mirror their parents' usage. Much of what you know anecdotally is further substantiated by cross-sectional and longitudinal research that has discovered patterns of language production and acquisition (VanPatten, 2014).

A rather extensive longitudinal study was conducted by Drs. Deb Roy and Rupal Patel, who began recording their life as new parents when they came home from the hospital with their first child (Roy *et al.*, 2006). Recording was assisted by a system of 11 cameras and microphones that captured continuous video and audio in most rooms and hallways of their house (no bathrooms!). In all, they amassed approximately 230,000 hours of audio–video recording that represented the first 3 years of their child's life. Using high-powered computers and mathematical modeling procedures, their research team has been able to track the progression of language use (articulation, word formation, etc.) and map this progress to how the child and his caretakers moved throughout the house (e.g., the word *water* was primarily used in the kitchen, and the word *bye* was primarily used by the front door).[1]

1 Readers are encouraged to watch Dr. Roy's TedTalk, "The Birth of a Word," found here: https://www.ted.com/talks/deb_roy_the_birth_of_a_word?language=en.

As extensive as this project was, however, there were still choices. Most obvious is that these data came from a single household and may or may not generalize to how all children in all situations across all time and in all cultures acquire language. For instance, Drs. Roy and Patel are both highly educated and employed a full-time nanny to look after their son. Perhaps language acquisition is dependent on structures made possible by the child spending most of his first 3 years in the home as opposed to, for instance, in daycare centers. Another choice centered on how to record speech and movement. The research team decided to record 8–10 hours per day, not the full 24-hour period, and they chose to record in a subset of the rooms in the house, not the entire house. Moreover, the cameras were installed on the ceiling, providing a bird's-eye view rather than close focus on individual faces or bodies. When data were analyzed, the research team had another set of choices, namely, on what behaviors to focus and how to represent those behaviors. In general, then, although their data collection effort is intense and impressive, it did not fully capture life-as-lived but instead represented a (very large) sample of behavior that was used to create general models of language acquisition, at least for one child in one particular caretaking situation (that of a first-born son of a heterosexual married couple, both of whom had full-time, professional jobs and employed a nanny).

One of the more impressive aspects of this data collection effort, aside from the 200 terabytes of data, is that behaviors were observed in their natural setting, in this case in the very environment where the child was learning language. If you are interested in, for instance, how people spend their time listening, it would be advantageous to actually observe people listening in their daily lives (perhaps in addition to reporting general tendencies; see Time Studies, Profile 60). But even studies that capture naturally occurring events make choices; you cannot observe people every waking moment of their lives, and if participants know you are observing they may possibly refrain from behaving in certain ways or change their "natural" behavior patterns to conform to your presence (i.e., the Hawthorne effect; Landsberger, 1958). In the case of the Human Speechome Project spearheaded by Drs. Roy and Patel, all participants knew they were being observed, and issues related to privacy and informed consent had to be addressed. For instance, cameras were controlled by a wall-mounted device that allowed tenants to turn on or off certain cameras and microphones. Moreover, although their data can provide great insights into antecedents and consequences of language learning, they are immensely complex. Decisions regarding transcription and annotation as well as how to represent body movement are all covered in this chapter, but it is important to state here that this research team took approaches to these issues that are far from universally appropriate for all studies, even for large-scale studies such as theirs.

Imagine, for instance, that someone on the research team is interested in how parents provide emotional support to their children. An initial glimpse at these types of interactions can be garnered from these data, but questions about specific listening contexts like support are often aided by observing behavior in more structured settings. For one thing, emotional support is a potentially low-occurring behavior; that is, its provision may only happen in 10% or fewer of the interactions we have on a daily basis (Goldsmith & Baxter, 1996; Mehl, Vazire, Holleran, & Clark, 2010). And even when it does occur, what happens if interlocutors do not agree that the conversation was "about support"—do you, as the researcher, define the conversation in those terms regardless? Much of our daily interactions are more mundane or superficial in nature than substantive (Mehl *et al.*, 2010). Moreover, many types of listening often occur in conversations that "rapidly shift topic, [where] people come and go, discussions are interrupted by phone calls and small emergencies, and numerous other irregularities occur" (Sillars, 1991, p. 201). Although these nuances of everyday conversation are interesting and should be recognized and studied systematically, the nature of some research questions necessitates control of observations for comparative purposes.

One example of a laboratory paradigm for behavioral listening research profiled in Section Three is the Couples Helping Exercise (CHE)—a method that helps to structure supportive conversations. To summarize, the CHE requires two participants who are asked to talk about one or more stressful events. Participants can be assigned randomly to the role of either problem discloser (the person whose problem is discussed) or supportive listener (the person who attempts to help the discloser cope with his or her problem) (see Jones & Wirtz, 2006). Alternatively, one of the participants can be instructed to "introduce a problem topic into conversation" without preassigning clear roles of "provider" and "recipient" of support (Goldsmith, 2004, p. 120). Participants can know each other well, as when researchers recruit married, cohabiting, or dating partners (e.g., Afifi, Afifi, Merrill, Denes, & Davis, 2013; Collins & Feeney, 2004), or they can be total strangers (e.g., Bodie, Vickery, *et al.*, 2015; High & Solomon, 2014). If preassigning roles, researchers can train one or both of the individuals to disclose problems in particular ways (e.g., to uncover whether listeners change strategy as a function of disclosure style; see Keaton, Bodie, & Keteyian, 2015) and/or to listen in ways that have been theorized to be more or less responsive to the needs of stressed individuals (e.g., Jones & Guerrero, 2001). The conversation that ensues can be limited to last 5 minutes or can be allowed to continue for 20 or 30 minutes or longer. Most of the time, conversations are video-recorded for later transcription and analysis. After the interaction, participants are asked to complete evaluations of their own and/or their partner's behavior.

As the CHE illustrates, observing behavior in a laboratory setting is far from a heterogeneous activity. Just as there is no one natural setting from which we can generalize about listening behaviors, there is no one naturalistic laboratory setting either.[2] Researchers must always make a principled (and justifiable) decision to restrict the observations of listening behaviors within particular contexts and settings. Although laboratory work is more structured than observations outside the lab in many ways, there are still various grades of structure that can be applied to lab work. At the high end of the structure spectrum are experimental studies that constrict the behavior of one or more of the participants, usually the listener, to observe how these manipulations affect the conversation and the other interlocutor, usually the discloser.[3] For instance, Jones and Guerrero (2001) trained advanced undergraduate students to enact three forms of verbal person centeredness and three forms of nonverbal immediacy, creating nine cells (3×3) in their experimental design. These listeners interacted with several participants who thought they were disclosing a stressful event to another student participant (i.e., they were unaware of the nature of the experimental manipulation). Results showed that listeners who validated expressed emotions (high person-centered comfort) with a high degree of nonverbal warmth (high nonverbal immediacy) helped disclosers feel better about their problematic

---

2 The terms *natural* and *naturalistic* are being used purposefully and not interchangeably here. Naturalistic means "imitating nature." In many ways, laboratory research can imitate conditions of the natural environment, and some work that observes behavior in natural settings is more "naturalistic" than "natural."
3 Experimental methods are not covered in depth in this chapter. The interested reader is directed to read, Cook, Shadish and Campbell (2002); Smith (2000); Keppel and Wickens (2009); and Kerlinger and Lee (2000, Chap. 24).

events, thus helping to replicate other research that has used hypothetical designs to study the role of listener behaviors in the coping process. In other words, listening styles or strategies seem to influence how people report and feel about personal problems.

Just because work is conducted in the lab, however, does not mean it is experimental in nature; that is, some laboratory work refrains from manipulating the behavior of any of the participants. For instance, Bavelas, Gerwing, Healing, and Tomori have created a paradigm for the study of how listeners co-construct close-call stories, a personal event that has a surprising twist but that turns out positive in the end (see the microanalysis of face-to-face dialogue [MFD] profile, Profile 42). Two unacquainted individuals are recruited and assigned to either disclose or listen to a close-call story. In a frequently cited example, the "Sleeper Story," a woman tells the story of falling asleep but leaving her reading lamp on. Although clamped to her headboard, the lamp came loose and landed on her pillow, catching it on fire. Her room filled with smoke, waking the woman and leading her to play the role of firefighter. Neither the listener nor the storyteller is told how to behave, and thus behaviors mimic naturalistic dialogue (i.e., unscripted behavior that is allowed to freely vary). Other work using this paradigm does involve some level of manipulation, assigning some listeners to count the number of words uttered by the storyteller that begin with the letter *t* while they simultaneously try to be attentive. These "t-counting" experiments have shown that distracted listeners cause various problems in the storytelling process (for review, see Bavelas & Gerwing, 2011).

All work using the CHE and the work on close-call stories is structured in the sense that it calls for the participants to have a particular type of conversation, one that may or may not occur outside of the lab exactly as it occurs inside the lab. If you reflect for a moment on the types of conversations you have with others, it is likely that you have told some close-call stories and that you have disclosed one or more problematic events. But do your conversations share anything in common with how these conversations occur in the lab? Work by Jefferson (1978, 1980), for instance, has shown that troubles talk (talk about problems) occurs in the context of talk about more superficial matters; that is, when we talk about our problems with others, we often change topics, waxing and waning in and out of talking about one particular stressful event. Consequently, asking people to talk about problems may feel "strange" or "unnatural" and thus may somehow change different aspects of those conversations when they occur in the lab. A method developed by Ickes and colleagues (see the Empathic Accuracy, Profiles 17 and 18) attempts to mimic how people talk about more superficial topics. After recruits are seated, the experimenter invents a reason to excuse him or herself from the room (e.g., to run an errand or check on the equipment), leaving participants to converse with little or no direction. Thus, the observations one might make with data generated with this "unstructured dyadic interaction paradigm" might be closer to what happens in "get to know you" conversations than if participants were asked to "get to know each other" (see McLaughlin, Cody, Kane, & Robey, 1981). Inventive readers will likely now be thinking of ways to incorporate this basic idea into laboratory paradigms that seek to foster supportive or other types of conversations that can more closely mirror how such conversations might happen in the proverbial "real" world.[4]

---

4 For instance, in a study of social support, Derlega, Barbee, and Winstead (1994) introduced participants to a common stressor, a public speaking task, then asked them to sit in a waiting area. Soon after they sat, another participant (who can be a friend, romantic partner, confederate, or untrained stranger) entered the room and sat down. Because they are stressed, participants are likely to begin talking about their stressor. It is possible that if the conversation continues long enough, the participants will shift in and out of troubles talk.

Use of Recording Equipment to Capture Behavior for Later Analysis

Regardless of whether you decide to capture behavior in its "natural" environment or to use some form of laboratory setup, your next decision is whether and how to capture behavior for later analysis. Researchers in the conversation analysis tradition typically rely on audio-recorded conversations that are then transcribed using standard rules (discussed further in this chapter). Researchers who employ other discourse analytic procedures rely on both audio- and video-recorded conversations, each of which poses challenges depending on the nature of the setting. For instance, if you are interested in how students listen in a classroom setting, various institutional review board and other logistical challenges might present barriers to access or to recording. At the least, the school and teacher will have to give permission, and in many cases each parent will have to provide permission to use data gathered from their child(ren). Thus, many classroom observations rely on standard rubrics that can be consistently marked by trained clinicians or hired research assistants (see the "Coding and Rating Behaviors" section for a discussion of coding rubrics; for an exemplar study, see Imhof, 2008). By not recording in a classroom setting, you can bypass many of the logistical hoops necessary for working with "special classes of subjects."

Although exceptions exist (e.g., Roy *et al.*, 2006), recording is generally made easier in the laboratory setting (perhaps indicating one reason why researchers decide to bring people to the lab as opposed to traversing their natural environments). At the very least, the lab setting should be equipped with recording equipment that allows the researcher to capture all individuals involved in an interaction (or other task). One example of a laboratory setup is provided in Figure 6.1.

There are three distinct rooms in this configuration. Moving from left to right on Figure 6.1, the interaction room (items 1–5) is where participants are seated and in which they interact. Ickes's team has made a decision to hide the recording equipment in this room, allowing for unobtrusive video recording of the interaction. Whether to record in an unobtrusive manner depends on the design. For Ickes and colleagues, their

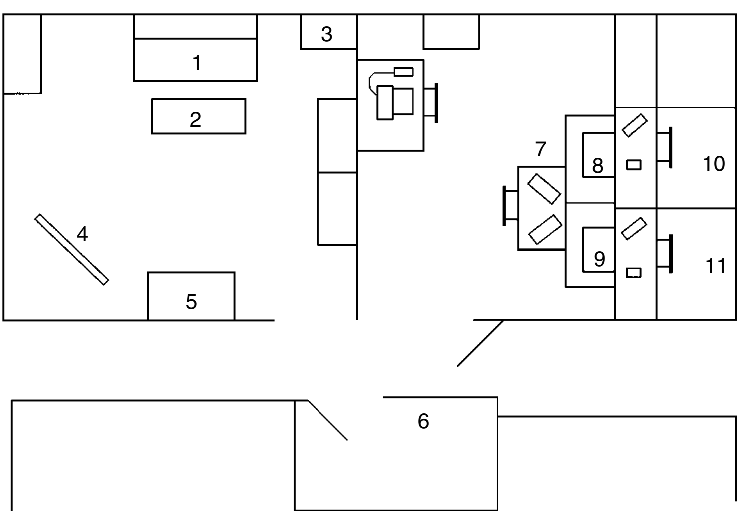


Figure 6.1 Laboratory setup used in the Unstructured Dyadic Interaction Paradigm of Ickes (see Profile 18). Used with permission.

backstory of the project is such that if participants knew they were being recorded, it is possible that their "unstructured interaction" would be less natural. Other work that informs participants of recording equipment finds that people acclimate to its presence rather quickly. The next room (item 6) is an observation room, in which the experimenter can monitor the interaction. The final room (items 7–11) contains equipment for postinteraction interviews and video playback. In their paradigm, Ickes and colleagues ask participants to watch their interaction and answer questions about that interaction. Other equipment for this room includes various software for recording survey data (e.g., impressions of the conversation) and debriefing stations.

Depending on the type of research you are conducting and the detail needed, your equipment needs may be more or less than those of some other research team. For audiotaping, several options are available. Relatively inexpensive digital recorders can be used, although the placement of the recorder may make it difficult to interpret portions of a conversation that are whispered or soft-spoken. Lapel microphones can help to ensure that more of the spoken language is captured, but these can be more expensive and potentially cumbersome for participants. For behavioral listening researchers wishing to make claims about a broader range of listening behaviors, the obvious choice is to video-record. With video-recorded interactions, researchers can watch (and rewatch) in order to code or rate various behaviors given concerns of a particular research project. Video recordings allow access to not only verbal listening behaviors but also nonverbal behaviors. If the choice is to video-record, however, the researcher must decide how many cameras to use and, if multiple cameras are used, where to position these cameras. If choosing to conduct microanalytic techniques, at least two cameras are needed, one for each interlocutor (see MFD, Profile 42). If your project requires, for instance, measuring facial movement in fine detail, the camera equipment will need to be higher quality than if you are only interested in general body orientation. In the concluding section of this chapter, we introduce several new developments including technologies that allow researchers to map facial contours as well as body position with over two dozen points of precision and those that allow for textual analysis of large datasets. In general, when choosing to video-record, it is often more convenient to conduct the study in a laboratory setting, although several companies offer portable lab solutions that, with advances in technology, make video recording away from the lab rather seamless.[5] Finally, it is also advisable to seek out readily available datasets prior to conducting a study from scratch. For instance, the Santa Barbara Corpus of Spoken American English contains several hours of audiotaped conversations (nearly 250,000 words).[6]

Representing Behavior That Is Captured

If you decide to capture behavior with some form of audio or video recording, the next decision addresses how the behaviors should be represented. One option is found in the work of Bavelas and colleagues, who watch the video-recorded interactions at slow speeds, sometimes for several hours (even for a 30- to 60-second interaction).

5 For example, Noldus has created a portable laboratory setup that can be found here: http://www.noldus.com/human-behavior-research/solutions/portable-observation-lab.

6 The Santa Barbara Corpus is managed by John DuBois and currently archived here: http://www.linguistics.ucsb.edu/research/santa-barbara-corpus#Intro.

Their microanalysis of face-to-face dialogue requires minimal transcription; indeed, Bavelas's team only transcribes selections of conversations for illustrative purposes after data are analyzed. Other work, however, requires transcription prior to data analysis. Transcription rules are as varied as research traditions, with conversation analysts operating from the most standard set of rules that require details on pausing, inflection, overlapping speech, and the like.[7] Other ways of transcribing require only that words be represented (e.g., see Language Style Matching, Profile 29). Regardless of how you decide to represent behavior, it is important to note that these decisions should be acknowledged and questioned.

Thinking About Listening Behavior in Concrete or Abstract Terms

When a listener responds to a speaker (i.e., provides "feedback" in Barker's [1971] terms), he or she enacts specific behaviors. A wink to indicate "Just kidding" or "I'm with you" or a smile to signal "I know, ridiculous, right?" are just two examples of specific behaviors that have the potential to communicate vast amounts of meaning. Listening responses can be seen (e.g., a wink), and they can be heard (e.g., saying "Mhm" to indicate interest). Regardless of whether the response is voiced or is a bodily action, all listening behaviors are concrete actions and thus lie on a basic or micro level of abstraction. In a sense, behaviors all exist at the nominal level—they can either occur or not (see Chapter 2; and MFD, Profile 42).

Microscopic skills such as asking questions or maintaining eye contact cluster into more mesoscopic skills that exist at higher levels of abstraction (Spitzberg & Cupach, 2002). So, for instance, training people to maintain a certain level of eye contact is usually in the service of helping them appear more attentive or friendly in conversation, that is, to assist them in developing more general listening competencies (Wolvin & Coakley, 1994). Figure 6.2 presents a model illustrating how specific listening behaviors (microlevel skills) might map onto more abstract skill clusters (mesolevel skills). The microscopic skills, represented by boxes on the outside of the middle rectangle, are differentially related to the five mesoscopic skills represented by ovals (attentive, responsive, etc.). These mesoscopic skills can be thought of as attributes that contribute to even more general impressions of listening competence, a macroscopic skill set (see Bodie, St. Cyr, Pence, Rold, & Honeycutt, 2012).

For example, consider nodding one's head or saying "Mhm" and "Yeah" (backchannel responses) in response to something interesting. When listeners enact these behaviors, their conversational partners make abstract judgments of them, such that listeners are labeled "friendly" and "attentive" (Bodie *et al.*, 2012). These more abstract judgments, in turn, cause speakers to infer that the listener is a "good" (or competent) listener (and/or potentially a competent communicator or socially skilled individual more generally; see Bodie, Pence, *et al.*, 2015).

---

7 The Department of Linguistics at the University of California, Santa Barbara, maintains a useful website that includes a bibliography and other sources for those interested in reading more about transcription rules. See http://www.linguistics.ucsb.edu/projects/transcription/.

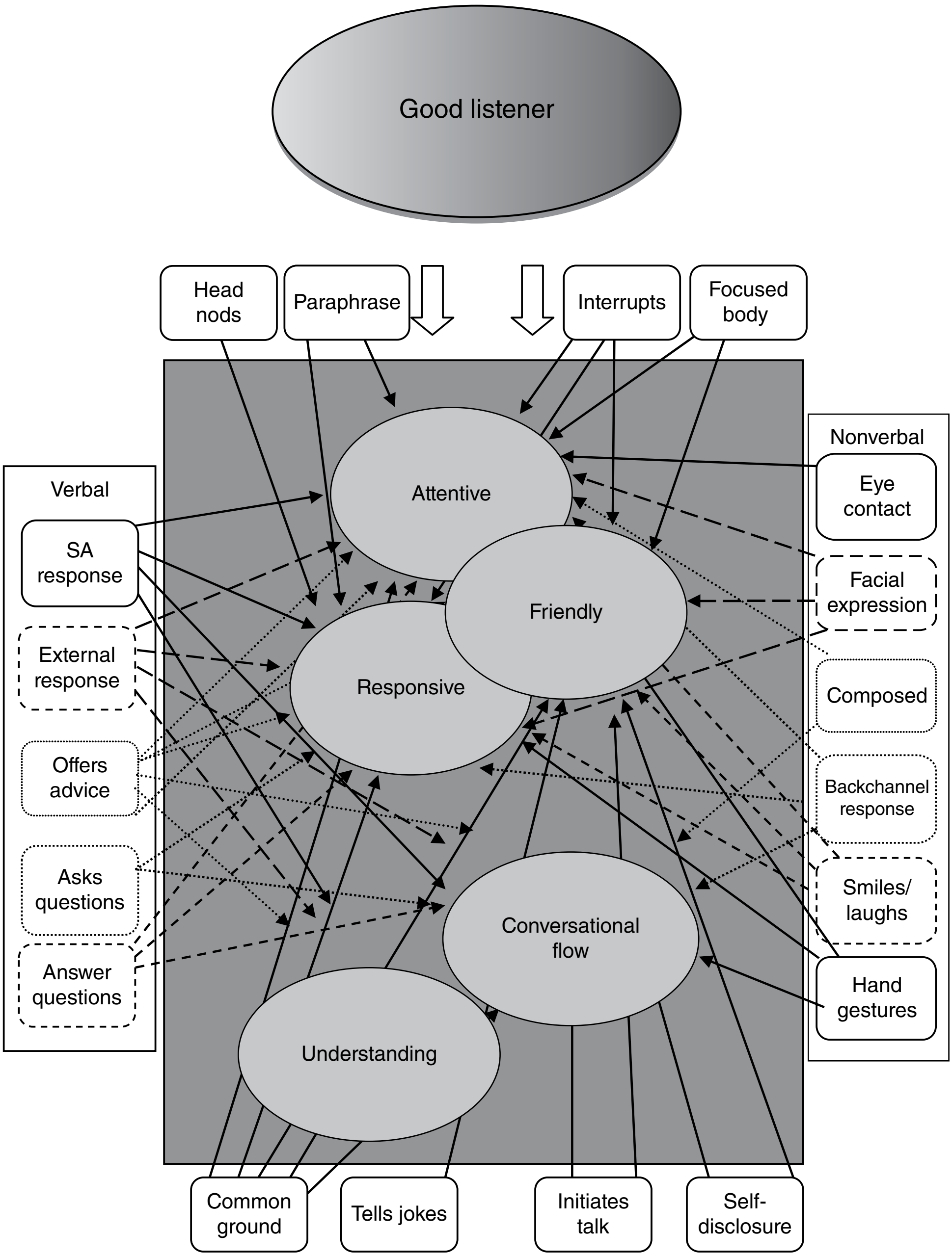


Figure 6.2 Graphical depiction of an implicit theory of listening, from Bodie *et al.* (2012). *Source*: Bodie (2012). Reproduced with permission of Taylor & Francis.

The consistent enactment of competent listening is thought to serve a variety of functions to individuals, relationships, and society at large (Bodie, 2012). The functions of competent listening, like the skills that comprise it, can also lie at different levels of abstraction. Examples of microscopic functions include immediacy, empathy, support, and relaxation; mesoscopic functions include intimacy, relational satisfaction, and openness. At the highest level of abstraction, "skills can be employed to move with or toward another person … to move away from another person … [or] to move against another" (Spitzberg & Cupach, 2002, pp. 589–590).

Because listening behavior occurs at a microscopic level of abstraction with implications at higher levels (in both form and function), choices have to be made with respect to the specific behaviors that will concern any individual study as well as how these behaviors are represented on measures designed to capture variability in their occurrence. These two choices are detailed in this section.

Choice 1: Sampling Behavior

Just like we must sample participants from a larger population of people or scale items from a larger population of possible representations of a latent construct (see Chapter 2), behavioral listening scholars must make sampling decisions with respect to: (a) which behaviors to include from a larger population of possible actions of interest, and (b) the level of abstraction at which to capture behavior.

Which Behaviors to Include?

Even a cursory look at behavioral research shows great variability in which behaviors are the focus of a particular study. Although some studies focus on one or a small number of specific behaviors like mutual eye gaze (e.g., Bavelas, Coates, & Johnson, 2002), other studies focus on classes of behavior like nonverbal immediacy (see Profile 47). The choice of which behaviors to study is, like a definition (see Chapter 1), a theoretical one, contingent on the specific interests of the study under consideration. Researchers must develop a rationale for the inclusion (and exclusion) of specific listening behaviors. A recent example from research conducted in the Louisiana State University Listening Lab (3L) is used to illustrate this process.

The 3L research team developed an argument for exploring a set of behaviors labeled *active listening* in the context of people talking about stressful events (Bodie, Vickery, *et al.*, 2015). Thus, their study was concerned with a sample of possible behaviors relevant to one particular type of conversation. For this particular manuscript, they chose to focus on a cluster of behaviors known as *active listening*. In particular, given the context, the 3L team needed to choose behaviors that best map onto the mesoscopic and macroscopic skills and functions of supportive conversations. Figure 6.3 illustrates how these behaviors reflect the micro, meso, and macro levels of abstraction.

Bodie, Vickery, *et al.* (2015) focused on two mesoscopic skills, verbal and nonverbal. In terms of nonverbal behaviors, they focused on nonverbal immediacy (NVI) and represented this emphasis with nine microscopic skills—four *facial cues* (smiling, eye contact, head nods, and facial pleasantness), four *body cues* (forward lean, body orientation [open/closed], body orientation [toward/away], and gestural animation), and one *vocal cue* (vocal pleasantness). Active listeners also signal attentiveness through a range of verbal behaviors, the most common of which are paraphrasing, reflecting feelings, assumption checking, and asking questions.

The decision to limit the analysis of active listening to these specific behaviors was driven by theory. All of these behaviors are proposed by other scholars (and by most textbooks that cover supportive communication) as highly relevant to supportive conversations. They are thought to influence how individuals think and feel about everyday

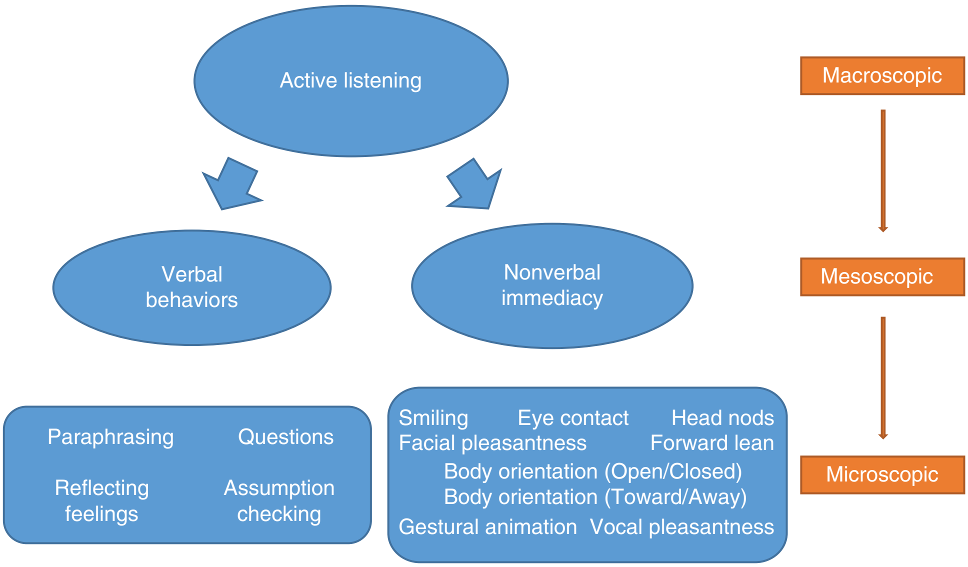


Figure 6.3 Conceptual map of "active listening" as composed of specific microscopic "verbal" and "nonverbal" actions.

stressful events, in particular by helping to create a warm environment conducive to opening up in a way that can help foster the reappraisal of problematic events and concomitant emotions (Burleson & Goldsmith, 1998; Jones & Wirtz, 2006). The explicit purpose was to provide an empirical test of the active listening paradigm in the context of talking about everyday stressors, a purpose that ultimately dictated the behaviors up for analysis.

Behavior at What Level of Analysis?

In addition to the specific behaviors analyzed in the active listening study, we also chose to focus on behavior produced by only one individual in the interaction. Other studies have focused on both interlocutors. For instance, we analyzed the co-occurrence of nonverbal immediacy behaviors produced by both individuals in the interaction, finding that disclosers who matched highly immediate listeners reported feeling better about their problems (Bodie, Cannava, Vickery, & Jones, 2016). Matching can be considered a behavior that occurs at the level of the dyad; although matching is made up of behaviors enacted by individuals, it only occurs if there are two or more people in interaction. Other behaviors such as the demand–withdraw pattern of marital conflict (Caughlin, 2002) or reciprocity and compensation (Burgoon, Stern, & Dillman, 1995) also occur at the level of the dyad. Whatever the decision, it is imperative to justify it in a way that makes clear the focus of behavioral coding and analysis.

Choice 2: Representing Behavior

After the choice is made to focus on a particular set of behaviors, the next decision has to do with representing behavior: Exactly how do we measure the occurrence, strength of operation, and other features of these behaviors? Do we simply count them? Or is it better to rate each behavior in some way? If the latter, do we use the same scale for all

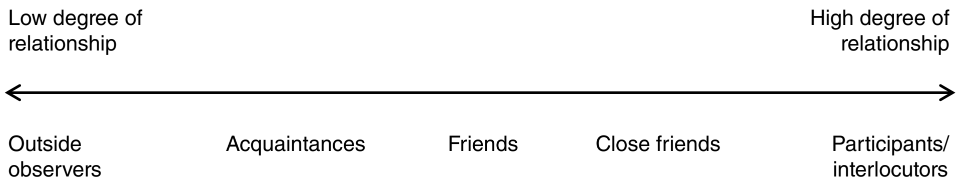


Figure 6.4 Continuum of judge "insiderness" based on the closeness between the judge(s) and actor(s).

behaviors or create separate scales? In either case, who gets to assess the behavior? Behavioral researchers will face several issues related to representing behavior. We address some of the most notable here.

Behavior from Whose Perspective?

Although Herman Hupfeld (1931) in "As Time Goes By" asserted, "a kiss is still a kiss, a sigh is just a sigh," it is more accurate, though much less poetic, to say that the same behavior can take on myriad meanings depending on the perspective from which it is judged. A primary decision facing behavioral researchers is, who gets to offer judgments of enacted behaviors? Are participants asked their opinions or to retrospectively recall the frequency, duration, or patterns of behavior (see Time Studies, Profile 60)? Or does the researcher attempt to train more "objective" coders or raters to make judgments from live or video-/audio-recorded interaction? Alternatively, friends or acquaintances of the interlocutors, or maybe untrained strangers, should judge the interaction (either live or from recordings).

Most basically, behavior can be judged from either an "insider" or "outsider" perspective, a dichotomy Levinger (1977) replaced with a continuum of "insiderness" or "varying degrees of distance and closeness" in the relation between the observer and observed (p. 151). Complete insiders to enacted behavior are the individual interlocutors, the participants in the conversation. Complete outsiders are those with relatively low degrees of relationship to the interlocutors. Between these two extremes, moving from low to high degrees of relationship, are acquaintances, friends, and close friends (Surra & Ridley, 1991; see Figure 6.4).

Observers have more or less insider status based on their closeness with the interacting participants. Some observers have more insider knowledge of the interacting individuals, whereas other observers are relatively unfamiliar with the people producing behaviors. For instance, a couple who has been married 20 years will have more insider knowledge about how each person typically uses language or about what each person typically means when certain phrases or bodily actions are used than a dyad composed of people who have just recently met.

In the study of supportive listening mentioned in the "Natural and Laboratory Observation" section, the team chose to train outside judges, advanced undergraduate students enrolled in an independent study course. Questions driving the study were the extent to which the occurrence of active listening behaviors had an impact on impressions of the conversation and on the coping potential of the discloser, not whether perceptions of those behaviors (how listeners or disclosers rated behaviors) correlated with perceptions of the conversation (how listeners or disclosers rated conversations). Research that asks participants to make judgments of their own and their partners' behavior typically finds that these judgments carry a heavy load of idiosyncratic and subjective information (e.g., Priem, Solomon, & Steuber, 2009). Indeed, we perceive behavior from a variety of ingrained cognitive structures that are built up over time in our relationships (Honeycutt & Cantrill, 2001). As we evolve in relationships with others, we learn to expect certain behaviors from them and to evaluate their behaviors in light of other relational knowledge. Often referred to as *bias* in perception (Kenny & Acitelli, 2001), judgments we make of behavior are colored by how well and in what ways we know the actor, and this applies whether we are judging behaviors directed at us or whether we are judging an interaction between two known others.

When observers who are completely outside of the relationship between the interacting others (e.g., trained or untrained coders or raters) are asked to make judgments of an interaction, these judgments are likely driven primarily by cultural scripts and general knowledge of the interaction context (e.g., observing a conflict or supportive interaction will likely frame judgments in different ways). People with no relational knowledge of the interacting participants "have no unique subjective knowledge about the relationship they are observing, [but that] does not mean that they have no knowledge about it at all" (Surra & Ridley, 1991, p. 40). Outside observers are "cultural informants who presumably understand the meaning of behaviors because they use the same cultural and social filters to interpret them as participants use" (Surra & Ridley, 1991, p. 41).

One important implication of choosing one perspective over others is that study results may be (wholly or partially) a function of the perspective chosen to represent the behavior. Using data from the same active listening conversations, the 3 L research team published an additional article that showed sufficient discrepancy between ways of representing active listening behaviors (Bodie *et al.*, 2014). In addition to asking trained undergraduate outsiders to assess active listening behaviors, we also gathered measures of active listening from the two insiders of the interaction. Listeners were asked to self-report how they think they typically listen. Each discloser was asked to judge the active listening behaviors of his or her listener immediately after the interaction. Results showed that these three ways of assessing active listening exhibited only a small degree of correspondence (see Chapter 5).

How does the Judge do the Judging?

Related to the choice of who gets to make the judgment is how the people, whoever they are, will be judging. The most cost- and time-efficient way to assess behavioral judgments is to ask individuals to retrospectively recall their (or a recent conversational partner's) behavior. This approach was taken by Lynn Cooper and her colleagues as they developed the Organizational Listening Survey (OLS; see Profile 50). The OLS asks participants to think about their listening behaviors in general, and most studies also ask some other person or group of people like managers or coworkers to provide impressions of the same individual's listening behavior. When self-reported data are compared with impressions gathered from coworkers, subordinates, or managers, results show a large degree of discrepancy (very similar to what we find when coding behaviors from multiple perspectives). Other scales that have been used as self-reported behavioral tendencies that are profiled in Section Three include the Active-Empathic Listening Scale (AELS), the Self-Perceived Listening Competence Scale (SPLCS), the Interaction Involvement Scale (IIS), and the Conversational Sensitivity Scale (CSS). Of course, as already mentioned, self-reports of behavior merely represent what people think they do, not what they actually do. The only way to measure listening behavior, at least as presently defined, is to capture observations of people as they engage as listeners (e.g., in video-recorded interactions).

When observations are made, they can be cast at various levels of measurement (see Chapter 2). Although behavioral enactment is technically a nominal-level variable (i.e., present or absent), other facets of behavior occur at ordinal, interval, and ratio levels. For instance, although whether a person looks up while talking can be cast as *yes* (1) or *no* (0), more complex ways of representing glancing include whether glancing is of a particular type (e.g., sustained, brief, or mutual) and the length of time the gaze is held (a ratio variable that can be represented in, e.g., milliseconds). All of these decisions are complex, to be sure, but they can be generally cast as a choice between coding and rating.

Coding and Rating Behaviors

Both coding and rating represent particular views from which behavior can be assessed, and each involves a theoretical choice on the part of the research team. The term *coding* is commonly reserved for the process of classifying visible behaviors into nominal or ordinal categories. Coding is conducted at a granular level, most often focusing on the frequency, duration, and patterning of specific microscopic behaviors. When coding, the concern is with the manifest contents of behavior, what the individual actor did or said, rather than what the individual intended or meant by his or her action (see VRM profile, Profile 63).

Whereas coding provides a microlevel focus, rating is a slightly more abstract measurement technique. Rating rules indicate "the approximate quantity of behavior within the segment" and are often defined along scales from *never* to *always* or from *none of the time* to *all or most of the time*. In the active listening study introduced in the "Choice 1: Sampling Behavior" section, all behaviors (all nine nonverbal immediacy cues and four verbal responses) were rated, not coded. The nonverbal immediacy cues were rated using a modified version of Andersen, Andersen, and Jensen's (1979) nonverbal immediacy scale (see Profile 47), and the verbal cues were rated using a modified version of the Active Listening Observation Scale (see Profile 4). For each set of behaviors, although observers watched the videotaped interactions and made decisions based on the enactment (or lack thereof) of specific, microscopic actions, their rating was not an aggregate of frequency counts but a gestalt impression of that behavior.

Training Coders/Raters

To accurately represent the degree of nonverbal immediacy and verbal active listening in the supportive listening data, it was important to have highly skilled and trained raters. Research assistants were upper-level communication studies majors who had been exposed to supportive communication research and the behaviors that make up active listening. These individuals were enrolled in an independent study course under the advisement of a principle investigator. Each rater was guided through approximately 4 hours of training over two meeting sessions that included: (a) a theoretical discussion of the relevant construct (e.g., immediacy, paraphrasing, and reflecting feelings), (b) discussing and visually demonstrating the level of each verbal and nonverbal cue, (c) coding video-recorded interactions, and (d) discussing and adjusting differences in coding. Other research has successfully rated conversational behavior using outsiders with little to no training (Priem *et al.*, 2009), a decision that has much to do with the complexity of the coding–rating system and the nature of the interaction being observed.

If the coding–rating system or interaction context is highly complex, it is quite important to provide judges with "good examples of what constitutes behavior at the high and low ends of each scale" (Guerrero, 2005, p. 229). For example, in the active listening study, a high level of eye contact was conceptualized as exhibiting eye contact 80% or more of the time, whereas a low level of eye contact was conceptualized as exhibiting eye contact only 20% of the time or less. In addition, judges should be provided examples from the same dataset they are asked to code, both examples that have been coded and rated as well as examples that they can work through and discuss any discrepancies with the principle investigator. Depending on the length of the interaction, it might also be advisable to divide judgments in some way. In this study, judges rated behaviors during the first and second halves of the conversations. Other work has asked judges to rate each minute of an interaction (e.g., Worley & Samp, 2015) or smaller time units (which has vast implications for the type of analyses that can be run on data, as discussed further in this chapter). In this case, there were two sets of coders, one assigned to code nonverbal immediacy and the other assigned to code verbal behaviors. Within each set, judges were asked to assess behaviors individually, which required them to watch each video-recorded conversation multiple times to generate all data necessary for the project.

Reliability of Rating and Coding

An extremely important part of any coding–rating project is the extent to which independent judges are generating internally consistent scores. For coding projects, researchers are typically interested in separating longer portions of data (e.g., conversation) into smaller units and then placing each of those units into a distinct category. An example of a coding scheme that involves extensive unitization and categorization is found in the Verbal Response Modes Typology (VRM). Bavelas and her colleagues also use metrics of unitization and categorization in their microanalysis of face-to-face dialogue.

A typical measure for unitization reliability is Guetzkow's *U* (Guetzkow, 1950), computed as:

$$\frac{(\#\,\mathit{Units\ Identified\ by\ Coder}\ 1)-(\#\,\mathit{Units\ Identified\ by\ Coder}\ 2)}{(\#\,\mathit{Units\ Identified\ by\ Coder}\ 1)+(\#\,\mathit{Units\ Identified\ by\ Coder}\ 2)}$$

So, for example, if one coder identified 19 units in a string of behavior and a second coder identified 20:

$$\mathrm{U}=\frac{20-19}{20+19}=\frac{1}{39}=.025.$$

A Guetzkow's *U* of 0.025 is equivalent to 97.5% agreement (1 – 0.025). Values of *U* below .05 (meaning 95% agreement between two coders on the number of units) are likely acceptable for most purposes. When a research team employs more than two coders, it is important to establish agreement statistics for all pairs of coders and report each metric along with the average.

After coders have determined the number of units in a stream of behavior, the next step is to place each unit into a category. Sometimes, researchers derive categorization schemes inductively, from the specific set of data with which they are working (e.g., microanalysis of face-to-face dialogue). Others use preexisting categories, an existing rubric, or a modified version of a scale (e.g., the active listening in supportive conversations study). Regardless of whether the coding proceeds inductively or deductively, it is important to establish agreement among coders with respect to categorization; that is, are all coders assigning units to the same category of action?

A common measure of categorization reliability is Cohen's kappa (κ) (Cohen, 1960), which estimates agreement between two coders who classify the same items into a set of categories. Kappa is calculated as:

$$\frac{p_o - p_e}{1 - p_e}$$

where p_o is the proportion of observations in agreement, and p_e is the hypothetical probability of chance agreement. When raters are in complete agreement, $\kappa = 1$. When there is no agreement among the raters, other than what would be expected by chance, κ approximates zero. The calculation of kappa is assisted, particularly for large projects, with freely available macros designed for Excel.[8] Kappa values above .70 are generally acceptable (adjusting for chance agreement, two coders agree on 70% of their categorizations), although slightly lower values may be sufficient for overly complex categorization matrices. Like Guetzkow's *U*, kappa is agreement between two coders; when more than two coders are employed, either kappa has to be calculated for all pairs of coders or a statistic such as Krippendorff's alpha (Hayes & Krippendorff, 2007; Krippendorff, 2007) can be used. Krippendorff's alpha also is useful for rating projects, when judges rate behavior as a gestalt.

For the most part, raters/coders judge only a portion of a total dataset; that is, after training and assessment of reliability, coders/raters are asked to evaluate a subset of the data independently. Thus, there is no way to know if coders/raters continue to exhibit high reliability or whether there is some element of "coder drift." Even though coders/raters may have begun the process with good reliability, as they work through a dataset they may develop idiosyncratic coding methods. For coding/rating projects that involve only a few video recordings, the easy solution is to have coders/raters make observations as close as possible to training (e.g., within a few days). For projects that involve more intensive coding/rating, one solution involves requiring all raters to rate all interactions and to model these ratings using structural equation modeling (individual raters are like items on a multi-item scale; see Kotowski, Levine, Baker, & Bolt, 2009). The 3 L research team chose a third option, meeting periodically during each coding–rating project in order to regroup and retrain, ensuring all judges were not only on the same proverbial page but also reading the same proverbial paragraph (and preferably the same proverbial sentence). A similar method is used when employing MFD. When taking multiple measures of intercoder reliability, you should report both the initial and any subsequent measures of intercoder and interrater reliability.

Analyzing Behavioral Data

To this point, we have discussed decisions related to choosing a location for observation, whether to record or code/rate from live interaction, and to what extent we rely on trained judges (and how many and from which perspectives). We also covered specifics related to training and establishing interrater agreement as well as other logistical

8 One example is found here: http://www.real-statistics.com/free-download/real-statistics-resource-pack/.

matters. Clearly, we could not fully represent all issues, but to this point you should have a good, general idea about how to design a behavioral listening study. In this section, we cover what happens after data are collected and how to analyze them to make claims about listening. Although space limitations preclude an extended exploration of all data analytic concerns, it is necessary to introduce some of the more common issues that behavioral researchers are likely to face.

Types of Questions for Behavioral Listening Research

A useful organizational framework for unpacking different ways of analyzing behavioral data comes from Cappella (1991), who argued researchers can ask questions about:

> (0) the types and structures of behaviors enacted in interpersonal encounters,
> (1) the processes of encoding and decoding such behaviors …
> (2) the magnitude and type of influence, if any, that one person's overt behavior has on the partner… [and]
> (3) the association between *patterns* of message interchange between partners and the partners' experienced state of the relationship. (p. 103)

Cappella (1987) referred to these questions as zero-, first-, second-, and third-order questions. Each of these question levels are explored in this section, and examples of listening research and data analytic techniques likely to be employed are provided.

Zero-order questions involve defining behaviors and their structure. We have already covered much about how to sample behaviors, and Figures 6.2 and 6.3 provide examples of behavioral sets that can be studied. But each of these behaviors happens in time, thus necessitating a decision about the time units the researcher will use when assessing listening behavior. If, for instance, you are interested in whether pausing indicates attention or reflexivity or whether interruptions indicate a lack thereof, you will first have to define what constitutes a pause and an interruption. Part of the definition will deal with time—how long does a speaker have to be silent to constitute "pausing" compared to "taking a breath" or some other behavior? How much overlapping speech has to be present for an utterance to count as an interruption? Other issues with respect to timing involve the sequence of events. For some, sampling at the level of conversational turn is fine-grained enough, but for others the turn is not specific enough (e.g., an independent clause or thought unit; compare the Verbal Response Mode, Profile 63, and Memory for Conversation, Profile 38, respectively).

When answering zero-order questions, researchers will most likely report descriptive statistics, namely, measures of central tendency (e.g., mean, median, and mode) and variability (e.g., standard deviation and range). If counting the occurrence of behaviors, frequency distributions are valuable. Although best practice dictates the reporting of descriptive data, much of the listening research falls short of adequately describing sample data (Keaton & Bodie, 2013). Other strategies include developing typologies and models of listening behavior, both those that are unique to particular research projects and those that are able to generalize beyond to inform broader theories. Finally, choices at this order of analysis have implications for subsequent levels. For instance, in order to detect nuanced patterns of behavior (second-order), researchers will have to divide streams of behavior into more granular units. In order to discover quadratic trends, at least three data points are needed; however, more complex patterns of data (e.g., lag-sequential analysis) require 50 or more observations.

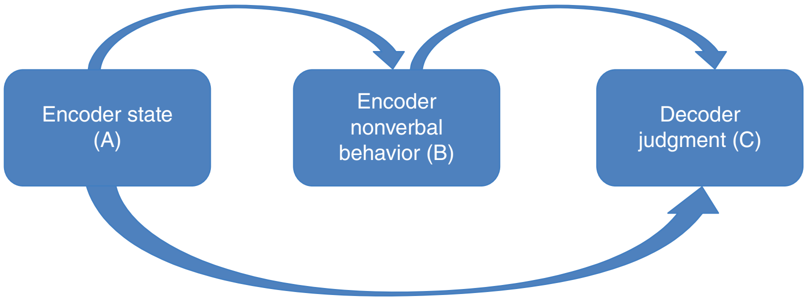


Figure 6.5 A simple model of judgment studies (see Rosenthal, 2005, p. 200). Used with permission.

First-order questions are those involving the production and processing of behavior. For these questions, it is useful to introduce Rosenthal's (1987) model of judgment studies (see Figure 6.5).

A judgment study involves "one or more encoders characterized by one or more attributes (A) observed by one or more decoders who make one or more judgments (C) about the encoders on the basis of selectively presented behavior (B)" (Rosenthal, 1987, p. 4). The typical design of judgment studies involves the observation or manipulation of one or more listening behaviors (B) produced by one or more actual or imagined help providers (A) in the service of assessing the evaluation of that behavior (C). First-order questions are about the A–B (encoder attributes–observed behavior) link: What individual differences might predispose some listeners to enact certain behaviors, or what situational facets might influence some listeners to react in certain ways? Several affective orientations toward listening are discussed in Chapter 5. Other work has shown that distracting listeners from paying full attention to a narrative causes speakers to be less fluent (Bavelas *et al.*, 2000) and to score more poorly on measures of long-term memory for events (Pasupathi, Stallworth, & Murdoch, 1998).

When determining the individual and situational variables to include in a judgment study, theory should guide the way. Theories such as affection exchange theory and constructivism propose specific classes of variables that ought to affect how people behave as listeners (see Bodie, 2012; and Chapter 1). These theories can help focus attempts to map patterns of individual and situational variability. During data analysis, researchers are likely to employ one or more methods from the general linear model (e.g., bivariate correlation or multiple regression) appropriate for answering questions generated from largely cross-sectional and experimental research designs. Exploring group differences and patterns of variability due to explicit manipulation of the environment can assist theory-building efforts and efforts geared toward how to best train people to behave as "good" listeners.

Second-order questions involve exploring patterns of interaction and necessitate interacting partners engaged in some form of conversation or activity. To answer second-order questions, it is not enough to represent the behavior of only a subset of interlocutors (e.g., only one person in a dyad). In dyadic interaction, it is possible to code the behavior of both individuals separately, then aggregate those data to capture mimicry, matching, reciprocity, and compensation, or to rate dyads for these features. Alternatively (or in addition to separate coding), researchers may choose to rate dyadic-level variables (e.g., the degree of mutual eye gaze).

With dyadic data, standard linear modeling techniques (e.g., multiple regression) are not sufficient. At the very least, you have to account for the contingent nature of the data. Given that the behavior of Person A is influenced by and influences Person B's behavior, dyadic data are not likely to meet the independence of error terms assumption. Various ways of handling nested data are outlined in several sources (e.g., Kenny, Kashy, & Cook, 2006).

The final set of questions, *third-order questions*, "takes the next logical step in the study of [listening], asking about the association between relationship states (and relationship outcomes) and [listening] patterns, whether conceived as adaptive patterns or message patterns" (Cappella, 1987, p. 221). Examples of third-order questions about listening include whether and to what extent certain patterns of adaptation, matching, or empathic accuracy influence relationship satisfaction, closeness, intimacy, or coping with problematic events. When behaviors are nested within individuals who are nested within dyads who are nested within larger structures (e.g., patients–physicians within certain healthcare settings), some type of multilevel modeling technique is needed to account for these different levels of analysis (see Dagne, Brown, & Howe, 2007).

New Horizons for Behavioral Listening Research

By recognizing listening as an active and not a passive activity, listening scholars have made great strides in documenting the powerful influence that competent listening can have in our daily lives. Behavioral listening research can take many forms, ranging from the observation of naturally occurring behavior to laboratory work that imposes some sort of structure on listening behavior in order to more closely examine its contours. This type of work is an important component of the larger landscape of listening scholarship, providing an empirical base for theory building and sound advice for educators and other practitioners who wish to improve how people attend to others.

Unfortunately, however, behavioral listening research is underrepresented in the literature. Part of the reason for this is the vast amounts of time and resources needed to adequately capture these data. When a researcher decides to embark on a study that explores listening behavior, she opens a proverbial can of worms including decisions regarding what behaviors to capture, how to represent them, and whether to code or rate them (and from whose perspective). Even when time and resources are available to collect these data, the complexity of data analytic methods appropriate for unpacking patterns and structures causes an additional layer of stress. Whether you want to answer zero-, first-, second-, or third-order questions will suggest the specific techniques you should apply and the degree to which you will need additional training in statistical methods or mathematical modeling.

As illustrated with the Human Speechome Project, however, it is possible to collect and analyze complex streams of data in efficient and time-sensitive ways. The work of Drs. Roy and Patel has taken advantage of recent advances in automated transcription technology as well as automated unitization of speech segments that can take several hours for a small speech sample if done by hand.[9] Other technologies allow researchers to map facial contours (Girard, Cohn, Jeni, Sayette, & De la Torre, 2015) as well as body position with over two dozen points of precision (Han, Shao, Xu, & Shotton, 2013), and then

9 For a list of recent publications from this project, see http://dkroy.media.mit.edu.

match these movements with speech and other data points of interest (e.g., physiological reactivity). Still other advances include devices that can sample human speech in its natural environment (Choudhury & Pentland, 2015; Mehl, Pennebaker, Crow, Dabbs, & Price, 2001) or those that allow for textual analysis of large datasets (Pennebaker & Francis, 1999), each of which is an exciting development for researchers who want to move from the laboratory to the places where people are enacting roles as listeners in their daily lives. Technology aside, the study of listening behavior is vital to understanding how people enact relationships, social roles, and identities; how they cope with problems; how they handle conflict; and how they obtain proper and satisfying medical care, just to name a few. Hopefully, this chapter has provided some insight into how one might appropriately design a behavioral listening study and some motivation to do just that.

References

Afifi, T. D., Afifi, W. A., Merrill, A. F., Denes, A., & Davis, S. (2013). "You need to stop talking about this!" Verbal rumination and the costs of social support. *Human Communication Research, 39*, 395–421. doi:10.1111/hcre.12012

Andersen, J. F., Andersen, P. A., & Jensen, A. D. (1979). The measurement of nonverbal immediacy. *Journal of Applied Communication Research, 7*, 153–180. doi:10.1080/00909887909365204

Barker, L. L. (1971). *Listening behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology, 79*, 941–952. doi:10.1037/0022-3514.79.6.941

Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. *Journal of Communication, 52*, 566–580. doi:10.1111/j.1460-2466.2002.tb02562.x

Bavelas, J. B., & Gerwing, J. (2011). The listener as addressee in face-to-face dialogue. *International Journal of Listening, 25*, 178–198. doi:10.1080/10904018.2010.508675

Beard, D., & Bodie, G. D. (2014). Listening research in the communication discipline. In P. J. Gehrke & W. M. Keith (Eds.), *The unfinished conversation: 100 years of communication studies* (pp. 207–233). New York: Routledge.

Blumer, H. (1969). Sociological implications of the thought of George Herbert Mead. In G. P. Stone & H. A. Farberman (Eds.), *Social psychology through symbolic interaction* (pp. 282–293). Waltham, MA: Xerox College Publishing.

Bodie, G. D. (2012). Listening as positive communication. In T. Socha & M. Pitts (Eds.), *The positive side of interpersonal communication* (pp. 109–125). New York: Peter Lang.

Bodie, G. D., Cannava, K., Vickery, A. J., & Jones, S. M. (2016). Patterns of nonverbal adaptation in supportive interactions. *Communication Studies, 67*, 3–19. doi:10.1080/10510974.2015.1036168

Bodie, G. D., Jones, S. M., Vickery, A. J., Hatcher, L., & Cannava, K. (2014). Examining the construct validity of enacted support: A multitrait-multimethod analysis of three perspectives for judging immediacy and listening behaviors. *Communication Monographs, 81*, 495–523. doi:10.1080/03637751.2014.957223

Bodie, G. D., Pence, M. E., Rold, M., Chapman, M. D., Lejune, J., & Anzalone, L. (2015). Listening competency in initial interactions, II: Applying trait centrality to discover the relative placement of listening competence among implicit competency theories. *Communication Studies, 66*, 528–548. doi:10.1080/10510974.2015.1039657

Bodie, G. D., St. Cyr, K., Pence, M., Rold, M., & Honeycutt, J. M. (2012). Listening competence in initial interactions I: Distinguishing between what listening is and what listeners do. *International Journal of Listening, 26*, 1–28. doi:10.1080/10904018.2012.639645

Bodie, G. D., Vickery, A. J., Cannava, K., & Jones, S. M. (2015). The role of "active listening" in informal helping conversations: Impact on perceptions of listener helpfulness, sensitivity, and supportiveness and discloser emotional improvement. *Western Journal of Communication, 79*, 151–173. doi:10.1080/10570314.2014.943429

Brownell, J. (2013). *Listening: Attitudes, principles, and skills* (5th ed.). Upper Saddle Ridge, NJ: Pearson.

Burgoon, J. K., Stern, L. A., & Dillman, L. (1995). *Interpersonal adaptation: Dyadic interaction patterns*. Cambridge: Cambridge University Press.

Burleson, B. R., & Goldsmith, D. J. (1998). How the comforting process works: Alleviating emotional distress through conversationally induced reappraisals. In P. A. Andersen & L. K. Guerrero (Eds.), *Handbook of communication and emotion: Research, theory, applications, and contexts* (pp. 245–280). San Diego, CA: Academic Press.

Cappella, J. N. (1987). Interpersonal communication: Definitions and fundamental questions. In C. R. Berger & S. H. Chafee (Eds.), *Handbook of communication science* (pp. 184–238). Newbury Park, CA: Sage.

Cappella, J. N. (1991). Mutual adaptation and relativity of measurement. In B. M. Montgomery & S. Duck (Eds.), *Studying interpersonal interaction* (pp. 103–117). New York: Guilford.

Caughlin, J. P. (2002). The demand/withdraw pattern of communication as a predictor of marital satisfaction over time: Unresolved issues and future directions. *Human Communication Research, 28*, 49–85. doi:10.1111/j.1468-2958.2002.tb00798.x

Choudhury, T., & Pentland, A. (2015). *The sociometer: A wearable device for understanding human networks*. Cambridge, MA: Human Design Group.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.

Collins, N. L., & Feeney, B. C. (2004). Working models of attachment shape perceptions of social support: Evidence from experimental and observational studies. *Journal of Personality and Social Psychology, 87*, 363–383.

Dagne, G. A., Brown, C. H., & Howe, G. W. (2007). Hierarchical modeling of sequential behavioral data: Examining complex associatiion patterns in mediation models. *Psychological Methods, 12*, 298–316. doi:10.1037/1082-989X.12.3.298

Derlega, V. J., Barbee, A. P., & Winstead, B. A. (1994). Friendship, gender, and social support: Laboratory studies of supportive interactions. In B. R. Burleson, T. L. Albrecht, & I. G. Sarason (Eds.), *Communication of social support: Messages, interactions, relationships, and community* (pp. 136–151). Thousand Oaks, CA: Sage.

Girard, J. M., Cohn, J. F., Jeni, L. A., Sayette, M. A., & De la Torre, F. (2015). Spontaneous facial expression in unscripted social interactions can be measured automatically. *Behavior Research Methods, 47*, 1136–1147. doi:10.3758/s13428-014-0536-1

Goldsmith, D. J. (2004). *Communicating social support*. New York: Cambridge University Press.

Goldsmith, D. J., & Baxter, L. A. (1996). Constituting relationships in talk: A taxonomy of speech events in social and personal relationships. *Human Communication Research, 23*, 87–114. doi:10.1111/j.1468-2958.1996.tb00388.x

Guerrero, L. K. (2005). Observer ratings of nonverbal involvement and immediacy. In V. Manusov (Ed.), *The sourcebook of nonverbal measures: Going beyond words* (pp. 221–236). Mahwah, NJ: Erlbaum.

Guetzkow, H. (1950). Unitizing and categorizing problems in coding qualitative data. *Journal of Clinical Psychology, 6*, 47–58. doi:10.1002/1097-4679(195001)6:1<47::AID-JCLP2270060111>3.0.CO;2-I

Han, J., Shao, L., Xu, D., & Shotton, J. (2013). Enhanced computer vision with Microsoft Kinect Sensor: A review. *IEEE Transactions on Cybernetics, 43*, 1318–1334.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*, 77–89.

High, A. C., & Solomon, D. H. (2014). Communication channel, sex, and the immediate and longitudinal outcomes of verbal person-centered support. *Communication Monographs, 81*, 439–468. doi:10.1080/03637751.2014.933245

Honeycutt, J. M., & Cantrill, J. G. (2001). *Cognition, communication, and romantic relationships*. Mahwah, NJ: Erlbaum.

Hupfeld, H. (1931). As time goes by. In *Everybody's welcome* [motion picture]. Los Angeles: Warner Bros. Music.

Imhof, M. (2008). What have you listened to in school today? *International Journal of Listening, 22*, 1–22. doi:10.1080/10904010701802121

Jefferson, G. (1978). Sequential aspects of storytelling in conversation. In J. Schenkein (Ed.), *Studies in the organization of conversational interaction* (pp. 219–248). New York: Academic Press.

Jefferson, G. (1980). On "trouble-premonitory" response to inquiry. *Sociological Inquiry, 50*, 153–185. doi:10.1111/j.1475-682X.1980.tb00019.x

Jones, S. M., & Guerrero, L. K. (2001). Nonverbal immediacy and verbal person-centeredness in the emotional support process. *Human Communication Research, 4*, 567–596. doi:10.1111/j.1468-2958.2001.tb00793.x

Jones, S. M., & Wirtz, J. (2006). How does the comforting process work? An empirical test of an appraisal-based model of comforting. *Human Communication Research, 32*, 217–243. doi:10.1111/j.1468-2958.2006.00274.x

Keaton, S. A., & Bodie, G. D. (2013). The statistical and methodological acuity of scholarship appearing in the *International Journal of Listening* (1987–2011). *International Journal of Listening, 23*, 115–135. doi:10.1080/10904018.2013.813206

Keaton, S. A., Bodie, G. D., & Keteyian, R. V. (2015). Relational listening goals influence how people report preferring to talk about problems. *Communication Quarterly, 63*, 480–494. doi:10.1080/01463373.2015.1062407

Kenny, D. A., & Acitelli, L. K. (2001). Accuracy and bias in the perception of the partner in a close relationship. *Journal of Personality and Social Psychology, 80*, 439–448. doi:10.1037/0022-3514.80.3.439

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: Guilford.

Keppel, G., & Wickens, T. D. (2009). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Fort Worth, TX: Harcourt.

Kotowski, M. R., Levine, T. R., Baker, C. R., & Bolt, J. M. (2009). A multitrait-multimethod assessment of the Verbal Aggressiveness and Argumentativeness Scales. *Communication Monographs, 76*, 443–462. doi:10.1080/03637750903300247

Krippendorff, K. (2007). *Computing Krippendorff's alpha-reliability*. Philadelphia: University of Pennsylvania. Retrieved from http://www.asc.upenn.edu/usr/krippendorff/dogs.html

Landsberger, H. A. (1958). *Hawthorne revisited: Management and the worker, its critics, and developments in human relations in industry*. Ithaca, NY: Cornell University Press.
Levinger, G. (1977). Re-viewing the close relationship. In G. Levinger & H. L. Raush (Eds.), *Close relationships: Perspectives on the meaning of intimacy* (pp. 137–161). Amherst, MA: University of Massachusetts Press.
Littlejohn, S. W. (1977). Symbolic interactionism as an approach to the study of human communication. *Quarterly Journal of Speech, 63*, 84–91. doi:10.1080/00335637709383369
Lundsteen, S. W. (1971). *Listening, its impact on all levels on reading and the other language arts*. Urbana, IL: ERIC.
McLaughlin, M., Cody, M. J., Kane, M., & Robey, C. (1981). Sex differences in story receipt and story sequencing behaviors in dyadic conversations. *Human Communication Research, 7*, 99–116. doi:10.1111/j.1468-2958.1981.tb00563.x
Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers, 33*, 517–523.
Mehl, M. R., Vazire, S., Holleran, S. E., & Clark, C. S. (2010). Eavesdropping on happiness: Well-being is related to having less small talk and more substantive conversations. *Psychological Science, 21*, 539–541. doi:10.1177/0956797610362675
Pasupathi, M., & Billitteri, J. (2015). Being and becoming through being heard: Listener effects on stories and selves. *International Journal of Listening, 29*, 67–84. doi:10.1080/10904018.2015.1029363
Pasupathi, M., Stallworth, L. M., & Murdoch, K. (1998). How what we tell becomes what we know: Listener effects on speakers' long-term memory for events. *Discourse Processes, 26*, 1–25. doi:10.1080/01638539809545035
Pearce, W. B. (2009). Communication and social construction: Claiming our birthright. In W. Leeds-Hurwitz & G. Galanes (Eds.), *Socially constructing communication* (pp. 33–56). Cresskill, NJ: Hampton.
Pennebaker, J. W., & Francis, M. E. (1999). *Linguistic Inquiry and Word Count (LIWC): A computer-based text analysis program*. Mahwah, NJ: Erlbaum.
Priem, J. S., Solomon, D. H., & Steuber, K. R. (2009). Accuracy and bias in perceptions of emotionally supportive communication in marriage. *Personal Relationships, 16*, 531–552. doi:10.1111/j.1475-6811.2009.01238.x
Ridge, A. (1993). A perspective on listening skills. In A. D. Wolvin & C. G. Coakley (Eds.), *Perspectives on listening* (pp. 1–14). Norwood, NJ: Ablex.
Rosenthal, R. (1987). *Judgment studies: Design, analysis, and meta-analysis*. Cambridge: Cambridge University Press.
Rosenthal, R. (2005). Conducting judgment studies: Some methodological issues. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research* (pp. 199–234). Oxford: Oxford University Press.
Roy, D., Patel, R., DeCamp, P., Kubat, R., Fleischman, M., Roy, B., … Gorniak, P. (2006, July). *The Human Speechome Project*. Paper presented at the Cognitive Science Society, Vancouver, BC, Canada.
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
Sillars, A. L. (1991). Behavioral observations. In B. M. Montgomery & S. Duck (Eds.), *Studying interpersonal interaction* (pp. 197–218). New York: Guilford.

Smith, E. R. (2000). Research design. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 17–39). Cambridge: Cambridge University Press.

Spitzberg, B., & Cupach, W. R. (2002). Interpersonal skills. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 564–612). Thousand Oaks, CA: Sage.

Surra, C. A., & Ridley, C. A. (1991). Multiple perspectives on interaction: Participants, peers, and observers. In B. M. Montgomery & S. Duck (Eds.), *Studying interpersonal interaction* (pp. 35–55). New York: Guilford.

VanPatten, B. (2014). Language acquisition theories. In C. Facke (Ed.), *Manual of language acquisition* (pp. 103–120). Berlin: Walter de Gruyter.

Weaver, C. (1972). *Human listening: Process and behavior*. Indianapolis, IN: Bobbs-Merrill.

Wolvin, A. D. (1989). Models of the listening process. In C. W. Roberts & K. W. Watson (Eds.), *Intrapersonal communication processes: Original essays* (pp. 508–527). New Orleans, LA: SPECTRA.

Wolvin, A. D., & Coakley, C. G. (1994). Listening competency. *Journal of the International Listening Association, 8*, 148–160. doi:10.1080/10904018.1994.10499135

Worley, T. R., & Samp, J. (2015). Goal variability and perceived resolvability in serial argumentation. *Communication Research*. Advance online publication. doi:10.1177/0093650214565918

Section Three

Measurement Profiles

> *Tests are always made for a purpose, and the purpose has considerable influence on the construct we want to measure.*
>
> Buck (2001, p. 95)

> *Measurement tools are certainly not neutral devices through which information is acquired. Rather, they are researchers' theory-driven constructions of the social world.*
>
> Cappella (1991, p. 106)

This final section profiles both common and emerging listening and listening-related instruments. As you review each profile, keep in mind the above quotes: Measures are designed with particular purposes, likely from one or more theoretical frameworks.

Some measures, especially those used in early research, have not received the scrutiny they deserve; others have been more fully tested but are lesser known to listening investigators. For an instrument to be profiled, it had to have sufficient research publications to be effectively evaluated, or it had to, in our opinion, be a potentially useful tool for future work. In the 65 profiles presented in the following pages, some directly address listening processes (e.g., the Listening Styles Profile), and others are more aptly described as measurements of listening-related characteristics (e.g., Language Style Matching). Most profiles address one specific measure, whereas others examine promising techniques (e.g., functional magnetic resonance imaging [fMRI]) or present multiple scales assessing the same construct (e.g., Nonverbal Immediacy).

Each profile follows a standard format that addresses conceptual and operational definitions, and provides a description of the measure, standard procedures for administration, and assessments of validity and reliability. Whenever possible, scale items are presented in their entirety. When not possible, sample items are presented, and availability information is provided. From our perspective, one of the most important elements of a profile is the critique. For too long, listening scholars have not always followed best practices for scale development and testing.

Keep in mind that space limitations meant that authors often had to make difficult decisions and hard choices when considering the depth of their review, which sample studies to include, what information to exclude, and what elements deserved criticism. We appreciate their hard work and dedication to this project and believe you will too.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

References

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

Cappella, J. N. (1991). Mutual adaptation and relativity of measurement. In B. M. Montgomery & S. Duck (Eds.), *Studying interpersonal interaction* (pp. 103–117). New York: Guilford.

Profile 1

The Academic Listening Self-rating Questionnaire (ALSA)

(Aryadoust, Goh, & Lee, 2012)

Profiled by: Vahid Aryadoust, PhD and Christine C. M. Goh, PhD

National Institute of Education, Nanyang Technological University, Singapore

Construct

The Academic Listening Self-rating Questionnaire (ALSA) is an instrument for the self-appraisal of academic English listening skills.[1]

Instrument Type

Self-Report

Description

The ALSA is a self-report instrument that academic English language learners can use to evaluate their own listening skills and to identify areas of weakness. It is founded on a model of academic listening defined and operationalized by Aryadoust, Goh, and Lee (2012). The model takes into consideration the structure of academic discourse where speaker-related, listener-related, text-related, and situation-related variables play parts in listeners' comprehension processes. Based on these components, Aryadoust *et al.* (2012) posited: (a) a general listening component comprising linguistic components and prosody (LCP) along with cognitive processing skills (CPSs); and (b) an academic listening component comprising memory and concentration (MC), note taking (NT), relating input to other materials (RIOM), and lecture structure (LS).

1 ALSA reflects the designation provided in the original article introducing the measure, which utilized Australian subjects (see Aryadoust, Goh, & Lee, 2012).

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

The ALSA consists of 47 items measuring these six major dimensions of academic listening. Each component is measured by multiple items with a four-category scale consisting of *poor* (1), *satisfactory* (2), *good* (3), and *excellent* (4). The LCP component is measured by 11 items, CPSs by 11 items, MC by three items, NT by four items, RIOM by three items, and LS by 15 items (see ALSA at the end of the profile).

LCP items refer to listeners' ability to process words and meaning, to manage prosodic features such as speed, and to understand details and surface information such as names and numbers. CPS items refer to listeners' ability to infer attitudes, emotions, and implicit meaning in speakers' verbal input. MC concerns listeners' memory capacity and ability to concentrate while listening. NT concerns listeners' attempt to take notes in lectures, tutorials, and seminars, and RIOM refers to listeners' ability to link verbal and nonverbal input. Finally, LS is an external component concerning the specific effects of lectures on the respondents' comprehension processes.

Administration

The ALSA is a self-administered or teacher-administered questionnaire that takes between 20 and 30 minutes to complete. It uses nontechnical English to describe listening processes so that users will have no difficulty rating their own listening abilities. To ensure greater reliability, students with low English proficiency should be briefed and assisted by the teacher.

Scoring

Students' numerical responses are added up to calculate raw scores on each component. Teachers can calculate the mean for each area across all the students in the class by adding up students' scores on each item (numerical responses) and dividing the result by the number of students. In addition, calculating the standard deviation allows the teacher and a student to see where he or she stands in relationship to everyone else in the class. Some students may either be genuinely weak in academic listening or have low confidence in their own listening; showing students how they stand with respect to their peers is one strategy to help them improve.

Development

The ALSA is a multicomponent, 47-item instrument designed to help English learners appraise their own academic listening skills. Aryadoust *et al.* (2012) surveyed the available literature investigating factors that influence English learners' listening comprehension in academic settings. These factors included general listening features (e.g., understanding details and making inferences; Buck, 2001) and academic listening features (e.g., subject knowledge and lecture structure; Flowerdew & Miller, 1992). Specifically, the structure and features of academic discourse are known to affect students' comprehension processes. For example, the presence of discourse markers, the level of formality, and lexico-grammatical features of listening texts and lectures play significant roles in students' comprehension (e.g., Flowerdew, 1994).

Based on their review, Aryadoust *et al.* (2012) posited a six-component model (LCP, CPSs, MC, LS, NT, and RIOM) and developed a pool of 62 items. After review, 47 items were retained in the final questionnaire. The 47-item instrument was piloted on a group of 30 English learners in Australia, and the finalized ALSA was then administered by Aryadoust *et al.* to a group of 119 English as a Second Language university students enrolled in six major Malaysian universities. It was later administered by Aryadoust (2013) to a larger group of students in one major university in Australia and the aforementioned universities in Malaysia (total $n = 255$). Finally, 67 university students took a sample of the International English Language Testing System (IELTS) listening test developed by Sawaki and Nissan (2009) and answered ALSA items (Aryadoust, 2013). Several psychometric and statistical analyses were performed to examine the underlying structure of the items and the correlations between factors and actual listening tests (see below). As a result of these validity examinations, MC, LS, and CPSs consist of a larger number of items than MC, NT, and RIOM. Overall, ALSA was found to generate reliable scores that correlated with the objective listening tests. As such, the scale has a fairly strong validity portfolio.

Reliability

Aryadoust *et al.* (2012) performed the Rating Scale Rasch model (RSM) (Andrich, 1978) analysis on each of the components to examine the reliability of ALSA scores. RSM is a family of the Rasch models that estimates reliability for both items and persons (listeners). The Rasch model reliability ranged from .51 (MC) to .92 (CPSs). Aryadoust *et al.* argued that the rather low reliability index for MC is attributed to the number of items. MC is measured by only three items because it is indirectly measured by items tapping other components, and therefore seems unnecessary to extend. Despite the low RSM reliability indices in a few components, Aryadoust (2013) reported high Cronbach's alpha internal constancy in a multinational sample ($n = 255$) ranging between .70 (MC) and .93 (CPSs).

Validity

Aryadoust *et al.* (2012) used the Rating Scale Rasch model (RSM) and structural equation modeling (SEM) to examine the psychometric features of the instrument. They used data collected from these analyses to develop a validity argument that comprises content-related validity evidence and substantive validity evidence. Content-related validity evidence was provided by the results of the RSM analyses as well as the survey of the available literature on academic listening and self-appraisal. To collect evidence supporting substantive validity, they followed Linacre's (2004) guidelines of psychometric validity. These guidelines address psychometric features and ordering of the rating scales; point-measure correlations (PMCs), which express the correlation between listeners' response on each item and their overall measured listening ability; and RSM fit statistics, which provide an indication of erratic patterns and aberrations in the data. In the pilot study with the Australian sample, seven items did not fit RSM well and were therefore scrutinized and reworded. In later phases, the reworded items functioned properly, with acceptable fit indices and PMCs.

In another study, Aryadoust (2013) presented evidence for structural validity of the instrument; using correlation analysis, he examined the relations between the components of ALSA and two objective listening tests (a sample of the IELTS listening test and a listening test developed by Sawaki and Nissan [2009] for the Educational Testing Service [ETS]). Aryadoust found moderate correlations between IELTS and LCP (.32), NT (.30), and RIOM (.29), as well as medium to high correlations between the ETS listening test and CPS (.50), LCP (.52), NT (.42), LS (.50), and RIOM (.40), lending support to the validity argument of ALSA.

Availability

The current version of ALSA is provided at the end of this profile with permission. The instrument is also available from Aryadoust (2013) and is located on Aryadoust's website.[2] The scale is free to use for research purposes.

Sample Studies

As earlier noted, the ALSA was developed by Aryadoust *et al.* (2012), who developed a validity argument for the instrument using some of Messick's (1989) components of unitary validity: *content-related validity* evidence (whether ALSA represents the construct of academic listening), *substantive validity* evidence (whether the operationalized construct or items elicit information concerning students' level of English academic listening), and *structural validity* evidence (whether the postulated components of ALSA are related to each other and can be predicted by students' scores on objective academic listening tests). Aryadoust *et al.* performed an extensive literature review and used RSM analysis along with SEM to collect evidence supporting content-related, substantive, and structural validity of the ALSA. Evidence backing the validity argument of ALSA outweighed potential rebuttals significantly.

In another study, Aryadoust (2013) collected external evidence of validity for the ALSA. As previously discussed, this study correlated ALSA with an IELTS listening test and a test developed by ETS researchers (Sawaki & Nissan, 2009), thereby providing support for the validity argument of ALSA.

ALSA is a newly developed instrument, and further research is ongoing in China and Turkey to examine its psychometric quality in different contexts.

2 https://www.academia.edu/1789106/Developing_an_Academic_Listening_Self-Assessment_Questionnaire_A_Study_of_Academic_Listening.

Critique

Written in an easy-to-understand language, the ALSA provides useful information to teachers, English language learners, and curriculum designers regarding the potential listening ability of students and their accuracy in rating themselves. It is specifically useful where objective listening tests are not available and the teacher or placement staff need to make quick decisions about the listening ability of candidates or students. The ALSA also will help promote self-appraisals in English language programs and transfer some part of the assessment responsibility to students.

Despite satisfying reliability and validity requirements, the ALSA has yet to be subjected to bias analysis and differential item functioning (DIF). These analyses will help determine whether items would maintain their psychometric features across different groups of listeners and whether the information elicited across different samples of listeners is equally reliable for all samples. In addition, future research can consider expanding the NT, RIOM, and MC dimensions.

Finally, what is categorized as *lecture structure* in the original study may be better referred to as *lecture-specific factors* to denote those lecture features that may affect comprehension. This is different from the general listening abilities that are referred to in other items.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*, 561–573. doi:10.1007/BF02293814

Aryadoust, V. (2013). *Building a validity argument for a listening test of academic proficiency*. Newcastle, UK: Cambridge Scholars Publishing.

Aryadoust, V., Goh, C., & Lee, O. K. (2012). Developing an academic listening self-assessment questionnaire: A study of modeling academic listening. *Psychological Test and Assessment Modeling*, *54*, 227–256.

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

Flowerdew, J. (1994). Research of relevance to second language lecture comprehension: An overview. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 7–29). Cambridge: Cambridge University Press.

Flowerdew, J., & Miller, L. (1992). Student perceptions, problems and strategies in second language lecture comprehension. *RELC Journal*, *23*, 60–80. doi:10.1177/003368829202300205

Linacre, J. M. (2004). Optimizing rating scale category effectiveness. In E. Smith & R. Smith (Eds.), *Introduction to Rasch measurement* (pp. 258–278). Maple Grove, MN: JAM Press.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.

Sawaki, Y., & Nissan, S. (2009). *Criterion-related validity of the TOEFL® iBT listening section* (TOEFL iBT™ Report No. iBT-08). Princeton, NJ: ETS.

Scale

The Academic Listening Self-rating Questionnaire (ALSA)
(Aryadoust, Goh, & Lee, 2012)

Source: Aryadoust, Goh, and Lee (2012). Reproduced with permission of Psychological Test and Assessment Modeling.

Dear Participant,

This questionnaire requires you to make a self-assessment of your academic listening comprehension ability. Please answer each question as accurately as you can. Your answers will be kept strictly confidential. Thank you.

Please fill in these personal particulars.

1) Gender: Male Female
2) Age: ___________

The following statements describe possible listening comprehension levels in English. For each statement, please indicate your self-assessment of your current ability level using the scale provided below:

1) = Not at all well
2) = Not well
3) = Well
4) = Very well

When I am listening in English, I can…

<u>Linguistic components and prosody (LCP)</u>

1) understand numbers, commonplace names, and short phrases in Standard English easily.
2) understand simple descriptions given by my professors about familiar persons, places, and objects.
3) understand the language of short oral reports of events and biographical information.
4) understand short and simple technical descriptions.
5) understand the main ideas and facts of lectures.
6) understand important names, dates, and numbers in lectures/seminars/tutorials.
7) understand details of short descriptions of places, people, and events that I know.
8) understand the lecturers who are non-native English speakers better than the native speakers.
9) keep up with and understand lecturers/tutors who speak fast.
10) recognize incorrect grammar and vocabulary when listening to my peers speaking English.
11) understand key vocabulary items when listening to a lecture/tutorial/seminar.

<u>Cognitive processing skills (CPSs)</u>

12) understand the language expressing personal likes and dislikes without reference to a dictionary.
13) understand oral reports about current and past events.

14) generally understand simple descriptions of feelings and wishes.
15) understand radio and TV news programs without major problems.
16) understand meanings that are not directly stated in lectures/seminars/tutorials.
17) understand the language relevant to professional needs without reference to a dictionary.
18) understand the meaning and the purpose of most idioms, cultural references, word play, and irony.
19) tell apart the language of humorous anecdotes and jokes from facts.
20) understand simple descriptions about familiar persons, places, and objects given by other students with a different first language than mine.
21) understand the language expressing spatial relationships and directions.
22) understand simple descriptions given in English about familiar persons, places, and objects by students with the same first language as me.

<u>Memory and concentration (MC)</u>

23) often remember much of the content of the lecture a day later.
24) concentrate on the lecture without being distracted by my own thoughts.
25) concentrate on lectures/tutorials/seminars without being distracted by people, things, and sounds around me in the room.

<u>Note taking (NT)</u>

26) easily take notes of important details of lectures/seminars/tutorials.
27) rephrase the content of the lecture and then take notes on it.
28) summarize the information from lectures/tutorials/seminars.
29) paraphrase the lecture/tutorial/seminar content to take notes of it.

<u>Relating input to other materials (RIOM)</u>

30) easily get clues from the slides to understand lectures/seminars/tutorials better.
31) relate the description of an object to a map.
32) connect the information of the lecture with my textbook and handouts.

<u>Lecture structure (LS)</u>

33) understand lectures/tutorials/seminars better whenever the lecturers signal when they are going to go on to another topic.
34) understand the relationships among the ideas in a lecture.
35) distinguish main points of lectures/tutorials/seminars from details.
36) understand the lecture/tutorial/seminar format—how it starts, continues, and ends.
37) understand facts without being concerned about distinguishing main points from details in a lecture/tutorial/seminar.
38) follow the hypothesis, persuasion, or argument in lectures/tutorials/seminars.
39) understand how different ideas in a lecture relate to each other.
40) distinguish between supporting examples and major points easily.
41) identify the main topic of the lecture.
42) understand lectures/tutorials/seminars better if they are delivered in formal language with fewer jokes and anecdotes.
43) correct my understanding of lectures/tutorials/seminars immediately if my understanding is incorrect.
44) identify the purpose and scope of lectures/tutorials/seminars.

45) distinguish between information that is relevant or irrelevant to the main points in lectures/tutorials/seminars.
46) tell when the lecturer/tutor is about to start a new topic.
47) understand the main ideas and important facts of conversations about academic subjects in lectures/tutorials/seminars.

Note: Labels should be removed and items randomized prior to administration.

Profile 2

Active - Empathic Listening Scale (AELS)

(Drollinger, Comer, & Warrington, 2006; also Bodie, 2011)

Profiled by: Shaughan A. Keaton, PhD

Young Harris College

Construct

The Active-Empathic Listening Scale (AELS) assesses three dimensions of listening: sensing, processing, and responding (Bodie, 2011; Drollinger, Comer, & Warrington, 2006).

Instrument Type

Self-Report; Other-Report; Behavioral Assessment

Description

The Active-Empathic Listening Scale (AELS) is an 11-item, three-factor scale measuring active-empathic listening across three dimensions: sensing ($n = 4$), processing ($n = 3$), and responding ($n = 4$) (Bodie, 2011; Drollinger *et al.*, 2006). *Sensing* describes a listener's ability to understand relational aspects of speech. *Processing* is the cognitive aspect of listening, and involves attending to, comprehending, receiving, and interpreting messages. *Responding* measures the behavioral output of listening, including verbal and nonverbal feedback.

Administration

The AELS can be administered via paper or online. All versions of the scale utilize 7-point scaling (1 = *never or almost never true*, 7 = *always or almost always true*). For the self-report version, participants indicate the extent to which each of 11 statements generally

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

applies to them, applies in a particular listening situation, or applied after experiencing a particular listening situation. The other-report version asks respondents to assess the statements with respect to a known other. The scale can also be modified to measure perceptions of AEL within a particular conversation. To do so, the prompt can read "My conversational partner" in the place of "I," making sure that the items are modified for grammar (e.g., change "ask" to "asked" for item 10). Finally, the scale can be used by trained or untrained raters asked to watch videotaped or live conversations. The survey takes fewer than 5 minutes to complete for any version. The wording of the instructions can be changed to cover a wide range of specific contexts (e.g., "think of a salesperson," "think of your physician," or "think of a family member").

Scoring

The items within subscales are averaged allowing four scores per participant: sensing, processing, responding, and total AEL.

Development

AEL was originally defined as a form of listening employed by salespeople, where customary active listening is merged with empathy to realize a "higher form of listening" (Comer & Drollinger, 1999; Drollinger *et al.*, 2006, p. 161). The scale was designed to assess effective versus ineffective listening from the points of view of customers. It drew from previous scales developed to measure empathy (Davis, 1980, 1983) and active listening (Ramsey & Sohi, 1997). The Drollinger *et al.* (2006) version was called the active-empathetic listening scale.

Items for the original version of the scale were generated from previous practitioner studies as well as from previous listening and empathy measures. Key informants with 10 years of sales experience were interviewed to provide insight into the role of listening in sales. Trained coders were then asked to sort the items into the three categories of sensing, processing, and responding. Items that did not clearly fit into a category were removed. Four studies were conducted to build a validity portfolio for the scale, each time resulting in item removal. In the last study, an exploratory factor analysis revealed the final 11 items.

Bodie (2011) refined and adapted this 11-item scale to a more general social context. The revised scale includes both cognitive and behavioral items; active listening involves not only processing information conveyed by one's conversational partner but also responding to those messages verbally and nonverbally. The scale has since been adapted to measure interlocutor perceptions of AEL after a conversation (Bodie, Jones, Vickery, Hatcher, & Cannava, 2014) as well as to rate AEL from an objective observer's perspective (Bodie & Jones, 2012).

Reliability

As reported in the studies cited within this profile, the reliability of the subscale scores—sensing ($.73 < \alpha < .85$), processing ($.66 < \alpha < .77$), and responding ($.74 < \alpha < .89$)-display modest to good evidence of internal consistency. Bodie, Gearhart, Denham, and Vickery

(2013) offered evidence for test–retest reliability for sensing ($r = .77$), processing ($r = .73$), and responding ($r = .79$), as well as the scale as a whole ($r = .70$), providing evidence that the AELS was invariant over a span of 14 to 45 days.

Validity

Drollinger *et al.* (2006) reported evidence of the original scale's dimensionality through confirmatory factor analysis (CFA), χ^2 (41) = 95.11, $p < .001$, CFI = .95, RMSEA = .19. They also provided evidence of convergent validity: All three subscales were related to a measure of empathy. Specifically, sensing ($r = .28$), processing ($r = .24$), and responding ($r = .17$) were associated with the Perspective Taking factor of the Interpersonal Reactivity Index (Davis, 1980, 1983); and sensing ($r = .18$) and processing ($r = .20$) were correlated with Empathic Concern (see Profile 28 for the IRI). Furthermore, the subscales of the AELS were related to a similar measure of active listening (Ramsey & Sohi, 1997): sensing ($r = .44$), processing ($r = .44$), and responding ($r = .57$).

Bodie (2011) provided evidence of construct validity for the more general version of the scale, $\chi^2(41, N = 416) = 119.10$, $p < .001$, GFI = .95, CFI = .95, RMR = .062, RMSEA = .06, CI 90% = .05, .08. Evidence of convergent validity has also been provided. Bodie (2011) found the AELS associated with a variety of theoretically relevant constructs: perspective taking ($.28 \le r \le .44$), empathic responsiveness ($.15 \le r \le .18$), sympathetic responsiveness ($.18 \le r \le .40$), and Interaction Involvement ($.19 \le r \le .67$; see Profile 25 for Interaction Involvement). He concluded that the constructs overlap but are not isomorphic (see Chapter 5 discussion of construct proliferation). Gearhart and Bodie (2011) provided further validity evidence by comparing the AELS to the Social Skills Inventory (SSI) (Riggio, 1986). In general, individuals with higher sensing, processing, and responding scores reported being more skilled in SSI-Emotional Sensitivity and the verbal dimensions of the SSI.

Although the AELS can be used to measure individual tendencies to enact AEL (or perhaps the motivation to do so; see Chapter 5), Bodie *et al.* (2013) reported that responses to AELS items vary as a function of situational prompts. The AELS was found to be time invariant and situationally stable; therefore, it can be used as either a time-invariant trait measure or a socially fluctuating state measure.

Pence and James (2014) offered evidence of construct validity through CFA, $\chi^2(41, N = 162) = 114.65$, $p < .001$, TLI = .87, CFI = .91, RMSEA = .12, CI 90% = .08, .13. Although the fit statistics were adequate in this report, the error (RMSEA) is somewhat above commonly accepted parameters, perhaps due to lower sample size. Pence and Vickery (2012) also provided evidence of model fit, $\chi^2(41) = 117.19$, $p < .001$, TLI = .94, CFI = .96, RMSEA = .07, CI 90% = .06, .09.

Availability

The generalized version of the scale is presented here (Bodie, 2011) and is free to use for research purposes with appropriate citation. All other reproduction requires written permission.

Sample Studies

Researchers have investigated the relationship of AEL to a wide variety of characteristics and in a number of contexts: interpersonal communication (Bodie, 2011), intrapersonal communication (Vickery, Keaton, & Bodie, 2015), supportive communication (Bodie & Jones, 2012; Bodie *et al.*, 2014), sales and marketing (Comer & Drollinger, 1999; Drollinger *et al.*, 2006), personality and emotional intelligence (Pence & Vickery, 2012), and biological sex differences (Pence & James, 2014).

Comer and Drollinger (1999) first conceptualized a model of AEL in regard to salespeople and their relationships with customers. They argued that effective listening includes empathy, and the combination of listening and empathy assists the personal selling process. Drollinger *et al.* (2006) later developed a scale to reflect this conceptual notion, developing the three factors in the process (sensing, processing, and receiving).

Bodie, as noted above, expanded this scale to include general conversational settings. Bodie, Vickery, and Gearhart (2013) found that supportive people and good listeners are described similarly, and supportive listening is best defined as a set of behaviors. Bodie and Jones (2012) used an other-report version of the AELS and reported that AEL is a crucial part of supportive communication: Helpers who utilized more person-centered and immediate support were rated as better listeners, although the effects were small in magnitude. Pence and Vickery (2012) examined AEL in regard to emotional intelligence (EI) and personality, finding that EI predicted each AELS dimension. Furthermore, there was a small, negative association between psychoticism and the AELS subscales. Vickery *et al.* (2015) reported associations between AEL and the attributes and functions of imagined interactions (IIs), a form of mental imagery where an individual imagines conversations with others (Honeycutt, 2010). Those not prone to use IIs for rehearsal and self-understanding, and those likely to use imagined conversations with others to compensate for lack of actual interaction with others, were less likely to report responding actively to a conversational other. Furthermore, those who do not engage in IIs before conversations or imagine a range of possible conversations to gain comprehension were not prone to report engaging in acts that acknowledge conversational partners.

Critique

As seen above, Bodie's general version of the AELS, although new, exhibits satisfactory evidence of reliability and validity, including temporal validity. However, the subscales are highly correlated, and further evidence is needed to determine whether three factors or just one are necessary to explain relations with other constructs. Because of the high correlations between the three latent factors, many scholars have used a composite score rather than scores for the individual dimensions of sensing, processing, and responding.

In addition, Bodie *et al.* (2014) found that reports of AEL from an individual listener, a conversational partner interacting with that listener, and a rater trained to assess AEL behaviors were not highly correlated—suggesting the perspective from which one views listening behavior influences scores. This finding seems to call into question the method of using self- and other-reported behaviors rather than observation of actual listening when researchers are interested in behaviors. Given that a component of AEL is behavioral—enacted within a particular conversation—and given that participants may under-, over-, or otherwise misreport their own or others' behaviors, the lack of association suggests that individuals might not be able to discern accurately how they or others generally listen. Consequently, researchers should make an effort to include behavioral data alongside the AELS.

References

Bodie, G. D. (2011). The Active-Empathic Listening Scale (AELS): Conceptualization and evidence of validity within the interpersonal domain. *Communication Quarterly*, *59*, 277–295. doi:10.1080/01463373.2011.583495

Bodie, G. D., Gearhart, C. C., Denham, J. P., & Vickery, A. J. (2013). The temporal stability and situational contingency of active-empathic listening. *Western Journal of Communication*, *77*, 113–138. doi:10.1080/10570314.2012.656216

Bodie, G. D., & Jones, S. M. (2012). The nature of supportive listening II: The role of verbal person centeredness and nonverbal immediacy. *Western Journal of Communication*, *76*, 250–269. doi:10.1080/10570314.2011.651255

Bodie, G. D., Jones, S. M., Vickery, A. J., Hatcher, L., & Cannava, K. (2014). Examining the construct validity of enacted support: A multitrait-multimethod analysis of three perspectives for judging immediacy and listening behaviors. *Communication Monographs*, *81*, 495–523. doi:10.1080/03637751.2014.957223

Bodie, G. D., Vickery, A. J., & Gearhart, C. C. (2013). The nature of supportive listening, I: Exploring the relation between supportive listeners and supportive people. *International Journal of Listening*, *27*, 39–49. doi:10.1080/10904018.2013.732408

Comer, L. B., & Drollinger, T. (1999). Active empathetic listening and selling success: A conceptual framework. *Journal of Personal Selling & Sales Management*, *19*, 15–29. doi:10.1080/08853134.1999.10754156

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, (*10*), 1–19.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113. doi:10.1037/0022-3514.44.1.113

Drollinger, T., Comer, L. B., & Warrington, P. T. (2006). Development and validation of the active empathetic listening scale. *Psychology & Marketing*, *23*, 161–180. doi:10.1002/mar.20105

Honeycutt, J. M. (2010). Introduction to imagined interactions. In J. M. Honeycutt (Ed.), *Imagine that: Studies in imagined interaction* (pp. 1–14). Cresskill, NJ: Hampton.

Pence, M. E., & James, T. A. (2014). The role of sex differences in the examination of personality and active empathic listening: An initial exploration. *International Journal of Listening*, *29*, 1–10. doi:10.1080/10904018.2014.965390

Pence, M. E., & Vickery, A. J. (2012). The roles of personality and trait emotional intelligence in the active-empathic listening process: Evidence from correlational and regression analyses. *International Journal of Listening*, *26*, 159–174. doi:10.1080/10904018.2012.712810

Ramsey, R. P., & Sohi, R. S. (1997). Listening to your customers: The impact of perceived salesperson listening behavior on relationship outcomes. *Journal of the Academy of Marketing Science*, *25*, 127–137. doi:10.1007/BF02894348

Riggio, R. E. (1986). Assessment of basic social skills. *Journal of Personality and Social Psychology*, *51*, 649–660.

Vickery, A. J., Keaton, S. A., & Bodie, G. B. (2015). Intrapersonal communication and listening goals: An examination of attributes and functions of imagined interactions and active-empathic listening behaviors. *Southern Communication Journal*, *80*, 20–38. doi:10.1080/1041794X.2014.939295

Scale

The Active-Empathic Listening Scale (Bodie, 2011)

Source: Bodie (2011). Reproduced with permission of Taylor & Francis.

AELS: Self-Report Version

Instructions: Please indicate how frequently you perceive these statements to be true about yourself, using the following scale:

1	2	3	4	5	6	7
Never or Almost never true						*Always or Almost always true*

Sensing

1) I am sensitive to what others are not saying.
2) I am aware of what others imply but do not say.
3) I understand how others feel.
4) I listen for more than just the spoken words.

Processing

5) I assure others that I will remember what they say.
6) I summarize points of agreement and disagreement when appropriate.
7) I keep track of points others make.

Responding

8) I assure others that I am listening by using verbal acknowledgements.
9) I assure others that I am receptive to their ideas.
10) I ask questions that show my understanding of others' positions.
11) I show others that I am listening by my body language (e.g., head nods).

Note: Other versions can be created by changing "I" to some other prompt such as "My friend" or "My conversational partner" and then adjusting the verb tense (e.g., change "I assure" to "My friend assures"). Items are specified to load on the three latent constructs of sensing, processing, and responding and should be randomized prior to administration. Researchers investigating situational listening should supply a context. Scores should be tested for adherence to model parameters prior to further statistical analysis. Labels should be removed and items randomized prior to administration.

Profile 3

Active Listening Attitude Scale (ALAS)

(Mishima, Kubota, & Nagata, 2000)

Profiled by: Brock Bybee, MA and Jonathon Frost, MA

Louisiana State University and Agricultural & Mechanical College

Construct

The Active Listening Attitude Scale (ALAS) attempts to measure three facets of self-reported attitude toward listening in an active fashion: *Listening Attitude*, *Listening Skill*, and *Conversation Opportunity*.

Instrument Type

Self-Report

Description

Developed by Mishima, Kubota, and Nagata (2000), the ALAS is a 31-item, self-report scale that assesses an individual's attitude toward listening in active ways. Three subscales constitute the ALAS. There are 13 questions categorized under *Listening Attitude*, 11 questions categorized under *Listening Skill*, and 7 questions categorized under *Conversation Opportunity*.

Administration

The ALAS is a self-administered questionnaire that can be completed in 5 to 10 minutes. Respondents are instructed to read individual items and mark their level of agreement along four points bounded by: *agree* (3), *rather agree* (2), *rather disagree* (1), and *disagree* (0). Standard instructions ask respondents to consider their regular listening style in the

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

workplace over the past month. Alternative instructions can be introduced if researchers are interested in studying attitudes toward active listening (AL) in other contexts or relationships as well as over other time periods (e.g., post conversation).

Scoring

The 31 items that constitute the ALAS are divided into the three factors: Listening Attitude, Listening Skill, and Conversation Opportunity. Items within each factor are averaged to form three subscales. Higher scores on Listening Attitude actually reflect less of a person-centered approach to listening; users may wish to reverse-code these items to aid in interpretation. Higher scores on the remaining two subscales reflect more self-perceived listening skill and conversation opportunity.

Development

There are a number of different elements that can influence workplace stress. One element of particular concern to the developers of the ALAS is the human relationships that one establishes at work. The ALAS was created to measure the attitudes of AL among general workers. Factors considered for the development of the scale were taken from Rogers's three elements of PCA—*empathic understanding*, *unconditional positive regard*, and *congruence*—as well as the technical aspects of utilizing AL, and the consideration to whether AL is cognitively employed. Originally, the ALAS questionnaire featured 47 items, which were later reduced to 31 through exploratory factor analysis (Mishima *et al.*, 2000). Although the extraction method was not specified, the authors did comment that they investigated the scree plot to determine the final number of factors.

The first factor, Listening Attitude, contains 13 items referring to how prone a person is to become irritated with a coworker. Respondents who answered positively to these questions were described as having tendencies opposing what the authors, drawing from Rogers (1957), called person-centered attitude. The second factor, Listening Skill, is an 11-item subscale measuring how participants view their own competency in listening. The skills measured are those thought to help promote conversation. The third factor, Conversation Opportunity, contains 7 items that measure how prone respondents are to have conversational opportunities with colleagues and associates.

A fourth factor was initially suggested from the original data; however, the item variance explained by this factor was too low to warrant retaining it as part of the final scale. Mishima *et al.* (2000) later developed a more condensed ALAS scale, using only the Listening Attitude and Listening Skill subscales, with 10 items for each subscale. This 20-item scale has been used in recent studies (Kluger & Zaidel, 2013; Kubota, Mishima, & Nagata, 2004).

Reliability

Two forms of reliability have been reported for the ALAS (Mishima *et al.*, 2000). First, internal consistency as measured by Cronbach's alpha has been reported to range between 0.74 and 0.84. Second, test–retest correlations between administrations separated by 2 weeks using data collected from 61 local government employees ranged between 0.79 and 0.83.

Validity

Mishima *et al.* (2000) provided initial evidence of construct validity by comparing (a) scores reported by doctors and psychologists ($N = 39$) working in hospitals and clinics with a specialty in psychosomatic medicine and (b) scores received by the original 536 workers tested. Based on the assumption that the therapists had received training in psychotherapy, they predicted significant differences in the subscale scores between the two groups. Indeed, therapists scored higher on all three subtests. For Listening Attitude, the mean score for therapists was 29.4 ($SD = 4.9$); for workers, $M = 24.4$ ($SD = 5.4$), $r^2 = .19$. For Listening Skill, the mean score for therapists was 25.0 ($SD = 3.3$), and the mean score for workers was 20.3 ($SD = 4.0$), $r^2 = .29$. Finally, for Conversation Opportunity, the mean score for therapists was 13.2 ($SD = 4.1$), and the mean for workers was 11.4 ($SD = 3.5$), $r^2 = .20$. Thus, the ALAS discriminates among groups known to have higher (versus lower) levels of empathy and willingness to suspend judgment while listening.

Availability

All 47 items that were used to construct the original ALAS are provided at the end of this profile and are reproduced with permission. The items are organized by subscale with markers for the 31 items constituting the original scale and the 20 items used in subsequent studies. The scale is free to use for research purposes.

Sample Studies

Mineyama, Tsutsumi, Takao, Nishiuchi, and Kawakami (2007) explored whether the reported listening attitudes and skills of supervisors were related to stressful working conditions and negative reactions of their employees. Forty-one supervisors and their direct employees ($N = 203$) filled out a 20-item version of the ALAS, only focusing on the Listening Attitude and Listening Skill subscales. Scores were then compared with ratings of working conditions and psychological stress in the workplace by using the Job Content Questionnaire (JCQ) and the Brief Job Stress Questionnaire (BJSQ). Employees of supervisors who had higher LA and LS scores showed a more favorable reaction to stress in the workplace than those employees who worked for supervisors with lower LA and LS scores.

Kubota *et al.* (2004) focused on the direct effects of AL training given to midlevel managers at a government facility. A combination of the condensed ALAS scale was used, as well as an Inventive Experimental Learning (IEL) method. IEL is constructed with two main parts: role-play and overall discussion. Participants were separated into small groups and then asked to discuss their own ideas. In role-playing, one person played the speaker while two played the roles of listeners: One was an active listener, and the other was "just listening." The remainder of the participants played the role of observers. In the overall discussion segment, group members were asked to give a synopsis of their findings from the role-playing session. Middle managers ($N = 284$) completed the ALAS and were divided into three separate categories using a lower quartile and an upper quartile of their ALAS score values: low-score group (−24%), medium-score group (25–75%), and high-score group (+76%). Groups then went through the IEL methods and discussed their findings from the exercise. Results showed that a small percentage of managers decreased in the low-score group, and a small percentage of managers in the high-score group increased. Adjustments in the group led the researchers to believe that AL training can improve workplace AL in just one session.

There have been alternate scales created by researchers with AL being a fundamental factor. Del Piccolo, Angela Mazzi, Scardoni, Gobbi, and Zimmermann (2008) developed the Verona Patient-Centered Communication Evaluation (VR-COPE), a nine-item scale assessing the relational aspects of medical consultations between doctor and patient. AL was also used in the development of the Science Technology Engineering and Mathematics-Active Listening Skills Assessment (STEM-ALSA). Wilkins, Bernstein, Harrison, Bekki, and Atkinson (2012) created the STEM-ALSA, which comprised three separate subscales that measure a person's knowledge, their ability to apply, and their self-efficacy in relation to AL.

Critique

Given that ALAS is a self-report instrument, concerns regarding social desirability bias are warranted. In addition, although the scale claims to measure one's attitude toward AL, it is curious that some items reference behaviors (e.g., "I begin to talk before the other person finished talking"). For a measure to tap only attitudes, items should be worded in a way to reference internal states and dispositions, rather than observable behaviors.

No data have been reported that help substantiate the factor structure of the scale. Indeed, after the initial development study, it is more common that items are cherrypicked and used in studies than the entire scale. Using different versions of the scale causes concerns regarding comparability of data. Likewise, little convergent or discriminant validity evidence has been offered for the scale.

References

Del Piccolo, L., Angela Mazzi, M., Scardoni, S., Gobbi, M., & Zimmermann, C. (2008). A theory-based proposal to evaluate patient-centered communication in medical consultations: The Verona Patient-Centered Communication Evaluation scale (VR-COPE). *Health Education*, *108*, 355–372. doi:10.1108/09654280810899984

Kluger, A. N., & Zaidel, K. (2013). Are listeners perceived as leaders? *International Journal of Listening*, *27*, 73–84. doi:10.1080/10904018.2013.754283

Kubota, S., Mishima, N., & Nagata, S. (2004). A study of the effects of active listening on listening attitudes of middle managers. *Journal of Occupational Health*, *46*, 60–67. doi:10.1539/joh.46.60

Mineyama, S., Tsutsumi, A., Takao, S., Nishiuchi, K., & Kawakami, N. (2007). Supervisors' attitudes and skills for active listening with regard to working conditions and psychological stress reactions among subordinate workers. *Journal of Occupational Health*, *49*, 81–87. doi:10.1539/joh.49.81
Mishima, N., Kubota, S., & Nagata, S. (2000). The development of a questionnaire to assess the attitude of active listening. *Journal of Occupational Health*, *42*, 111–118. doi:10.1539/joh.42.111
Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, *21*, 95. doi:10.1037/h0045357
Wilkins, K. G., Bernstein, B. L., Harrison, C. J., Bekki, J. M., & Atkinson, R. K. (2012). Development of the science technology engineering and mathematics—Active listening skills assessment (STEM-ALSA). In *Proceedings – Frontiers in Education Conference FIE* [6462248]. doi:10.1109/fie.2012.6462248

Further Reading

Ikemi, A., Kubota, S., Noda, E., Tomita, S., & Hayashida, Y. (1992). Person-centered approach in occupational mental health: Theory, research and practice. *Sangyo Igaku: Japanese Journal of Industrial Health*, *34*, 18–29. doi:10.1539/joh1959.34.18
Japan Industrial Safety and Health Association (1990). *Textbook for training psychological consultants: Health promotion for mind and body*. Tokyo: Author.
Rogers, C. R., & Farson, R. E. (1955). *Active listening*. Chicago: Industrial Relations Center, University of Chicago.

Scale

Active Listening Attitude Scale (ALAS)

Please answer your level of agreement/disagreement with the items below using the following scale:

3 = Agree
2 = Rather Agree
1 = Rather Disagree
0 = Disagree

For each item, chose one of the four alternatives based on your ordinary style of listening in the workplace during the last one month.

<u>Listening Attitude</u>

1) I tend to persist in my opinion, while talking with others.*
2) I tend to hurry the other person into talking faster.*
3) I tend to talk in a directive and persuasive way, while talking with others.*
4) While listening, I get irritated from not understanding the other person's feelings.*
5) I inadvertently see the other person from a critical viewpoint.*
6) I tend to deny the other person's opinion, when it's different from mine.
7) When I want to say something, I talk about it, even if I interrupt the other person.*

8) I begin arguing with the other person before I know it, while I'm listening to him/her.*
9) I talk offensively, when I'm in a bad mood.*
10) I begin to talk before the other person finishes talking.*
11) While listening, I tend to talk to the other person, sticking to his/her trivial words.*
12) I'm actually talking longer than the other person in spite of my intention to listen to him/her.*
13) I can listen to the other person, even if he/she has a different opinion from mine.*

Listening Skill

1) I listen to the other person, paying attention to his/her unexpressed feelings.*
2) I tend to listen to others seriously.*
3) I listen to the other person, summarizing in my mind what he/she has said.*
4) I sometimes give the other person a brief summary of what he/she has said.
5) I listen to the other person, putting myself in his/her shoes.*
6) When the other person is hesitating, I give him/her a chance by saying "For example, is it like this?"*
7) I listen to the other person calmly, while he/she is speaking.*
8) I'm pleased that I have given some advice to the other person.*
9) I listen to the other person, paying more attention to the changes of his/her feelings than to the contents of his/her talk.*
10) I'm aware of my own feelings, while I'm listening to others.*
11) I listen to others absent-mindedly.* (R)

Conversation Opportunity

1) I'm willing to say something to others usually.
2) I'm the kind of person whom other people feel easy to talk to.
3) I talk with others personally.
4) I'm asked my advice by other people.
5) I express my feelings straightforwardly.
6) I can listen to other persons' worries, but I can't confide mine. (R)
7) I don't talk with someone else unless I have something I have to talk about.

Remaining Items[1]

1) I should listen to others more seriously.
2) I understand a person as the stereotype of such and such.
3) I don't think I have a smooth conversation, when the other person becomes silent during talking.
4) When I can't follow what the other person is talking about, I pretend to understand it.
5) I never turn down another person's request to give him/her some advice.
6) I become emotional in spite of myself, while I'm talking.
7) I can mutually understand anybody.
8) I can take an interest in anybody.
9) I tell the other person whatever things I feel.
10) I can keep listening to the other person, even if I'm not interested in his/her talk.
11) When I began to talk at the same time as the other person did, I let him/her talk.

1 These items did not load on any of the three listed factors in the original study.

12) When the other person is hesitating, I wait for him/her to begin talking.
13) While listening, I'm careful not to interrupt the other person's talk.
14) After the conversation, I regret that I should not have listening to the other person.
15) I don't get tired from listening to others.
16) Listening to others arouses resistance in me.

Note: Items marked with an asterisk (*) have been used as a condensed version of the ALAS. Labels should be removed and items randomized prior to administration. Items marked with (R) should be reverse coded prior to calculating scores. Higher scores on Listening Attitude reflect less of a person-centered approach to listening; users may wish to reverse-code these items to aid in interpretation. Higher scores on the other two subscales reflect more self-perceived listening skill and conversation opportunity.

Profile 4

Active Listening Observation Scale (ALOS)

(Fassaert, van Dulmen, Schellevis, & Bensing, 2007)

Profiled by: Andrea J. Vickery, PhD

University of Richmond

Construct

The Active Listening Observation Scale (ALOS) was designed to allow ratings of how frequently a physician engaged in specific behaviors indicating attention, understanding, and involvement in face-to-face patient consultations.

Instrument Type

Observer Rating

Description

The ALOS is a measure capturing observer evaluations of a target's behavior in an interaction. It is composed of items representing observable verbal behavior (e.g., questions, acknowledgment of feelings and emotions), nonverbal behavior (e.g., nonverbal understanding, inviting body language), and general behavioral perceptions (e.g., relaxed, not distracted). Higher scores on individual items represent more frequent use of, for instance, questions, acknowledgments, and displays of understanding in a relaxed and open interactional style. Individual items are averaged together such that higher average scores reflect the more frequent use of active listening behaviors, and lower scores reflect the infrequent use of active listening behaviors.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

The ALOS is not directly administered to participants; instead, observers rate the frequency of particular active listening behaviors (items) demonstrated by a target (the listener). When used as developed in Fassaert, van Dulmen, Schellevis, and Bensing (2007), raters are first trained on how to recognize and appropriately rate ALOS items. Raters should demonstrate consistency at two levels. First, raters should assess behavioral items consistently (e.g., a "1" is rated the same way in different videos). Second, raters should evaluate targets similarly (i.e., rater scores for a particular video should correlate). Alternatively, untrained observers can be asked to complete items about another individual, either live or from a videotaped interaction, with minimal training.

Scoring

Each of the seven items is rated on a 5-point scale ranging from *never* (1) to *always* (5), representing how frequently (or infrequently) a listener engaged in these behaviors in the interaction. Scores on the seven items are then averaged together to generate an overall score; there are no subscores as the instrument is unidimensional.

Development

The ALOS was first developed by Thijs Fassaert, Sandra van Dulmen, François Schellevis, and Jozien Bensing (2007) with the practical goal of understanding how active listening during patient–physician consultations could improve patient satisfaction and reduce the number of repeat visits for minor ailments. Fassaert *et al.* (2007) recognized that how physicians express understanding and demonstrate attention helps to constitute physician active listening.

The ALOS includes various verbal, nonverbal, and general behaviors purported to represent active listening. The 14 items initially developed by Fassaert *et al.* (2007) include verbal behaviors such as asking open questions, adjusting language, and acknowledging feelings and emotions; nonverbal behaviors such as expressing understanding and using inviting body language; and general behaviors where listeners create an open atmosphere, do not act distracted, and give others time to talk. After removing six items with low intercoder agreement, eight items were submitted to a principal component analysis, after which an additional item was removed.

The original development of the ALOS focused on physician listening in patient–physician interactions, but recent research has extended the behaviors represented in the ALOS to supportive conversations between relational partners featuring nonmedical problems. In their study, Bodie and Jones (2012) found validity evidence for an 11-item version of the scale that was completed by untrained observers of supportive listeners.

Reliability

In the original development of the instrument, Fassaert *et al.* (2007) reported acceptable reliability on the seven items of the ALOS, estimating a Cronbach's alpha (α) of .83. Additionally, Fassaert *et al.* (2007) reported interrater reliability estimates comparing the two observers asked to rate a sample of videos on the ALOS items. Fassaert *et al.* (2007) reported an average kappa of .52 (range = .43 to .62). Fassaert *et al.* also reported inter-rater reliability estimates for six low-performing items (Items 1, 4, 5, 6, 10, and 14) specifying their criteria for excluding low-performing items as $r < .50$ and Cohen's kappa < .34. It should be noted that low reliability in their study is likely not the result of problematic items but of problematic training. Fassaert *et al.* (2007) performed a principal component analysis on the remaining eight items, seven of which loaded sufficiently on a single component that explained 47.5% of the item variance.

The ALOS has been used in studies of supportive listening, with results suggesting appropriate estimates of reliability. Bodie and Jones (2012) asked untrained observers to watch a video of a supportive conversation and rate the active listening behaviors of a target in the video; the authors estimated acceptable scale reliability of $\alpha = .91$ using the full 14-item ALOS. Bodie, Jones, Vickery, Hatcher, and Cannava (2014) used single items modeled after the ALOS to capture active-empathic listening behaviors in supportive conversations. Two of these items closely resemble items on the original ALOS: using exploring questions and expressing understanding verbally. Raters also completed two specific verbal behavior items as well as a "global rating of active listening that reflected their assessment of the four composite behaviors" (Bodie *et al.*, 2014, p. 505). This modification of the ALOS resulted in an appropriate estimate of inter-rater reliability, Krippendorff's $\alpha = .80$. The authors did not provide rater reliability for individual items. Krippendorff's alpha meets the same reliability criteria as Cohen's kappa but can be calculated for a group of raters versus the limitation of two raters in Cohen's kappa (Hayes & Krippendorf, 2007).

Validity

Fassaert *et al.* (2007) provided evidence for the construct validity of the ALOS. With concern for content validity, Fassaert *et al.* (2007) initially developed 14 items to represent various facets of active listening behaviors. Additionally, Fassaert *et al.* (2007) reported that individual items were moderately associated with the other items on the ALOS (range = .37–.71), providing additional support for the scale. The ALOS also is only moderately associated with ratings of physician verbal attention, patient-reported previsit anxiety, and patient-rated affective performance, suggesting divergent validity from such related constructs (e.g., affective performance). In recent research extending the ALOS to supportive conversations, Bodie and Jones (2012) employed the full 14-item measure and achieved acceptable measurement model fit after removing three items (e.g., "is not off hand or hasty," "uses exploring questions," and "spends time on social talk"). These results support the validity of the underlying unidimensional measurement model of the ALOS, even when additional ALOS items are included.

Availability

The ALOS instrument was first published in *Patient Education and Counseling* by Fassaert *et al.* (2007). All 14 original items are presented below with permission and are free to use for research purposes; the 7- and 11-item versions used in prior research are identified.

Sample Studies

Stemming from the original purpose of this instrument, the ALOS has been used to investigate outcomes of physician–patient interactions (Fassaert *et al.*, 2007, 2008; van Dulmen, Fassaert, van der Jagt, & Schellevis, 2010). Also within the medical context, the ALOS has recently been used as an evaluation of peer leaders in diabetes support groups. Tang, Funnell, Gillard, Nwankwo, and Heisler (2011) developed a training program for peer leaders who facilitate diabetes self-management support groups. In this training program, individuals engaged in partnered role-play activities and then assessed their partners' listening behaviors with the ALOS. Before graduation from the program, peer leaders engaged in a patient interview simulation, their behaviors were rated again with the ALOS, and peer leaders had to achieve an average score greater than 4.0 on the 5.0 scale. Peer leaders who did not pass the first time were allowed to retake the assessment (Tang *et al.*, 2011; Tang, Sohal, & Garg, 2013). No reliability or validity estimates were reported in these studies. Similarly, the ALOS was used to assess the development of listening skills in second-year medical students by Hulsman, Harmsen, and Fabriek (2009), but, again, reliability and validity estimates were not reported.

Outside the medical context, the ALOS has been used to operationalize active listening in supportive conversations; these studies support the frequency and occurrence of active listening behaviors in informal supportive conversations and the effects of these behaviors on a host of outcomes (Bodie & Jones, 2012; Bodie *et al.*, 2014).

Critique

The use of the ALOS in observing and rating a participant's active listening is recommended when study design, research questions, or hypotheses aim to investigate how particular active listening behaviors are displayed in conversation. This instrument is relatively new, with only a handful of studies available outside of work conducted by the original authors on the same dataset, but as Bodie and Jones (2012) argued, it "taps microlevel behaviors" important in active listening (p. 259). If researchers use the ALOS, reliability and validity estimates should continue to be reported to contribute to the validity profile of this instrument. Fassaert *et al.* (2007) do report lower interrater reliability on particular items, although training and discussion with raters during training can mitigate this concern.

The ALOS has been used in training and assessing active listening skills (i.e., Tang *et al.*, 2011, 2013). Although the importance of assessing active listening skills in this context is commended, caution should be exercised in employing this instrument as an assessment tool without addressing ecological validity and responsiveness validity concerns, as much of the present evidence for validity is based on measurement validity.

References

Bodie, G. D., & Jones, S. M. (2012). The nature of supportive listening II: The role of verbal person centeredness and nonverbal immediacy. *Western Journal of Communication*, *76*, 250–269. doi:10.1080/10570314.2011.651255

Bodie, G. D., Jones, S. M., Vickery, A. J., Hatcher, L. C., & Cannava, K. E. (2014). Examining the construct validity of enacted support: A multitrait-multimethod analysis of three perspectives for judging immediacy and listening behaviors. *Communication Monographs*, *81*, 495–523. doi:10.1080/03637751.2014.957223

Fassaert, T., van Dulmen, S., Schellevis, F., & Bensing, J. (2007). Active listening in medical consultations: Development of the active listening observation scale (ALOS-global). *Patient Education and Couseling*, *68*, 258–264. doi:10.1016/j.pec.2007.06.011

Fassaert, T., van Dulmen, S., Schellevis, F., van der Jagt, L., & Bensing, J. (2008). Raising positive expectations helps patients with minor ailments: A cross-sectional study. *BMC Family Practice*, *9*, 1471–2296. doi:10.1186/1471-2296-9-38

Hayes, A. F., & Krippendorf, K. (2007). Answering the call for a standard reliability measures for coding data. *Communication Methods and Measures*, *1*, 77–89. doi:10.1080/19312450709336664

Hulsman, R. L., Harmsen, A. B., & Fabriek, M. (2009). Reflective teaching of medical communication skills with DiViDU: Assessing the level of student reflection on recorded consultations with simulated patients. *Patient Education and Couseling*, *74*, 142–149. doi:10.1016/j.pec.2008.10.009

Tang, T. S., Funnell, M. M., Gillard, M., Nwankwo, R., & Heisler, M. (2011). The development of a pilot training program for peer leaders in diabetes: Process and content. *The Diabetes Educator*, *37*, 67–77. doi:10.1177/0145721710387308

Tang, T. S., Sohal, P. S., & Garg, A. K. (2013). Evaluating a diabetes self-management support peer leader training programme for the English- and Punjabi-speaking South-Asian community in Vancouver. *Diabetic Medicine*, *30*, 746–752. doi:10.1111/dme.12179

van Dulmen, S., Fassaert, T., van der Jagt, L., & Schellevis, F. (2010). The association between positive communication strategies and the course of minor ailments. *Huisarts Wet*, *5*, 265–271. doi:10.1007/s12445-010-0117-1

Scale

Active Listening Observation Scale (ALOS-global) (Fassaert *et al.*, 2007)

Source: Fassaert *et al.* (2007). Reproduced with permission of Elsevier.

The GP…[1]

1) Uses inviting body language
2) Shows not to be distracted during the consultation[1]
3) Is not off-hand or hasty[2,3]
4) Is obviously relaxed and confident
5) Is not detached

1 This introduction to the scale was used in the original study by Fassaert *et al.* (2007). Participating physicians were evaluated by other medical experts.
2 These items are part of the reduced 7-item ALOS-global.
3 These items were removed to improve model fit in Bodie and Jones (2012).

6) Adjusts his/her language to that of the patient
7) listens attentively[2]
8) Gives patient time and space to present the problem[1]
9) Uses exploring questions[2,3]
10) Creates an open atmosphere during the conversation
11) Spends time on social talk[3]
12) Is good in leading the conversation[2]
13) Expresses understanding non-verbally[2]
14) Dilates verbally upon patient's feelings or emotions

Profile 5

Affectionate Communication Scale (TAS) (Trait-Given & Trait-Received)

(Floyd, 2002)

Profiled by: Kory Floyd, PhD[1] and Mark Alan Generous, PhD[2]

[1] *University of Arizona*
[2] *Saint Mary's College of California*

Construct

The Trait Affection Scale (TAS) assesses an individual's typical or trait-level tendency toward affectionate behavior. The scale comprises separate subscales to assess the amount of affection an individual typically gives to others (TAS-G) and typically receives from others (TAS-R).

Instrument Type

Self-Report

Description

The TAS was constructed to assess a participant's trait level of affectionate communication. Although most people are more or less affectionate in certain relationships and under certain circumstances (see ACI profile, Profile 6), individuals also evidence a typical level or trait-like tendency, wherein some are more affectionate than others regardless of the context. Participants completing the TAS are presented with 16 statements and are asked to indicate, on a scale from 1 to 7, how much they agree or disagree with each statement as a description of themselves. Ten items comprise the TAS-G, which measures an individual's tendency to give or express affection to others. Six items comprise the TAS-R, which assesses a person's tendency to receive expressions of affection from others.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

Participants indicate their level of agreement with 16 statements about themselves. The measure takes approximately 5 minutes to complete and can be administered with pencil and paper, in an online questionnaire, over the phone, or during a face-to-face interview.

Scoring

TAS-G and TAS-R are usually administered using a 7-point scale (from 1 = *strongly disagree* to 7 = *strongly agree*), although some have used a 5-point scale (e.g., Lewis, Heisel, Reinhart, & Tian, 2011). The TAS-G is made up of 10 items, with 5 positively worded items (e.g., "I consider myself to be a very affectionate person") and 5 negatively worded items (e.g., "I'm not a very affectionate person"). The TAS-R includes six items, with four positively worded items (e.g., "People are always telling me that they like me, love me, or care about me") and two negatively worded items (e.g., "Most of the people I know don't express affection to me very often"). After reverse-coding specified items, scale scores are typically computed by aggregating individual item scores. Scores from the two subscales can also be averaged together to provide one holistic measure of trait affectionate communication (see Floyd, Pauley, & Hesse, 2010; Hesse & Floyd, 2008; Pauley, Hesse, & Mikkelson, 2014).

Development

The TAS-G was initially developed for use as a manipulation check for a broader study in affectionate communication (see ACI profile, Profile 6). Floyd (2002) created known-divergent groups by instructing undergraduate students to distribute questionnaires to the "most affectionate and least affectionate persons they knew" (p. 140). Floyd created the TAS-G to determine whether the two groups differed in their trait-level tendencies to express affection. The comparison demonstrated that 92% of the variance in the TAS-G score was accounted for by group condition (i.e., whether the individuals were in the high- or low-affection condition). TAS-R served as a dependent variable in the analyses of the same study, with 49% of the variance in TAS-R accounted for by group condition. Horan and Booth-Butterfield (2010) modified both scales to assess affection given and received within a specific romantic relationship (e.g., item 2 on the TAS-G was reworded to "I am always telling my romantic partner how much I care about him or her," and item 1 on the TAS-R was reworded to "My partner hugs me a lot").

Reliability

Reliability estimates (Cronbach's alpha) of .79 or higher have been reported for the TAS-G (see Floyd *et al.*, 2005), and .84 or higher for the TAS-R (Floyd, Hesse, & Haynes, 2007). These values indicate that both the TAS-G and TAS-R produce internally consistent scores. Some researchers also have aggregated the scores from the TAS-R and the TAS-G to create a composite trait affection score, and high internal consistencies have been reported for this procedure (alphas have been .93 or higher; see Floyd *et al.*, 2010; Hesse & Floyd, 2008; Pauley *et al.*, 2014).

Validity

Whether using the composite trait affection score or individual subscales (TAS-G and TAS-R), multiple findings support the construct validity of the TAS. Floyd *et al.* (2005), for instance, found that TAS-G scores were positively associated with relationship satisfaction ($r = .36$), social activity ($r = .38$), and the likelihood of being in a romantic relationship ($r = .25$). At the same time, they were negatively associated with indicators of insecure attachment, such as discomfort with closeness ($r = -.67$), fear of intimacy ($r = -.53$), and the perception that relationships are of secondary importance ($r = -.50$). Similarly, Hesse and Trask (2014) found that secure attachment was positively related to both TAS-G ($\beta = .49$) and TAS-R ($\beta = .24$), whereas fearful attachment was negatively related to TAS-G ($\beta = -.51$) and TAS-R ($\beta = -.11$). Dismissive attachment was related to TAS-G ($\beta = -.43$), but not TAS-R. Hesse and Floyd (2008) also reported that trait affection levels (the aggregate of TAS-G and TAS-R) were negatively correlated with alexithymia, an individual characteristic suppressing the ability to encode and decode emotions ($r = -.56$). With respect to personality characteristics, Floyd *et al.* (2005, Study 1) reported that TAS-G scores were positively associated with extraversion ($r = .61$) and negatively associated with psychoticism ($r = -.56$) and neuroticism ($r = -.22$), and that all three correlations remained significant even after controlling for TAS-R scores. Floyd *et al.* (2005, Study 3) also found that TAS-G scores predicted liking ($r = .49$), love ($r = .50$), and relationship satisfaction ($r = .59$) in close relationships. Lewis *et al.* (2011) even found that TAS scores predicted relative electrical activity in the left anterior cortex of the brain versus the right anterior cortex.

Availability

The TAS is available in Floyd (2002) and is reproduced with permission at the end of this chapter. It is free for use with appropriate citations.

Sample Studies

Multiple studies have used the TAS (either as a composite measure or divided into TAS-G and TAS-R subscales) to examine aspects of close relationships and individual health. In a study of married couples, for instance, Pauley *et al.* (2014) found that trait affectionate communication predicted husbands' and wives' enactment of relational maintenance behaviors. For example, spouses' affectionate communication scores predicted their own maintenance behaviors, husbands' affectionate communication predicted their wives' enactment of positivity and network sharing, and husbands' and wives' affectionate communication scores mutually predicted their enactment of assurances. Hesse, Rauscher, Roberts, and Ortega (2014) also reported that trait affectionate communication mediated a relationship between a hurtful family environment and family satisfaction.

Trait affectionate communication also is associated with individual well-being and health. Using a composite trait affection score (i.e., the aggregate of TAS-G and TAS-R), Hesse and Floyd (2008) found that trait affectionate communication mediated the effects of alexithymia on mental health, relationship closeness, and nonverbal immediacy. In a later study, Hesse and Floyd (2011) also found that trait affectionate communication mediated the relationship between alexithymia and individuals' attachment behavior and number of close relationships.

Concerning physical health, Floyd *et al.* (2007) reported that TAS-G scores predicted lower systolic blood pressure ($r = -.61$) diastolic blood pressure ($r = -.54$), and glycosylated hemoglobin (HbA_{1c}), a measure of average blood sugar ($r = -.20$). Similarly, Floyd, Pauley *et al.* (2014) found that trait affectionate communication predicted higher immunoglobulin levels and higher toxicity of natural killer cells. Floyd *et al.* (2010) also reported that trait affectionate communication predicted elevation of oxytocin—a calming and pain-suppressing hormone—in the wake of acute stress. Collectively, these studies demonstrate the health benefits of having a high trait level of affectionate communication, although Floyd, Hesse, Boren, and Veksler (2014) also found that trait affectionate communication predicted higher antibody titers to latent Epstein–Barr virus, which suggests immunosuppression.

Critique

The TAS-G and TAS-R dependably produce adequate estimates of internal consistency and have amassed an impressive validity portfolio as measures of an individual's trait level of affection given and received; however, the factor structure of both scales (either unidimensional or multidimensional) has yet to be tested via confirmatory factor analysis, which would be a useful statistical procedure to further demonstrate validity. Additionally, some items in the scale are double-barreled (e.g., "I love giving people hugs or putting my arms around them" and "People are always telling me that they like me, love me, or care about me"), which has been identified as problematic in scale construction by other interpersonal communication scholars (see Stafford [2010] for a discussion of this issue with the Relational Maintenance Strategies Measure). Finally, no evidence speaks to the test–retest reliability of the TAS, which would be expected to be high if both subscales index a trait.

References

Floyd, K. (2002). Human affection exchange: V. Attributes of the highly affectionate. *Communication Quarterly*, *50*, 135–152. doi:10.1080/01463370209385653

Floyd, K., Hess, J. A., Miczo, L. A., Halone, K. K., Mikkelson, A. C., & Tusing, K. J. (2005). Human affection exchange: VIII. Further evidence of the benefits of expressed affection. *Communication Quarterly*, *53*, 285–303. doi:10.1080/01463370500101071

Floyd, K., Hesse, C., Boren, J. P., & Veksler, A. E. (2014). Affectionate communication can suppress immunity: Trait affection predicts antibodies to latent Epstein-Barr virus. *Southern Communication Journal*, *79*, 2–13. doi:10.1080/1041794X.2013.858178

Floyd, K., Hesse, C., & Haynes, M. T. (2007). Human affection exchange: XV. Metabolic and cardiovascular correlates of trait expressed affection. *Communication Quarterly*, *55*, 79–94. doi:10.1080/01463370600998715

Floyd, K., Pauley, P. M., & Hesse, C. (2010). State and trait affectionate communication buffer adults' stress reactions. *Communication Monographs*, *77*, 618–636. doi:10.1080/03637751.2010.498792

Floyd, K., Pauley, P. M., Hesse, C., Veksler, A. E., Eden, J., & Mikkelson, A. C. (2014). Affectionate communication is associated with markers of immune and cardiovascular system competence. In J. M. Honeycutt, C. Sawyer, & S. Keaton (Eds.), *The influence of communication on physiology and health status* (pp. 115–130). New York, NY: Peter Lang.

Hesse, C., & Floyd, K. (2008). Affectionate experience mediates the effects of alexithymia on mental health and interpersonal relationships. *Journal of Social and Personal Relationships*, *25*, 793–810. doi:10.1177/0265407508096696

Hesse, C., & Floyd, K. (2011). Affection mediates the impact of alexithymia on relationships. *Personality and Individual Differences*, *50*, 451–456. doi:10.1016/j.paid.2010.11.004

Hesse, C., Rauscher, E. A., Roberts, J. B., & Ortega, S. R. (2014). Investigating the role of hurtful family environment in the relationship between affectionate communication and family satisfaction. *Journal of Family Communication*, *14*, 112–128. doi:10.1080/15267431.2013.866453

Hesse, C., & Trask, S. L. (2014). Trait affection and adult attachment styles: Analyzing relationships and group differences. *Communication Research Reports*, *31*, 53–61. doi:10.1080.08824096.2013.844118

Horan, S. M., & Booth-Butterfield, M. (2010). Investing in affection: An investigation of affection exchange theory and relational qualities. *Communication Quarterly*, *58*, 394–413. doi:10.1080/01463373.2010.524876

Lewis, R. J., Heisel, A. D., Reinhart, A. M., & Tian, Y. (2011). Trait affection and asymmetry in the anterior brain. *Communication Research Reports*, *28*, 347–355. doi:10.1080/08824096.2011.615960

Pauley, P. M., Hesse, C., & Mikkelson, A. C. (2014). Trait affection predicts married couples' use of relational maintenance behaviors. *Journal of Family Communication*, *14*, 167–187. doi:10.1080/15267431.2013.864292

Stafford, L. (2010). Measuring relationship maintenance behaviors: Critique and development of the revised relationship maintenance behavior scale. *Journal of Social and Personal Relationships*, *28*, 278–303. doi:10.1177/0265407510378125

Scale

Affection Communication Scale – Trait (TAS) (Floyd, 2002)

Source: Floyd (2002). Reproduced with permission of Taylor & Francis.

How much would you say each of the following statements reflects you and the way you communicate? Indicate your level of agreement by writing the appropriate number on the line preceding each item, according to the scale below.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

<u>Trait Affection – Given</u>

1) I consider myself to be a very affectionate person.
2) I am always telling my loved ones how much I care about them.
3) When I feel affection for someone, I usually express it.
4) I have a hard time telling people that I love them or care about them.*
5) I'm not very good at expressing affection.*
6) I'm not a very affectionate person.*
7) I love giving people hugs or putting my arms around them.
8) I don't tend to express affection to other people very much.*
9) Anyone who knows me well would say that I'm pretty affectionate.
10) Expressing affection to other people makes me uncomfortable.*

<u>Trait Affection – Received</u>

11) People hug me quite a bit.
12) People are always telling me that they like me, love me, or care about me.
13) I don't get very much affection from other people.*
14) I get quite a bit of affection from others.
15) Many people I know are quite affectionate with me.
16) Most of the people I know don't express affection to me very often.*

Note: Labels should be removed and items randomized prior to administration. Items marked with an asterisk (*) should be reverse-scored. Scores for the given and received subscales are calculated by aggregating the scores of the individual items comprising each subscale. A total score for trait affectionate communication can be calculated by aggregating the scores of all 16 items on the TAS.

Profile 6

Affectionate Communication Index (ACI)

(Floyd & Morman, 1998)

Profiled by: Kory Floyd, PhD[1] and Lisa J. van Raalte, PhD[2]

[1] *University of Arizona*
[2] *Sam Houston State University*

Construct

The Affection Communication Index (ACI) was developed to measure verbal expressions of affection, direct nonverbal expressions of affection, and affectionate social support across various relationship types. It indexes an individual's perception of how much affection he or she expresses within a particular close relationship.

Instrument Type

Self-Report

Description

The Affectionate Communication Index (ACI) was constructed to measure how frequently an individual expresses affection in a specific target relationship (Floyd & Morman, 1998). Participants are presented with 18 behaviors and asked to indicate, on a scale from 1 (*never or almost never do this*) to 7 (*always or almost always do this*), how often they engage in each behavior within a target relationship as a means of expressing affection to the other person. These 18 affectionate behaviors are organized into three clusters: verbal statements (five items; e.g., saying "I love you" or "I care about you"), direct nonverbal gestures (eight items; e.g., kissing, hugging, and handholding), and socially supportive behaviors (five items; e.g., helping with problems and sharing private information).

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

Participants are asked to think of a specific relationship when completing the ACI. They then report how often they enact each listed behavior as a means of communicating affection to the target relational partner. The assessment takes approximately 3 minutes to complete and can be used with virtually any type of relationship, including romantic, platonic, and familial bonds.

Scoring

Responses are aggregated to produce a total expressed affectionate communication score. Scholars also can calculate separate scores for each of three subscales indexing verbal expressions of affection, direct nonverbal expressions of affection, and affectionate social support.

Development

When developed by Kory Floyd and Mark Morman (1998), the goal was to formulate a practical and psychometrically sound self-report measure of expressed affectionate communication from an otherwise inconsistent measure of affection in previous research. Adopting a grounded theory approach, 218 undergraduate students provided a list of 67 affectionate behaviors they shared with a close partner. An independent sample of 34 undergraduate students reviewed the list of 67 items. The authors eliminated items that these students indicated lacked face validity and retained items that were agreed upon by at least half of the respondents. The 34 remaining items were subjected to factor analysis (described here in the Reliability section), internal reliability tests, and construct validity checks that resulted in an 18-item scale.

Reliability

To reduce the number of items in the final scale, the 34 items were submitted to a principal components analysis (oblique rotation). Cattell's scree test was used to justify a three-component solution: nonverbal, verbal, and support (see Floyd & Morman, 1998). This structure replicated when independent data were submitted to a confirmatory factor analysis. In both instances, the resulting scale contained 18 items, with 8 items representing nonverbal affection, 5 items representing verbal affection, and 5 items representing support affection. Internal consistency reliability as assessed by Cronbach's alpha has consistently reached conventionally acceptable levels for the overall ACI and for each subscale of the ACI (on average yielding .70 or above). The ACI has also been subject to test–retest reliability over a 14-day period (see Floyd & Morman, 1998). Scores were significantly correlated from Time 1 to Time 2 ($r = .87$ for verbal expressions, .89 for nonverbal expressions, and .83 for socially supportive expressions) and did not change significantly over time.

Validity

The ACI was developed by asking a pool of respondents to generate referents of their own affectionate behavior. Self-reports of affectionate behavior provided real-world examples of affectionate communication, thus strengthening ecological validity in the ACI. In the development of the ACI, items that did not adequately reflect expressions of affectionate communication were removed, thereby boosting face validity. For example, the original list of affectionate communication behavior included getting drunk together or studying together. Whereas these examples may reflect bonding activities, they did not reflect expressions of affection and were, therefore, removed from the scale. Attending to construct validity, the ACI was correlated with multiple scales with specific theorized outcomes. Specifically, the ACI positively correlated with a measure of relational closeness ($r = .25$), negatively correlated with a measure of psychological distance ($r = -.22$), positively correlated with psychological affection ($r = .58$), and was uncorrelated with social desirability (.05; see Floyd & Morman, 1998). Scores on the ACI also discriminated between relationships known in advance to be highly affectionate and non-affectionate (Floyd & Morman, 1998), and between biological and nonbiological family relationships (Floyd & Morman, 2003b).

Availability

The scale was published in Floyd and Mikkelson (2005) and is free for use with appropriate citation. The 18 items are included here with instructions for administration.

Sample Studies

The ACI has been used to index affectionate communication in a variety of personal relationships, including marital and dating relationships (Punyanunt-Carter, 2004), siblings (Myers, Byrnes, Frisby, & Mansson, 2011), siblings-in-law (Floyd & Morr, 2003), father–son relationships (Floyd & Morman, 2003a), grandparents and grandchildren (Mansson, 2012), and parent–child relationships within Asian-American families (Park, Vo, & Tsong, 2009). In nearly every case, indicators of relational closeness—including trust, listening, and shared conversation—have been positively related to the reported expression of affection. Affectionate communication is also typically more frequent in genetic forms of a relationship (e.g., biological siblings) than in nongenetic forms (e.g., stepsiblings), even when controlling for differences in affective closeness. Several other studies have associated ACI scores with indices of the body's stress response, including cortisol (Floyd *et al.*, 2007) and dehydroepiandrosterone sulfate (Floyd & Riforgiate, 2008); with mental well-being (Schrodt, Ledbetter, & Ohrt, 2007); and with emotional impairments such as alexithymia (Hesse & Floyd, 2008). Collectively, these studies support the conclusion that both receiving and expressing affection—in close relationships, at least—contribute to individual health and wellness.

Critique

Despite evidencing multiple forms of psychometric adequacy, the ACI is limited to a constellation of behaviors that emerged as indicators of affection during an inductive analysis of data collected from a largely Caucasian, middle-class American sample. As such, it does not index idiosyncratic affectionate expressions (such as personal idioms) and may not include behaviors important for the expression of affection in other cultures. Moreover, the measure assesses the perceived frequency of various affectionate behaviors without accounting for potential differences in their potency. Although some behaviors (kissing on the lips) may have more intimate connotations than others (helping with a task), these differences are not accounted for.

References

Floyd, K., & Mikkelson, A. C. (2005). The affectionate communication index. In V. Manusov (Ed.), *The sourcebook of nonverbal measures: Going beyond words* (pp. 47–56). Mahwah, NJ: Lawrence Erlbaum Associates.

Floyd, K., Mikkelson, A. C., Tafoya, M. A., Farinelli, L., La Valley, A. G., Judd, J., … Wilson, J. (2007). Human affection exchange: XIV. Relational affection predicts resting heart rate and free cortisol secretion during acute stress. *Behavioral Medicine*, *32*, 151–156. doi:10.3200/BMED.32.4.151-156

Floyd, K., & Morman, M. T. (1998). The measurement of affectionate communication. *Communication Quarterly*, *46*, 144–162. doi:10.1080/01463379809370092

Floyd, K., & Morman, M. T. (2003a). Fathers' and sons' reports of fathers' affectionate communication: Implications of a naïve theory of affection. *Journal of Social and Personal Relationships*, *22*, 99–109. doi:10.1177/0265407505049323

Floyd, K., & Morman, M. T. (2003b). Human affection exchange: II. Effects of role and biological status on affection in father-son relationships. *Journal of Social Psychology*, *143*, 599–612. doi:10.1080/00224540309598466

Floyd, K., & Morr, M. C. (2003). Human affection exchange: VII. Affectionate communication in the sibling/spouse/sibling-in-law triad. *Communication Quarterly*, *51*, 247–261. doi:10.1080/01463370309370155

Floyd K., & Riforgiate, S. (2008). Affectionate communication received from spouses predicts stress hormone levels in healthy adults. *Communication Monographs*, *75*, 351–368. doi:10.1080/03637750802512371

Hesse, C., & Floyd, K. (2008). Affectionate experience mediates the effects of alexithymia on mental health and interpersonal relationships. *Journal of Social and Personal Relationships*, *25*, 793–810. doi:10.1177/0265407508096696

Mansson, D. H. (2012). A qualitative analysis of grandparents' expressions of affection for their young adult grandchildren. *North American Journal of Psychology*, *14*(2), 207–219.

Myers, S. A., Byrnes, K. A., Frisby, B. N., & Mansson, D. H. (2011). Adult siblings' use of affectionate communication as a strategic and routine relational maintenance behavior. *Communication Research Reports*, *28*, 151–158. doi:10.1080/08824096.2001.565276

Park, Y. S., Vo, L. P., & Tsong, Y. (2009). Family affection as a protective factor against the negative effects of perceived Asian values gap on the parent-child relationship for Asian American male and female college students. *Cultural Diversity and Ethnic Minority Psychology*, *15*, 18–26. doi:10.1037/a0013378

Punyanunt-Carter, N. M. (2004). Reported affectionate communication and satisfaction in marital and dating relationships. *Psychological Reports*, *95*, 1154–1160. doi:10.2466/pr0.95.3f.1154-1160

Schrodt, P., Ledbetter, A. M., & Ohrt, J. K. (2007). Parental confirmation and affection as mediators of family communication patterns and children's mental well-being. *Journal of Family Communication*, *7*, 23–46. doi:10.1080/15267430709336667

Scale

The Affectionate Communication Index (Floyd & Morman, 1998)

Source: Floyd and Morman (1998). Reproduced with permission of Taylor & Francis.

We would like you to think about how you express love or affection to this person. That is, how do you let this person know that you love him or her? To what extent would you say that you do each of the following things *as a way to express affection to him or her*? Indicate your response by writing the appropriate number on the line preceding each item, according to the scale below.

1	2	3	4	5	6	7
Never or Almost Never Do This						Always or Almost Always Do This

Verbal

1) Say "I love you"
2) Say how important he or she is to you
3) Say he or she is one of your best friends
4) Say "I care about you"
5) Say he or she is a good friend

Nonverbal

6) Kiss on lips
7) Hug him or her
8) Wink at him or her
9) Hold his or her hand
10) Kiss on cheek
11) Put your arm around him or her
12) Sit close to him or her
13) Give him or her a massage or backrub

Social Support

14) Help him or her with problems
15) Acknowledge his or her birthday
16) Praise his or her accomplishments
17) Share private information
18) Give him or her compliments

Note: Instructions are written to refer to a particular person or relationship type. The researcher should specify this person or relationship prior to introducing the included instructions. Scores for the verbal, nonverbal, and support subscales are calculated by aggregating the scores of the individual items comprising each subscale. A total score for affectionate communication can be calculated by aggregating the scores of all 18 items in the ACI.

Profile 7

Attributional Complexity Scale (ACS)

(Fletcher, Danilovics, Fernandez, Peterson, & Reeder, 1986)

Profiled by: Michael Navarro, MA

Louisiana State University and Agricultural & Mechanical College

Construct

Attributional complexity is the degree to which an individual prefers complex (to simple) explanations for human behavior (Fletcher, Danilovics, Fernandez, Peterson, & Reeder, 1986).

Instrument Type

Self-Report

Description

The Attributional Complexity Scale (ACS) is a 28-item, self-report scale designed to measure seven primary attributional constructs (Fletcher *et al.*, 1986). The *motivational component* refers to an individual's innate curiosity to explain and understand other people. *Preference for complex explanations* is a tendency to include multiple causes in one's explanation of behavior, rather than a single cause. *Metacognition* is the "tendency to think about the underlying processes involved in causal attribution" (p. 876). *Behavior as a function of interaction* refers to an awareness that others' behavior is partially a function of the social situation, including the specifics involved when interacting with another. *Complex internal explanations* comprise a tendency to explain behavior using a relatively abstract and/or causally complex set of internal dispositions like beliefs, attributes, and abilities. Similarly, *complex contemporary external explanations* refer to a tendency to explain behavior using external causes that are more removed from the

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

immediate environment. Finally, *use of temporal dimension* refers to attributionally complex individuals being more likely to generate causal attributions that are removed in time. Attributionally complex individuals are able to (and prefer) think(ing) deeply about causes for behavior, whereas attributionally simple individuals are limited to rudimentary explanations for behavior.

Administration

The ACS is a self-administered questionnaire that takes between 5 and 10 minutes to complete. In the original study, participants were asked to complete the Person Perception Questionnaire, which was described as an instrument "designed to investigate the different ways that people think about themselves and other people" (Fletcher *et al.*, 1986, p. 877). Each of the 28 items is followed by scaling options ranging from −3 (*strongly disagree*) to +3 (*strongly agree*).

Scoring

The ACS is intended to produce a single score that represents an individual's level of attributional complexity. After the 14 reverse-scored items are recoded, items are summed to produce a score that will range from −84 to +84, with higher scores corresponding to higher levels of attributional complexity. If a different scaling option is used (e.g., 1–7), then the range will change (e.g., 28–196).

Development

Fletcher *et al.* (1986) proposed the ACS to help reconcile the seemingly opposing views found, at the time, in the attribution literature. On one hand, some argued that individual perceivers were cognitive misers, relying on mental shortcuts or heuristics when making judgments. On the other hand, some argued that individual perceivers were akin to naïve scientists, carefully and systematically weighing judgment decisions to generate sophisticated attributions. The ACS was developed with the idea that both are true (i.e., people are "motivated tacticians"; Fiske & Taylor, 1991; Operario & Fiske, 1999) and that individual differences in attributional complexity can explain why some research finds perceivers operating as cognitive misers and other perceivers operating as naïve scientists. As Fletcher *et al.* put it, "'some people are simpletons and others are experts" (p. 882). The scale has since been used in hundreds of attribution-based research studies, mainly as a predictor of judgment type or as a moderator variable.

Fletcher *et al.* (1986) published results from five studies that generated and subsequently provided validity evidence for the ACS. The first study checked the face validity and comprehensibility of an initial set of 45 questions, narrowing the scale to 28 items. The second study was a check of convergent and discriminant validity. The first two studies used psychology student samples. To check for concurrent validity, a third study was conducted to compare the responses of psychology students and students of the natural sciences. The final two studies examined the external validity of the ACS by testing the hypothesis that "the more attributionally complex a person is, the more likely that person will be to generate complex explanations for human behavior" (Fletcher *et al.*, 1986, p. 881).

All items were written to tap one of the seven attributional components described here. In particular, people vary in their motivation to understand behavior; preference for complex over simple explanations for behavior; tendencies to think about their own thinking; awareness of how interactions influence individual behavior; and tendencies to infer behavior from complex (versus simple) internal, external, and temporal dimensions.

A total of 289 undergraduates at Illinois State University completed the final 28 items, with a subset of these participants (102 enrolled in a large introductory psychology course) asked to complete the scale a second time (18 days after the first administration). An initial principal component analysis generated a single component that accounted for 21.4% of the item variance. Although items were internally consistent (alpha = .85), factor loading values suggest several of the items are weaker indicators than others. In particular, items 5, 9, 14, and 25 had loadings below .30; these and other items also had relatively low item-total correlation coefficients. In addition, although reliability estimates are reported for Studies 2–5, no further factor analytic procedures were performed.

Reliability

As a unidimensional scale, the ACS generates scores with internal consistency values above .80 in most studies. Test–retest reliability was estimated by Fletcher *et al.* (1986) at .80 using an 18-day lag time between scale administrations ($N = 102$). Although most studies have treated the scale as unidimensional, some have generated subscales in line with the seven attributional components that make up the construct (e.g., Follett & Hess, 2002). Internal consistency estimates of subscale scores are generally lower than the total scale. Fletcher *et al.* reported subscale reliabilities between .39 (Complex Internal) and .68 (Motivational Component); and Fast, Reimer, and Funder (2008) reported subscale reliabilities between .50 and .65 (total scale = .88). Correlations between subscales are generally moderate in magnitude ($r_{ave} = .40–.48$).

Validity

In the original publication, Fletcher *et al.* (1986) provided evidence for several types of validity for the scale. In Study 2, he showed that the ACS correlated significantly with need for cognition but not with social desirability, internal-external locus of control, dogmatism, or ACT scores, providing evidence of convergent and discriminant validity. In Study 3, he reported that ACS scores were higher for psychology than natural science majors, providing construct validity evidence.

The ACS purports to measure the degree to which an individual makes complex attributions of others' behavior, suggesting that the key validity criterion should be whether scale scores correlate highly with actual attributions. Fletcher's fourth study found initial evidence for such convergent validity, showing that individuals scoring higher on the ACS produced more spontaneous causal statements in written descriptions of their friends than did individuals with lower ACS scores. Likewise, in his fifth study, Fletcher reported a correlation of .31 between ACS scores and the degree to which participants chose a more complex causal attribution for hypothetical behavioral descriptions. Other research has shown that attributionally complex individuals are less likely to commit social judgment errors (i.e., any judgment of an experimental stimulus that goes against the judgment process under question), can achieve greater accuracy when making judgments (Wilson, Levine, Cruz, & Rao, 1997), and are judged to be more socially skilled (Fast *et al.*, 2008).

Availability

The ACS is reprinted at the end of this profile with permission from the American Psychological Association, who holds the copyright for the ACS. The scale is free to use for research purposes.

Sample Studies

The construct of attributional complexity has been explored in a variety of different disciplines. One of its many uses has been in applied areas, such as the sales floor and the courtroom. Porter and Inks (2000) found that salespeople who rate high in attributional complexity are more likely to engage in a technique known as *adaptive selling* (see Active-Empathic Listening Scale profile, Profile 2). This concept refers to an individual's ability to gather information and tailor sales presentations for each individual customer, recognize the customer's reaction, and make immediate adjustments (Weitz, Sujan, & Sujan, 1986). The ACS also has been used in research focusing on jury selection processes. Pope and Meyer (1998) found that individuals scoring lower in attributional complexity were more likely to find a defendant guilty both before and after the presentation of evidence than individuals who scored higher.

Other work has focused on exploring possible links between attributional complexity and depression. Flett, Pliner, and Blankstein (1989) utilized individual components of the ACS in conjunction with the Beck Depression Inventory and noted that depressed individuals exhibited more desire to engage in complex, external attributional processing. Marsh and Weary (1989) found that the relationship between attributional complexity and depression exhibited an inverted-U pattern, with mildly depressed individuals reporting high ACS scores and nondepressed and severely depressed individuals having the lowest scores.

Attributional complexity also has been utilized as a possible explanation for the presence of other notable psychological effects. Devine (1989) found that individuals high in attributional complexity and motivation were less likely to fall victim to the overattribution effect, uncovering what the author claimed was a "created" bias in the traditional measurements of the overattribution effect in previous research. Similarly, Wilson *et al.* (1997) noted that attributional complexity had a moderating effect on individuals' susceptibility to the actor–observer bias in recalling unfulfilled obligations, with high-complexity individuals attributing these obligations more to the actor and less toward external factors. Other research has shown attributional complexity to be an indicator of emotional intelligence, particularly when combined with empathy (Fitness & Curtis, 2005).

Critique

Although the ACS was designed to measure behavioral attribution and thus should be conceptually distinct, it shares much in common with other constructs such as cognitive complexity (see the Role Category Questionnaire, Profile 56) and self-complexity. The degree to which the ACS measures something conceptually or operationally distinct is thus an open question. In addition, although some work shows ACS scores are largely independent from scores on standardized academic tests, the degree to which it is related to intelligence is an open question. Fletcher used the ACT as a measure of intelligence, but the ACT measures academic aptitude and is a function of schooling more than of general intelligence (even though its design *was* based on the IQ test). Finally, the correlation between ACS scores and actual judgments is typically small to moderate in magnitude, suggesting that actual behavior and reported behavior are not isomorphic. The degree to which the ACS is a measure of behavioral tendencies or what people think they do is thus uncertain.

References

Devine, P. G. (1989). Overattribution effect: The role of confidence and attributional complexity. *Social Psychology Quarterly*, *52*, 149–158. doi:10.2307/2786914

Fast, L. A., Reimer, H. M., & Funder, D. C. (2008). The social behavior and reputation of the attributionally complex. *Journal of Research in Personality*, *42*, 208–222. doi:10.1016/j.jrp.2007.05.009

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

Fitness, J., & Curtis, M. (2005). Emotional intelligence and the Trait Meta-Mood Scale: Relationships with empathy, attributional complexity, self-control, and responses to interpersonal conflict. *E-Journal of Applied Psychology: Social Section*, *1*(1), 50–62.

Fletcher, G. J. O., Danilovics, P., Fernandez, G., Peterson, D., & Reeder, G. D. (1986). Attributional complexity: An individual differences measure. *Journal of Personality and Social Psychology*, *51*, 875–884. doi:10.1037/0022-3514.51.4.875

Flett, G. L., Pliner, P., & Blankstein, K. R. (1989). Depression and components of attributional complexity. *Journal of Personality and Social Psychology*, *56*, 757–764. doi:10.1037/0022-3514.56.5.757

Follett, K. J., & Hess, T. M. (2002). Aging, cognitive complexity, and the fundamental attribution error. *The Journals of Gerontology: Social Sciences*, *57*, 312–323. doi:10.1093/geronb/57.4.P312

Marsh, K. L., & Weary, G. (1989). Depression and attributional complexity. *Personality and Social Psychology Bulletin*, *15*, 325–336. doi:10.1177/0146167289153004

Operario, D., & Fiske, S. T. (1999). Social cognition permeates social psychology: Motivated mental processes guide the study of human social behavior. *Asian Journal of Social Psychology*, *2*, 63–78.

Pope, J., & Meyer, M. (1998). Attributional analysis of jurors' judgements in a criminal case: A preliminary investigation. *Social Behavior and Personality: An International Journal*, *27*, 563–574. doi:10.2224/sbp.1999.27.6.563

Porter, S. S., & Inks, L. W. (2000). Cognitive complexity and salesperson adaptability: An exploratory investigation. *Journal of Personal Selling & Sales Management*, *20*, 15–21. doi:10.1080/08853134.2000.10754217

Weitz, B. A., Sujan, H., & Sujan, M. (1986). Knowledge, motivation, and adaptive behavior: A framework for improving selling effectiveness. *Journal of Marketing, 50*, 174–191. doi:10.2307/1251294
Wilson, S. R., Levine, K. J., Cruz, M. G., & Rao, N. (1997). Attribution complexity and actor-observer bias. *Journal of Social Behavior and Personality, 12*, 709–726.

Scale

Attributional Complexity Scale (Fletcher *et al.*, 1986)

Source: Fletcher *et al.* (1986). Reproduced with permission of American Psychological Association.

Instructions: This questionnaire has been designed to investigate the different ways that people think about themselves and other people. The questionnaire is anonymous, so there is no need to put your name on it. There are no right or wrong answers. We are interested in your own perceptions. Please answer each question as honestly and accurately as you can, but don't spend too much time thinking about each answer.

For each item listed below, please indicate your level of agreement or disagreement using the following scale:

−3 = strongly disagree
−2 = moderately disagree
−1 = slightly disagree
0 = neither agree nor disagree
+1 = slightly agree
+2 = moderately agree
+3 = strongly agree

<u>Motivational component</u>

1) I don't usually bother to analyze and explain people's behavior.
2) I don't enjoy getting into discussions where the causes for people's behavior are being talked over.*
3) I really enjoy analyzing the reasons or causes for people's behavior.
4) I am not really curious about human behavior.*

<u>Preference for complex explanations</u>

5) Once I have figured out a single cause for a person's behavior I don't usually go any further.*
6) I have found that the causes for people's behavior are usually complex rather than simple.
7) I usually find that complicated explanations for people's behavior are confusing rather than helpful.*
8) I prefer simple rather than complex explanations for people's behavior.*

<u>Metacognition</u>

9) I believe it is important to analyze and understand our own thinking processes.

10) I am very interested in understanding how my own thinking works when I make judgments about people or attach causes to their behavior.
11) I give little thought to how my thinking works in the process of understanding or explaining people's behavior.*
12) When the reasons I give for my own behavior are different from someone else's, this often makes me think about the thinking processes that lead to my explanations.

Behavior as a function of interaction

13) I think a lot about the influence that I have on other people's behavior.
14) I think very little about the different ways that people influence each other.*
15) I think very little about the influence that other people have on my behavior.*
16) I believe that to understand a person you need to understand the people who that person has close contact with.

Complex internal explanations

17) I have found that the relationships between a person's attitudes, beliefs, and character traits are usually simple and straightforward.*
18) To understand a person's personality/behavior I have found it is important to know how that person's attitudes, beliefs, and character traits fit together.
19) I have thought a lot about the way that different parts of my personality influence other parts (e.g., beliefs affecting attitudes or attitudes affecting character traits).
20) I tend to take people's behavior at face value and not worry about the inner causes for their behavior (e.g., attitudes, beliefs, etc.).*

Complex contemporary external explanations

21) If I see people behaving in a really strange or unusual manner I usually put it down to the fact that they are strange or unusual people and don't bother to explain it any further.*
22) When I try to explain other people's behavior I concentrate on the person and don't worry too much about all the existing external factors that might be affecting them.*
23) I think a lot about the influence that society has on other people.
24) I think a lot about the influence that society has on my behavior and personality.

Use of temporal dimension

25) I have thought a lot about the family background and personal history of people who are close to me, in order to understand why they are the sort of people they are.
26) I have often found that the basic cause for a person's behavior is located far back in time.
27) When I analyze a person's behavior I often find the causes form a chain that goes back in time, sometimes for years.
28) I have thought very little about my own family background and personal history in order to understand why I am the sort of person I am.*

Note: Labels should be removed and items randomized prior to administration. Item marked with an asterisk (*) should be scored in the reverse direction.

Profile 8

Audio Message Complexity: Audio Content Change (Acc) and Audio Information Introduced (Aii)

(Lang, Gao, Potter, Lee, Park, & Bailey, 2015)

Profiled by: Robert F. Potter, PhD and Annie Lang, PhD

Indiana University

Construct

Audio complexity is the amount of cognitive resources automatically called by structural features of a message and the extent to which those resources are required to process that message.

Instrument Type

Message Coding Scheme; Behavioral Observation

Description

Audio Content Change (Acc) and Audio Information Introduced (Aii) are based on the assumption that the best way to operationalize message complexity is by considering the human cognitive system processing the content. Complexity is therefore conceptualized as an assessment of the amount of cognitive resources automatically called by structural features of the message and the extent to which those resources are required to process the message. Acc identifies structural features within the message known to cause an orienting response (OR). The orienting response is an evolved, hard-wired reaction to environmental change or signal stimuli and automatically delivers cognitive resources to message encoding. For each occurrence of an Acc, the coder determines the extent to which the change introduces novelty, motivational relevance, and/or an auditory change that is a learned media-literacy signal. Each of eight Aii attributes are coded in a binary fashion, and the results summed to obtain a relative quantification of the level of resources required following each Acc. Thus, Acc scores range from

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

those with low local complexity (i.e., require few resources) to those with high local complexity (i.e., require many resources). Global message complexity measures also can be calculated by summing Acc and Aii across the entire message and dividing by the message duration in seconds. Acc provides a global index of the resources automatically allocated by the listener, and Aii provides the global index of resources required.

Administration

Prior to coding a message, coders listen to the entire message in order to get a general sense of the content. Acc is coded first. To do so, coders list each time one of six auditory structural features occur, namely voice changes, music onset, sound effect onset, production effect onset, silence onset, and voice onset. Each has been shown to cause orienting responses (Potter, Lang, & Bolls, 2008) and are thought to elicit the same amount of initial resource allocation to message processing. Next, to quantify Aii, coders listen closely to the six seconds (6 s) of audio immediately following each Acc and count the number of dimensions in which a change has occurred.

Scoring

Acc is scored by counting the number of times the following structural features happen in a message:

- *Voice change*: The replacement of one speaker by another in the auditory stream (Potter, 2000; Potter, Jamison-Koenig, Lynch, & Sites, 2016).
- *Music onset*: The beginning of music in the audio message.
- *Sound effect onset*: The initial sound representing a concrete object or action, such as a doorbell ringing or a vacuum cleaner.
- *Production effect onset*: Production effects are synthesized sounds that are used to provide auditory emphasis but are not associated with concrete objects or actions. Examples include echo, laser sounds, and noise gates (Potter *et al.*, 2008).
- *Emotional word onset*: Words with significant emotional connotation as defined by the Affective Norms of English Words database (ANEW; Bradley & Lang, 1999).
- *Silence onset*: The cessation of sound for at least 2 seconds.
- *Voice onset*: The replacement of any nonspoken content with a speaking voice.

After the Accs are identified, the 6 seconds of audio following each is listened to in order to quantify the amount of resources required to process the information it introduced. To do this, the 6 seconds prior to the Acc is compared to the 6 seconds following the Acc on each of eight dimensions to determine if, for each dimension, a change is present (1) or absent (0). The dimensions are:

- *New content*: A specific voice or piece of audio content occurs for the first time in the message.
- *Unrelated content*: The content after the Acc is not topically related to that before the Acc.
- *Form change across the Acc*: An electronically created tonal or timbral difference occurs between the two time windows that begin at the point of the Acc.

- *Form change in previous occurrence of sound*: If the tone or timbre of a previously heard sound or voice is different compared to the last time *that* sound or voice was heard in the message.
- *Ambient sound change*: A difference in the general environmental sounds on either side of the Acc.
- *Ambient sound change behind a previous occurrence of sound*: Whether the environmental sounds behind a focal sound are different post Acc compared to the last time the focal sound was heard.
- *Inherently emotional sound*: Whether the post-Acc sound is motivationally relevant. This can be a sound that is emotionally compelling—such as the onset of a scream, a gunshot, or a baby laughing—or sounds related to threat or opportunity as in growls or sexual sounds, respectively.
- *Emotional change*: A change in the intensity or direction (pleasant, unpleasant) of the emotional tone after the Acc.

Each dimension is scored as *present* (1) or *absent* (0), so the value of Aii for each Acc can range from 0 to 8.

Development

The development of the Acc construct was based on the well-known phenomenon of the orienting response, which is an evolved survival mechanism that automatically responds to novel and/or learned signal stimuli in the environment by allocating a small amount of processing resources to information encoding (Graham, 1979). Communication scholars conceptualize the human brain as being unable to initially distinguish mediated messages from real-life events (Reeves & Nass, 1996). From this perspective, it is assumed that people will respond to media like real life, and hence mediated content that signals relevant information or is novel should elicit orienting responses. Data support this assumption (Lang, 1990; Lang, Geiger, Strickwerda, & Sumner, 1993; Lang, Bolls, Potter, & Kawahara, 1999).

Reliability

Because audio message complexity is a product of a message, judges have to be trained on the task of rating messages for Acc and Aii. Due to the nascent nature of these measures, the published data on the reliability of the indices are limited. Lang *et al.* (2015) used Acc and Aii in two studies and reported high reliability (.91) in Study 2. The first study did not report reliability coefficients; instead, it reported that consensus rating of the radio stimuli was done in teams of three or six researchers.

Validity

Potter's (2000) work provided substantial convergent validity to the primary claim of the Acc construct—namely, that changes in the auditory stream result in orienting responses that deliver automatic resource allocation to message encoding. Potter showed that the most common of the auditory structural features, the voice change, results in decelerating heart rates indicative of the OR and that this result did not habituate after six repetitions in 2-minute messages. Potter *et al.* (2008) demonstrated similar heart rate patterns following many of the other Acc structural features, including commercial onsets, jingle onsets, sound effects, and emotional words.

The claim that ORs elicited by Accs deliver the same initial amount of cognitive resources was validated by Lang *et al.* (2015) using secondary task reaction times (STRTs) as a key dependent variable. Subjects were told to listen closely to audio messages (the primary task) and respond with a button press as quickly as they could when they heard a brief 1000 Hz auditory probe (the secondary task). They predicted that if, indeed, the automatic allocation of resources resulted in the same initial amount applied to the auditory novelty, then there would be no significant difference in reaction time to probes placed after Accs with identical resource requirements (e.g., ones with the same levels of Aii). This prediction was supported. Furthermore, Lang *et al.* (2015) provided additional support for the human-centered nature of the Aii construct by demonstrating that dimensions which are novel and emotional both call for, and require, additional automatic allocation of resources to encoding.

Availability

The measures are described in Lang *et al.* (2015). A detailed codebook is available by contacting the authors. Both Acc and Aii may be used for free with appropriate citations.

Sample Studies

To date, there has been only one published manuscript that utilized Acc and Aii to test hypotheses (Lang *et al.*, 2015). The first study reported in that article was designed to test the measures at a global level using a mixed 2 (Accs/s) × 2 (Aii/s) × 4 (Message) design. These were within-subject factors, with Acc/s and Aii/s completely crossed with two levels—high and low. The Message factor provided repetitions to counter message-specific results in the statistical analyses. Subjects were randomly assigned to one of four presentation orders, which served as the between-subjects factor.

Subjects ($N = 82$) heard each message through headphones and provided self-reported data following each message. After all the messages and a short distraction task, subjects completed a forced-choice yes/no recognition memory task for information within the messages. Three target and three foil sentences (with two words from the sentence altered) were provided for each message. Dependent measures were the percentage correct of targets and foils as well as response latency.

Latency results showed that, as predicted, when Acc/s was high the recognition probes were responded to more quickly than when Acc/s was low. Recognition accuracy analyses returned a significant interaction between Acc/s and Aii/s. At low levels of Acc/s, an increase in the Aii/s did not significantly affect accuracy. Subjects best recognized information when Acc/s was high and Aii/s low. In contrast, poorest recognition occurred when both global measures were at the highest levels.

Study 2 focused on the more common auditory structural features of the Voice Change (VC) and the Voice Onset (VO) to test Acc and Aii as local measures (Lang *et al.*, 2015). In other words, the goal of this experiment was to "assess the impact of the amount and type of information introduced by a specific type of Acc at a specific point in time in the message" (Lang *et al.*, 2015, p. 768). The experimental design was a 9 (Aii Type) × 5 (Repetition) within-subjects design. Nine different pairings of specific Aii combinations were created in this study: VO, VC, Voice change to new voice (VCN), Voice onset to new voice (VON), Voice change with emotion change (VCE), Voice change with a form change (VCF), Voice change with natural sound change (VCNS), and Voice change to a new voice with a form change (VCFN). The first prediction was that there would be no differences in STRTs following VCs and VOs—since both involved previously heard voices—and that recognition data would also not be significantly different following the two Accs, given that the resources required would be identical. This was supported for both dependent variables.

A second prediction in Lang *et al.* (2015) was that the introduction of novelty resulting from VCs or VOs involving new voices would result in an additional automatic allocation of processing resources (perhaps through a second OR elicited due to novelty as opposed to auditory change per se) but that those additional resources would then *be required* to fully process the novelty. This was supported by a non–statistically significant difference in the STRTs between VCs or VOs involving previously heard voices compared to the same Accs involving new voices coupled with significantly greater recognition data for the latter.

A similar third prediction suggested that a voice change that also involved a motivationally relevant emotion change (VCE) would automatically elicit additional resource allocation but that the emotional content would require the additional resources to be fully processed. Parallel to the findings for the second prediction, STRTs were not significantly different for VCs and VCEs, but recognition data were greater for the VCE content.

Critique

As mentioned in this profile, little empirical data exist demonstrating psychonomic properties of the Acc and Aii measures. Even the single study that does report reliability values (Lang *et al.*, 2015) only does so for the entire Aii construct and not for individual dimensions. It is possible, and perhaps likely, that the validity for these dimensions varies substantially—suggesting the need for further refinement of their conceptual definitions.

Furthermore, recent work done by Potter, Lynch, and Kraus (2015) suggests further development of Acc may be necessary. Results from Potter *et al.* (2015) showed that—unlike the voice change—orienting to some Accs may habituate after repeated occurrence in a message. Potter *et al.* repeated music onsets and production effect onsets across 40 minutes of a simulated radio broadcast and found that cardiac orienting was less reliable later in the message. Whether this is related to the much longer duration of the stimulus in this experiment compared to Lang *et al.* (2015) remains an open question. If, however, some Accs do not result in cardiac orienting after repeated presentation, then coding them as identical to the Voice Change in the Acc measure is not optimal.

Potter *et al.* (2016) demonstrated the need for further refinement of the Aii measure as well. Using voices of different fundamental frequencies to constitute voice changes, they showed that both cardiac orienting and recognition memory are affected by the similarity of the sound of the voices involved in the VC. This is unaccounted for in the Aii dimensions as presently constituted.

References

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words* [CD-ROM]. Gainesville, FL: NIMH Center for the Study of Emotion and Attention, University of Florida.

Graham, F. K. (1979). Distinguishing among orienting, defense, and startle reflexes. In H. D. Kimmel, E. H. V. Olst, & J. F. Orlebeke (Eds.), *The orienting reflex in humans* (pp. 137–167). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lang, A. (1990). Involuntary attention and physiological arousal evoked by structural features and emotional content in TV commercials. *Communication Research*, *3*, 275–299. doi:10.1177/009365090017003001

Lang, A., Bolls, P., Potter, R. F., & Kawahara, K. (1999). The effects of production pacing and arousing content on the information processing of television messages. *Journal of Broadcasting & Electronic Media*, *43*, 451–475. doi:10.1080/08838159909364504

Lang, A., Gao, Y., Potter, R. F., Lee, S., Park, B., & Bailey, R. L. (2015). Conceptualizing audio message complexity as available processing resources. *Communication Research*, *42*, 759–778. doi:0093650213490722.

Lang, A., Geiger, S., Strickwerda, M., & Sumner, J. (1993). The effects of related and unrelated cuts on television viewers' attention, processing capacity, and memory. *Communication Research*, *20*, 4–29. doi:10.1177/009365093020001001

Potter, R. F. (2000). The effects of voice changes on orienting and immediate cognitive overload in radio listeners. *Media Psychology*, *2*, 147–178. doi:10.1207/S1532785XMEP0202_3

Potter, R. F., Lang, A., & Bolls, P. D. (2008). Identifying structural features of audio: Orienting responses during radio messages and their impact on recognition. *Journal of Media Psychology: Theories, Methods, and Applications*, *20*(4), 168–177. doi:10.1027/1864-1105.20.4.168

Potter, R. F., Lynch, T., & Krause, A. (2015). I've heard *that* before: Habituation of the orienting response follows repeated presentation of audio structural features in radio. *Communication Monographs (82)*, 359–378. doi: 10.1080/03637751.2015.1019529

Potter, R. F., Jamison-Koenig, E., Lynch, T., & Sites, J. (2016). Effect of vocal-pitch difference on automatic attention to voice changes in audio messages. *Communication Research. Advance online publication.* doi:10.1177/0093650215623835

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge: Cambridge University Press.

Profile 9

Communicative Adaptability Scale (CAS)

(Duran, 1983)

Profiled by: Gwen A. Hullman, PhD

University of Nevada–Reno

Construct

The Communicative Adaptability Scale (CAS) was designed to measure an individual's ability to tailor communication behaviors. As a dispositional trait, "communicative adaptability is conceptualized as the ability to perceive socio-interpersonal relationships and adapt one's behaviors and goals accordingly" (Duran, 1992, p. 255).

Instrument Type

Self-Report; Other-Report

Description

The CAS is a six-dimension, 30-item self-report measure designed to assess social communicative competence. Respondents rate their agreement with statements on 5-point scales bounded by *never true* and *always true*. The items were designed to tap cognitive aspects of communicative competence. Five items represent each of the six dimensions: social confirmation, social experience, social composure, articulation, appropriate disclosure, and wit. *Social confirmation* is a skill that reflects one's ability to see a situation from another person's perspective (Duran & Wheeless, 1982). *Social experience* refers to the sum of different types of social experiences one has (Duran & Wheeless, 1982). *Social composure* evaluates the extent to which one operates in social situations without becoming anxious or nervous, whereas *articulation* refers to fluency and abilities related to vocal organization of ideas (Duran, 1983). *Appropriate disclosure* indicates awareness and adherence to social norms regarding disclosure (Chelune, 1975).

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Finally, *wit* refers to the ability to use humor when appropriate to diffuse uncomfortable feelings and tension (Duran, 1983). The other-report version of this test also is evaluated here, but typically, the measure is used as a self-report assessment.

Administration

Both the self-report and the other-report versions of the CAS are completed in less than 10 minutes. Items from the six dimensions should be randomized before administration. The scale needs no situational prompt.

Scoring

After reverse-coding designated items, a global score is calculated by averaging the scores of all 30 items. Individual averages for each of the six dimensions may be calculated to explore variability in specific areas of communicative adaptability.

Development

Duran and Wheeless (1982) initially developed the Social Management Scale (SMS), which measured self-perceived communication competence. Later, with the goal of increasing variance accounted for by the measure, Duran (1983) reconceptualized the SMS as the CAS, proposing four additional factors (social composure, wit, appropriate disclosure, and articulation). The original version of the CAS correlated to other established measures of communication flexibility and appropriateness (Duran, 1983). Other-report versions of the CAS have been used by revising the CAS with minor word changes (Duran & Zakahi, 1988; Hullman, 2007). Gareis, Merkin, and Goldman (2011) also utilized the CAS in both native and English language contexts.

Reliability

Cronbach's alphas across 14 samples demonstrate internal consistency of the CAS dimensions (Duran, 1983, 1992; Duran & Kelly, 1989, 1994; Duran & Zakahi, 1984, 1988; Hawken, Duran, & Kelly, 1991; Hullman, 2007; Martin & Rubin, 1994; Zakahi & Duran, 1985). Social experience alphas ranged from .76 to .89, $M = .82$; social confirmation alphas ranged from .75 to .92, $M = .85$; social composure alphas ranged from .75 to .88, $M = .81$; appropriate disclosure alphas ranged from .66 to .85, $M = .75$; articulation alphas ranged from .75 to .91, $M = .81$; and wit alphas ranged from .58 to .87, $M = .74$. Alpha ranges for dimensions stated in additional studies support these results (Zakahi, 1985, 1986).

Other-report alphas are similar to self-report alphas with ranges between .71 and .91 (Hawken, Duran, and Kelly, 1991; Hullman, 2007; Zakahi, 1985). Composite alphas also are acceptable (e.g., .85, Gareis *et al.*, 2011; .79, .80, Duran & Zakahi, 1988; .79, Zakahi & Duran, 1984).

Validity

Many studies have provided validity evidence for the CAS using adult and student samples. Spitzberg (2003) stated that "the CAS has displayed consistent factor structure, acceptable psychometrics, and has generally related to other measures as predicted" (p. 108). Duran (1992) summarized research that supported relations between the CAS and communication apprehension, shyness, loneliness, communication satisfaction, attraction, cognitive complexity, interaction involvement, and androgyny. For example, a composite score of the CAS positively correlated to scores of relational satisfaction (Duran & Zakahi, 1988; Zakahi & Duran, 1984), attraction (Zakahi & Duran, 1984), assertiveness (Zakahi, 1985), persistence in college (Hawken *et al.*, 1991), and loneliness (McCroskey, 1970; Zakahi, 1986). Higher social experience, articulation, and social composure scores correlated with participants' lack of shyness (Duran & Kelly, 1989). Social confirmation scores and appropriate disclosure scores correlated to responsiveness (Duran & Kelly, 1988). Later research concurs. Adaptability factors are related to cognitive flexibility (Hullman, 2007), appropriateness (Hullman, 2007), loneliness (Gareis *et al.*, 2011), and friendship satisfaction (Gareis *et al.*, 2011). The CAS scores relate to other measures as we would expect them to, which supports the convergent validity of the scale.

Although the convergent and predictive validity of the CAS are supported, the internal factor structure is not always confirmed. For example, Duran (1983) reported that the social composure items collapsed with social experience items for an adult population but remained separate for the student sample. Hullman (2007) reported that social experience items cross-loaded onto both social composure and social confirmation items for a student sample. Beatty, Marshall, and Rudd (2001) also reported that social experience items loaded with social composure items to produce a five-factor structure for a noncollege sample. It appears that the proposed factor structure is supported in most samples, but five dimensions are found for other samples. Studies using an adult sample are not as prolific as those that have sampled college students, so a comparison across sample types is not possible at this time.

Duran (1992) stated that social experience and social composure most likely influenced one another in a cyclical fashion, and perhaps that relationship is illustrated in their shared variance. Generally speaking, earlier studies supported the six-factor structure and that finding may be a result of the factor analysis procedures employed. Most early studies did not report a factor analysis procedure. A few datasets were subjected to principal component analysis or principal factor analysis with oblique rotation (Duran, 1983). Later examinations, however, of the CAS used confirmatory factor analysis on self- and other-report versions (Hullman, 2007), which also could explain differences in factor structure.

Availability

The scale is reproduced below (with permission) and can be found in Duran's (1992) original article published in *Communication Quarterly*. The scale is free to use for research purposes.

Sample Studies

Zelko (1954) stated that listening requires the ability to adapt to the physical situation. Witkin and Trochim (1997) explicitly stated clusters of skills that comprise the listening construct. These included, but were not limited to, sensory impressions, context, overt response, empathy, organization, interpretation, information storage, subliminal cues, and nonverbal stimuli. The skills here closely align with those found in the CAS. Furthermore, Imhof (2012) found that individuals vary widely in their listening profiles depending on context, which suggests that adaptability across contexts is important for listening competence (see also Gearhart, Denham, & Bodie, 2014). Scholars continue to illustrate the unique applicability of the CAS in a variety of interpersonal communication situations.

Higher adaptability scores generally relate positively to healthy relational outcomes, such as satisfaction measures and participation in social activities (Duran & Kelly, 1994). For example, Gareis *et al.* (2011) reported that international students coming to the United States who are high in English adaptability have more friends and are more satisfied with those friendships, a finding that echoes earlier conclusions that adaptive roommates are more satisfied with their roommates than are less adaptive roommates (Duran & Zakahi, 1988).

Adaptability scores are related to other communication traits in many studies. Gareis *et al.* (2011) reported that English proficiency and willingness to communicate are strongly correlated to English communicative adaptability. Furthermore, adaptabilities in one's native language and in English as a second language are positively related to one another. In addition, Hullman (2007) found that other-report adaptability and other-report appropriateness were related to one another, and that self-report adaptability is related to cognitive flexibility. Social experience, specifically, related to perceptiveness and attentiveness traits as well (Duran & Kelly, 1988).

Jebreen (2015) tested the CAS in a sample of novice software analysts who collaborated to determine the requirements of new software and hardware systems. Jebreen found that the analysts scored high on social experience, social composure, and social confirmation, but lower on articulation, appropriate disclosure, and composure. A possible explanation is the lack of training in those areas for those who study computer science.

Beatty *et al.* (2001) examined the heritability of communication competence by comparing scores of adaptability among twins. Based on the assumption that neurobiological systems are thought to underlie communication traits and behaviors, they concluded that social composure is 88% heritable (some social experience items loaded onto this factor), wit is 90% heritable, social confirmation is 37% heritable, and articulation and appropriate disclosure are 0% heritable.

Critique

The CAS seems to be more stable as a five-factor structure than as a six-factor structure. Wording of certain items and sample type may influence the stability of the factor structure. Covariations among social experience, social confirmation, and social composure items also present possible causes of factor instability in some recent analyses (Hullman, 2007). Across many studies, the social confirmation, social experience, and social composure dimensions seem to emerge as the most influential dimensions of the CAS. Although other dimensions display sufficient psychometric properties, the "big three" have the most predictive power.

Another explanation of the difference in significance across dimensions may lie in a point made by Duran (1992): Likert-type responses are fixed responses across a continuum of *always* doing something or *never* doing something. He stated that the response choice *sometimes* represents doing something only when it is called for, and may measure the most adaptive person. An examination of the items reveals wording differences in the social confirmation, social experience, and social composure items compared to the articulation, wit, and appropriate disclosure items. Most items in the "big three" are less likely to be considered inappropriate in any context, whereas some other items are clearly not always the most competent choice. For example, a social experience item is "I find it easy to get along with new people." It may be difficult to think of situations where answering *always* to this item is incompetent. Conversely, consider the items "I disclose at the same level that others disclose to me," "When I am anxious, I often make jokes," "I often make jokes in tense situations," and "At times, I don't use appropriate verb tense." A cursory review of the latter items highlights exceptions in which answering *always* or *never* would show a lack of adaptability.

As with all measures, limitations do exist with the CAS. The CAS is a self-report scale. As was illustrated in Hullman (2007) and Duran and Zakahi (1988), an individual's rating of self- adaptability differs (in the individual's favor) from a comparative other rating of the same target. Self-report CAS scores also are unrelated to observed conversational skill (Carrell & Wilmington, 1996). Therefore, researchers should be cognizant that the CAS self-report version measures the individual's perception of adaptability, which appears to be different from other peoples' perceptions of that individual's adaptability and perhaps actual adaptability in context (see also Bodie, Jones, Vickery, Hatcher, & Cannava, 2014). Many of the items are cognitively based, and one could argue that an individual's cognition is not always available to other people. Other people have only behavioral observations at their disposal when making judgments.

Rubin (1994) expressed concern about the content validity of the CAS. Given the vast set of skills competent communicators must have, six dimensions may not represent the complete set of adaptability skills. She further argued that wit and articulation very often are unrelated to other constructs in research, unlike the social dimensions of the CAS. This continues to be the case in more recent research and should be further examined.

Furthermore, although agreement with the items on the scale presents opportunities for people to develop adaptability, the items do not preclude a rigid, inflexible style. It may be difficult for those lacking competence to evaluate these elements as others would. For example, some people may enjoy novel social situations (an item representing social experience) even though they do not change their behavior across these situations. The CAS items serve as proxies to the notion of behavior changing across situations. The dimensions measure perceptions of skills that would enable a person to be adaptive, but they do not measure demonstrated behavioral skill across a multitude of different contexts. The items do, however, suitably represent the cognitive aspects of adaptability. A social confirmation item, for example, is "While I am talking, I think about how the other person feels" (Duran, 1992, p. 267).

Finally, the factor structure of the CAS may not be stable across all samples, but the individual items themselves do seem to provide a valuable contribution in measuring the cognitive aspects of communicative adaptability.

References

Beatty, M. J., Marshall, L. A., & Rudd, J. E. (2001). Twins study of communicative adaptability: Heritability of individual differences. *Quarterly Journal of Speech*, *87*, 366–377. doi:10.1080/00335630109384346

Bodie, G. D., Jones, S. M., Vickery, A. J., Hatcher, L., & Cannava, K. (2014). Examining the construct validity of enacted support: A multitrait-multimethod analysis of four perspectives for judging immediacy and listening behaviors. *Communication Monographs*, *80*, 495–523. doi:10.1080/03637751.2014.957223

Carrell, L. J., & Wilmington, S. C. (1996). A comparison of self-report and performance data in assessing speaking and listening competence. *Communication Reports*, *9*, 185–191. doi:10.1080/08934219609367650

Chelune, G. J. (1975). Self-disclosure: An elaboration of its basic dimensions. *Psychological Reports*, *36*, 79–85. doi:10.2466/pr0.1975.36.1.79

Duran, R. L. (1983). Communicative adaptability: A measure of social competence. *Communication Quarterly*, *31*, 320–326. doi:10.1080/01463378309369521

Duran, R. L. (1992). Communicative adaptability: A review of conceptualization and measurement. *Communication Quarterly*, *40*, 253–268. doi:10.1080/01463379209369840

Duran, R. L., & Kelly, L. (1988). An investigation into the cognitive domain of competence II: The relationship between communicative competence and interaction involvement. *Communication Research Reports*, *5*, 91–96. doi:10.1080/08824098809359806

Duran, R. L., & Kelly, L. (1989). The cycle of shyness: A study of self-perceptions of communication performance. *Communication Reports*, *2*, 30–38. doi:10.1080/08934218909367478

Duran, R. L., & Kelly, L. (1994). The role of social experience in the development of communication competence. *Communication Research Reports*, *11*, 119–126. doi:10.1080/08824099409359949

Duran, R. L. & Wheeless, V. (1982). Social management: Toward a theory based operationalization of communication competence. *Southern Communication Journal*, *48*, 51–64. doi:10.1080/10417948209372552

Duran, R. L., & Zakahi, W. R. (1984). Competence or style: What's in a name? *Communication Research Reports*, *1*, 42–47.

Duran, R. L., & Zakahi, W. R. (1988). The influence of communicative competence upon roommate satisfaction. *Western Journal of Speech Communication*, *52*, 135–146. doi:10.1080/10570318809389631

Gareis, E., Merkin, R., & Goldman, J. (2011). Intercultural friendship: Linking communication variables and friendship successes. *Journal of Intercultural Communication Research*, *40*, 153–171. doi:10.1080/17475759.2011.581034

Gearhart, C. C., Denham, J., & Bodie, G. D. (2014). Listening as a goal directed activity. *Western Journal of Communication*, *78*, 668–684. doi:10.1080/10570314.2014.910888

Hawken, L., Duran, R. L., & Kelly, L. (1991). The relationship of interpersonal communication variables to academic success and persistence in college. *Communication Quarterly*, *39*, 297–308. doi:10.1080/01463379109369807

Hullman, G. A. (2007). Communicative adaptability scale: Evaluating its use as an other-report measure. *Communication Reports*, *20*, 51–74. doi:10.1080/08934210701643693

Jebreen, I. (2015). Requirements determination as a social practice: Perceptions and Preferences of novice analysts. *Computer and Information Science, 8*, 134–144. http://dx.doi.org/10.5539/cis.v8n3p134

Imhof, M. (2012). Who are we as we listen? Individual listening profiles in varying contexts. *International Journal of Listening, 18*, 36–45. doi:10.1080/10904018.2004.10499061

Martin, M. M., & Rubin, R. B. (1994). Development of a communication flexibility measure. *Southern Communication Journal, 59*, 171–178. doi:10.1080/10417949409372934

McCroskey, J. C. (1970). Measures of communication-bound anxiety. *Speech Monographs, 37*, 269–277. doi:10.1080/03637757009375677

Rubin, R. B. (1994). Communicative Adaptability Scale. In R. B. Rubin, P. Palmgreen, & H. E. Sypher (Eds.), *Communication research measures: A sourcebook* (pp. 120–124). New York: Routledge.

Spitzberg, B. H. (2003). Methods of interpersonal assessment. In J. O. Greene & B. R. Burleson (Eds.), *Handbook of communication and social interaction skills* (pp. 93–176). Mahwah, NJ: Lawrence Erlbaum.

Witkin, B. R., & Trochim, W. W. K. (1997). Toward a synthesis of listening constructs: A concept map analysis. *International Journal of Listening, 11*, 69–87. doi:10.1207/s1932586xijl1101_5

Zakahi, W. R. (1985). The relationship of assertiveness to communication competence and communication satisfaction: A dyadic assessment. *Communication Research Reports, 2*, 36–40.

Zakahi, W. R. (1986). The effects of loneliness on perceptions of interaction: A cross-lag panel study. *Communication Research Reports, 3*, 94–99.

Zakahi, W. R., & Duran, R. L. (1984). Attraction, communicative competence, and communication satisfaction, *Communication Research Reports, 1*, 54–57.

Zakahi, W. R., & Duran, R. L. (1985). Loneliness, communication competence, and communication apprehension: Extension and replication. *Communication Quarterly, 33*, 50–60. doi:10.1080/01463378509369578

Zelko, H. P. (1954). An outline of the role of listening in communication. *Journal of Communication, 4*, 71–75. doi:10.1111/j.1460-2466.1954.tb00231.x

Scale

Communicative Adaptability Scale (Duran, 1992)

Source: Duran (1992). Reproduced with permission of Taylor & Francis.

Instructions: The following are statements about communication behaviors. Answer each item as it relates to your general style of communication (the type of communicator you are most often) in social situations.

Please indicate the degree to which each statement applies to you by marking the appropriate number according to the scale below.

5 = always true of me
4 = often true of me
3 = sometimes true of me
2 = rarely true of me
1 = never true of me

<u>Social Composure</u>

1) I feel nervous in social situations.
2) In most social situations I feel tense and constrained.
3) When talking, my posture seems awkward and tense.
4) My voice sounds nervous when I talk with others.
5) I am relaxed when talking with others.

<u>Social Confirmation</u>

6) I try to make the other person feel good.
7) I try to make the other person feel important.
8) I try to be warm when communicating with another.
9) While I'm talking I think about how the other person feels.
10) I am verbally and nonverbally supportive of other people.

<u>Social Experience</u>

11) I like to be active in different social groups.
12) I enjoy socializing with various groups of people.
13) I enjoy meeting new people.
14) I find it easy to get along with new people.
15) I do not "mix" well at social functions.

<u>Appropriate Disclosure</u>

16) I am aware of how intimate my disclosures are.
17) I am aware of how intimate the disclosures of others are.
18) I disclose at the same level that others disclose to me.
19) I know how appropriate my self-disclosures are.
20) When I self-disclose, I know what I am revealing.

<u>Articulation</u>

21) When speaking I have problems with grammar.
22) At times, I don't use appropriate verb tense.
23) I sometimes use one word when I mean to use another.
24) I sometimes use words incorrectly.
25) I have difficulty pronouncing some words.

<u>Wit</u>

26) When I am anxious, I often make jokes.
27) I often make jokes in tense situations.
28) When I embarrass myself, I often make a joke about it.
29) When someone makes a negative comment about me, I respond with a witty comeback.
30) People think I am witty.

Note: Items 1, 2, 3, 4, 15, 21, 22, 23, 24, and 25 are reverse-coded. Labels should be removed and items ordered randomly before use.

Profile 10

Communication Competency Assessment Instrument (CCAI)

(Rubin, 1982a, 1982b)

Profiled by: Phil Backlund, PhD

Central Washington University

Construct

The Communication Competency Assessment Instrument (CCAI) was designed for the college environment as a comprehensive assessment of an individual's communication competence. The instrument assesses speaking, interpersonal, and listening skills.

Instrument Type

Behavioral Assessment

Description

The CCAI was developed to provide assessments of the appropriateness of communication behaviors. The goal of the instrument is to identify students who may have difficulties with both sending and receiving communication in an educational setting (Rubin, 1982a). The CCAI is based on the premise that impressions of communication competence are centered on observer evaluations and impressions of actual behaviors enacted by a communicator. The instrument does not assess motivation (or other affective aspects of learning) or knowledge about basic communication principles. Part of the instrument is paper and pencil, and part is an oral interview based on viewing a 6-and-a-half-minute videotape of a representation of a class lecture.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

The CCAI is administered in three main sections. The first section asks the student to present a 3-minute extemporaneous persuasive talk on a topic of interest. Six judgments about the student's speaking ability are made: pronunciation, facial expression/tone of voice, speech clarity, informative/persuasive distinction, clarity of ideas, and ability to express and defend a point of view. An additional question assesses the student's ability to recognize a lack of understanding in the audience.

Next, the student views a videotaped, 6-and-a-half-minute representation of a class lecture. In the open-ended version of the CCAI, four questions are asked about the lecture. The questions assess competencies associated with listening: the ability to tell the difference between fact and opinion, the ability to understand suggestions, the ability to identify work needed to complete as assignment, and the ability to summarize.

Finally, the student is asked to respond in various ways to statements about experiences within the educational environment. These items allow for a self-assessment of ritual performance, asking questions, answering questions, expressing feelings, using an organizational pattern, giving directions, describing someone else's opinion, and describing differences of opinion. For example, three of the nine competencies assessed in this section include: "summarize oral instruction given by an instructor," "introduce yourself at the beginning of class," and "obtain information about requirements for your major." All responses are either oral or nonverbal in nature. The CCAI takes approximately 30 minutes per student to administer (Rubin, 1982a).

With respect to listening, the CCAI requires students to view a videotaped representation of a first day in a listening class. The instructor explains course requirements, factors that affect listening including suggestions for improvement, and the first class assignment. The CCAI has two versions for the listening portion—oral open-ended (see above) and a multiple-choice version. In the oral version, students are orally asked the four questions about the listening videotape, and they respond orally to these questions. The influence of reading ability was considered in the creation of the multiple-choice version of the CCAI. It is completed only after students have listened to the videotape. Students are allowed as much time as they need to answer the questions.

Scoring

The rater evaluates each of the 19 competencies using 5-point Likert scaling. Scores are summed and range from 19 to 95. Norms were not developed, so a baseline level of "competence" was not established. Results are used for comparison and research purposes, not for a definitive evaluation of competence.

Development

The CCAI was first pilot-tested at the University of Wisconsin–Parkside during August 1979. Communication faculty evaluated face validity. The instrument was further refined through more administrations of the instrument, and a third version was expanded and refined. The third version was analyzed by the communication faculty to determine readiness for reliability and validity assessment, and questions that did not meet the interrater reliability level of .80 were dropped from the instrument. The CCAI also was pilot-tested with students who helped identify confusing or ambiguous questions. The instrument was further refined based on this feedback.

At least three variations of the CCAI were subsequently developed. The original version of the listening section of the instrument included open-ended questions; the next version included multiple-choice questions in place of the open-ended ones (Rubin, 1982). In addition, a scaled-down college version was developed for high school students (CCAI-HS) (Rubin, 1995). The CCAI-HS contains 15 skill assessments in Communication Codes (CC), Oral Message Evaluation (OM), Basic Speech Communication Skills (BS), and Human Relations (HM) (Rubin, Welch, & Buerkel, 1995). The CCAI-HS does not include the listening component, a decision made to reduce testing time.

Reliability

In the initial development of the instrument, the coefficient alpha for the CCAI, based on 50 first-year scores, was .86. In addition, interrater reliability was established for a group of nine faculty members, .92, from various fields after a 4-hour training session (Rubin, 1982b), and between two raters, .97, after over a month of assessing students (Rubin, 1982a). In a separate study, the CCAI demonstrated a coefficient alpha of .78, and interrater reliability scores have ranged from .92 to .97 (Rubin, 1985).

Validity

Rubin detailed the development of the instrument and reported initial face and content validity data (Rubin, 1981, 1982a). The CCAI was developed to assess the communication abilities of college students, and a range of communication situations were described that could exist within the college setting that related to learning. Another faculty expert panel confirmed that these competencies sampled the domain of communication in educational contexts, thus providing evident in favor of content validity (Rubin, 1981). The initial information on face and content validity (together with interrater reliability) formed the basis for further validation studies. Concurrent validity was examined, and low (but statistically significant) correlation coefficients were found between the CCAI and past speaking experience, $r = .31$; grade point average, $r = .28$; number of credits completed, $r = .35$; and number of communication courses completed, $r = .28$. Also, the listening portion of the CCAI correlated, $r = .69$, with a separate (but unnamed) listening test (Rubin, 1982a). Investigations concerned with predictive validity suggest that the CCAI is a useful tool for predicting student-teacher success (McCaleb, 1983; Rubin & Feezel, 1984). Convergent validity was investigated through comparisons with instructor ratings of the same students. Instructor impressions were internally consistent ($\alpha = .90$) and correlated with the CCAI, $r = .65$, as well as instructors' perceptions of the students' public speaking (Rubin, 1982a).

Availability

For information on the availability of the CCAI, researchers are directed to its developer, Dr. Rebecca Rubin.

Sample Studies

An early issue in the conception of communication competence was the source of the judgment of competence. Who is in a better position to judge a person's competence—the person themselves or an observer (Wiemann & Backlund, 1980)? Rubin (1985) developed the CCAI as an observer-centered evaluation of competence, and her 1985 study was one of the first to show that self-report and other-reports of competence did not correlate highly. The CCAI correlated strongly with holistic impressions of competence and with student grades.

The CCAI was one of the few early assessment tools that included a comprehensive listening test as part of its conception of competence. Rubin and Roberts (1987) compared the conceptual and methodological similarities and differences of three listening measures—the Watson-Barker Listening Test (WBLT; see Profile 64), Kentucky Comprehensive Listening Test (KCLT), and the CCAI. The results provided information on the concepts being assessed in each and brought to light important methodological issues for listening test users. Similarities were found between the KCLT and WBLT. The CCAI forms were related more strongly to the WBLT than to the Kentucky test. The strongest relationship among the listening tests existed between the WBLT and the CCAI-OE (Open-Ended). The CCM-OE form appears to be superior to the CCAI-MC (Multiple Choice) form. Even though researchers and teachers seek a quick and easy format for rating listening ability, the open-ended responses seem superior (in the form of concurrent validity) to the multiple-choice format. This finding seems to be supported also by work showing that the WBLT does not factor appropriately (see Profile 64).

Rubin and colleagues (Rubin & Graham, 1988; Rubin, Graham, & Mignerey, 1990) embarked on a series of studies to determine the relation of communication competence to college success. Results indicated that communication competence is linked to success in college, that high school communication experience is related to higher GPAs and higher communication competence ratings, and that communication apprehension is related to perceptions of communication competence. Implications of these results are discussed in comparison to current theories of communication competence. Results of the second study suggested that communication competence decreased significantly during the sophomore year but then increased in the junior and senior years; communication apprehension and interaction involvement (see Profile 25) scores remained steady. The CCAI (including the listening portion) was shown to be a partial predictor of success in college.

Finally, Rubin and colleagues focused on high school speech instruction and the validity and reliability information on the CCAI–HS (Rubin *et al.*, 1995). The study examined the role of standardized, performance-based assessment measures of communication competence in the high school context, the applicability of using the CCAI-HS to assess student speaking performance, and the level of improvement as a result of instruction. The CCAI-HS did not include the listening portion of the assessment instrument.

Critique

The CCAI was one of the few measures that evaluated performed communication behavior rather than self-reporting of student abilities. A primary drawback is the amount of time necessary to use the measure with any sizeable number of students.

The main criticism of the oral, open-ended recall approach is that it was not practical for large-scale testing situations and, therefore, not cost-effective. This fact limited its applicability and likely its adoption as a research measure. One final drawback is the fact that that the public speaking portion is not given to an audience.

References

McCaleb, J. L. (1983). *The predictive validity of two measures of teachers' communication ability*. Unpublished manuscript, Department of Communication Arts & Theatre, University of Maryland.

Rubin, R. B. (1981, November). The development and refinement of the Communication Competency Assessment Instrument. Paper presented at the annual meeting of the Speech Communication Association, Anaheim, CA.

Rubin, R. B. (1982a) Assessing speaking and listening competence at the college level: The communication competency assessment instrument, *Communication Education*, *31*, 19–32. doi:10.1080/03634528209384656

Rubin, R. B. (1982b). *Communication Competency Assessment Instrument*. Annandale, VA: Speech Communication Association.

Rubin, R. B. (1985). The validity of the Communication Competency Assessment Instrument. *Communication Monographs*, *52*, 173–185. doi:10.1080/03637758509376103

Rubin, R. B., & Feezel, J. (1984, November). Elements of teacher communication competence: An examination of skills, knowledge and motivation to communicate. Paper presented at the annual meeting of the Speech Communication Association, Chicago.

Rubin, R. B., & Graham, E. E. (1988). Communication correlates of college success: An exploratory investigation, *Communication Education*, *37*, 14–27. doi:10.1080/03634528809378700

Rubin, R. B., Graham, E. E., & Mignerey, J. T. (1990). A longitudinal study of college students' communication competence, *Communication Education*, *39*, 1–14. doi:10.1080/03634529009378783

Rubin, R. B., & Roberts, C. V. (1987). A comparative examination and analysis of three listening tests, *Communication Education*, *36*, 142–153. doi:10.1080/03634528709378655

Rubin, R. B., Welch, S. A., & Buerkel, R. (1995). Performance-based assessment of high school speech instruction, *Communication Education*, *44*, 30–39. doi:10.1080/03634529509378995

Wiemann, M., & Backlund, P. M. (1980). Current theory and research in communicative competence. *Review of Educational Research*, *50*, 185–199.

Profile 11

Communication Functions Questionnaire (CFQ-30)

(Samter & Burleson, 1990)

Profiled by: Graham D. Bodie, PhD

Meek School of Journalism and New Media, The University of Mississippi

Construct

The Communication Functions Questionnaire (CFQ) was designed to measure the importance people place on specific communication skills in specified relationships.

Instrument Type

Self-Report

Description

The CFQ is a measure of the importance people place on a set of 10 communication skills for the functioning of close, personal relationships. The skills are organized under two broad classifications. Affectively oriented skills are those relevant to the management of emotions and include *comforting* skills (assisting others perceived as needing aid), *conflict management* skills (effective problem solving), *ego support* skills (boosting feeling of self-worth), *regulatory* skills (assistance in recognizing and remedying mistakes), *expressiveness* skills (ability to express emotions appropriately), and *listening* skills (ability to be attentive to others). Instrumentally oriented skills are those relevant to the management of behavior and include *referential* skills (ability to provide information in clear and concise manners), *conversation* skills (ability to start and maintain a conversation), *narrative* skills (ability to tell stories in entertaining ways), and *persuasion* skills (ability to influence others and gain compliance). Each skill is measured with three items, resulting in a 30-item scale.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

The CFQ is a self-administered questionnaire that takes approximately 5 minutes to complete. Participants complete the CFQ with respect to a particular relationship (e.g., best friend, cross-sex friend, work acquaintance, romantic partner, or spouse). The relationship of interest can be inserted into the instructions as needed. Participants read descriptions of communication behaviors typically performed in the relationship of interest and are asked to indicate how important that behavior is for people of that relationship type to display. Each described behavior is rated from *somewhat important* (1) to *extremely important* (5).

Scoring

There are three items for each of 10 skills. To score, compute the mean of the three items tapping the relevant skill.

Development

The CFQ was first introduced by Samter and Burleson (1990) in a study exploring the relationship between communication values and peer acceptance. Their original version of the CFQ was composed of 31 items organized by eight skills. Burleson and Samter (1990) used an 8-factor, 40-item version of the scale in their study of the relation between communication values and cognitive complexity, measured by the Role Category Questionnaire (see Profile 56). Although Burleson, Kunkel, Samter, and Werking (1996) used the same 40 items (5 items per factor), only 30 were retained after an exploratory factor analysis. Although the CFQ has undergone refinement over the last decade, all versions of the scale have included four affectively oriented communication skills (comforting, conflict management, ego support, and regulation) and four instrumental skills (conversing, informing, persuading, and narrative skill). The most recent version of the CFQ includes two additional affectively oriented skills, expressiveness and listening (see Jones, 2005); all 10 skills are assessed with three items each, resulting in a 30-item measure.

Reliability

Internal consistencies for each of the primary subscales have been found acceptable, generally exceeding .75 and often exceeding .80, as reported in the studies cited in this profile.

Validity

Factor analyses of the CFQ items indicate that they load as intended (Burleson & Samter, 1990; Samter & Burleson, 1990). In addition, confirmatory factor analyses of the CFQ subscales have regularly found that they load on two secondary factors, appropriately characterized as Affectively Oriented Skills and Instrumental Skills (e.g., Burleson, Kunkel, & Birch, 1994; Burleson, Kunkel, Samter, & Werking, 1996; Burleson & Samter, 1990; Burleson, Samter, & Lucchetti, 1992).

Evaluations of communication skills have been found associated with several characteristics of individuals, including cognitive complexity (Burleson & Samter, 1990), "love styles" or love attitudes (Kunkel & Burleson, 2003), expressivity-instrumentality (MacGeorge, Feng, & Butler, 2003), and individualism-collectivism (Mortenson, 2002). In sum, evaluations of partner communication skills obtained with the CFQ exhibit theoretically appropriate sensitivity to variations in (a) skill types (some skills are valued more than others), (b) relationship types (skills are valued in some relationships more than others), (c) skills within particular relationships (certain skills are more valued in some relationships than in others), and (d) several individual differences.

Availability

The Very Close Friend's version of the CFQ is provided at the end of this profile. Other versions can be created by changing the wording of the relationship of interest (e.g., same-sex friend, romantic partner, or parent).

Sample Studies

Communication values (i.e., evaluations of the importance of these communication skills) have been found associated with several relationship characteristics and outcomes. For example, Samter and Burleson (1990) discovered that college students who placed relatively high value on affectively oriented skills such as comforting and ego support were better liked and more accepted by their housemates than students who evaluated these skills less highly. Similarly, Samter (1992) found that persons who highly valued affectively oriented communication skills reported lower levels of loneliness than those viewing these skills as less important. Moreover, friends tend to be more similar than nonfriends in their evaluations of communication skills (Burleson *et al.*, 1992). Significantly, similarity in skill evaluations has been shown to lead to higher levels of relationship satisfaction for both friends (Burleson *et al.*, 1992) and romantic partners (Burleson *et al.*, 1994).

The communication skill evaluations tapped by the CFQ are influenced by several factors. In general, the importance accorded to a partner's communication skills increases linearly with relationship intimacy. For example, Westmyer and Myers (1996) found that the communication skills of partners were rated as more important in best friendships than in casual friendships or acquaintanceships, and this was especially true for the affectively oriented skills of conflict management, ego support, comforting, and regulation. Similar results were obtained for relationship closeness in a study of coworkers (Myers, Knox, Pawlowski, & Roog, 1999). Burleson *et al.* (1996) found that most of the communication skills tapped by the CFQ were considered more important in the context of romance than friendship. These results are consistent with the findings for friends and coworkers if it is assumed that romance is generally experienced as a more intimate relationship than friendship, an assumption directly supported by some research (e.g., Rubin, 1970).

The major source of variance in communication skill evaluations is type of skill. In virtually every study reported to date (e.g., Burleson *et al.*, 1996; Burleson & Samter, 1990; MacGeorge *et al.*, 2003), affectively oriented skills, especially ego support, comforting, and conflict management, have been rated as more important than the instrumental/interactional skills of conversing, informing, persuading, and narrative (i.e., storytelling). This finding is almost certainly a function of most research having focused on close relationships wherein concerns with emotion and relational issues assume prominence. Moreover, some research indicates that relationship type moderates the effect of skill type with respect to the evaluation of skill importance. For example, Burleson *et al.* (1996) found affectively oriented skills (ego support, comforting, and conflict management) were seen as substantially more important in romantic partners than friends, whereas the interactional skills (conversational and narrative skills) of romantic partners were seen as only somewhat more important than those of friends. These results appear consistent with the notions that romance is a more intimate relationship than friendship, and that affectively oriented communication skills are especially relevant in highly intimate relationships.

Critique

The CFQ has been used in a variety of populations with estimates of reliability consistently reaching acceptable levels. Validity evidence suggests the scale factors as intended and that communication values are related to conceptually similar phenomenon. Although the CFQ provides important information as to the value people place on communication skills in a range of relationships, samples tend to be overly represented by white, middle-class adults. The degree to which ethnic identity and other individual and cultural-level factors might influence the importance placed on skills or even on the generation of new skill-based factors should be investigated by future work. In addition, the most recent version of the CFQ has not been submitted to rigorous validity assessment. Finally, the classification of communication skills into 10 categories implies that they exist in tight, non-overlapping conceptual space. The degree to which participants do, indeed, make distinctions among these skills in more than a loose fashion is an empirical question that is only answered partially by past factor analytic work. More research should explore relevant differences and similarities in the classification presented by the CFQ and how everyday communicators classify skills.

References

Burleson, B. R., Kunkel, A. W., & Birch, J. D. (1994). Thoughts about talk in romantic relationships: Similarity makes for attraction (and happiness, too). *Communication Quarterly*, *42*, 259–273. doi:10.1080/01463379409369933

Burleson, B. R., Kunkel, A. W., Samter, W., & Werking, K. J. (1996). Men's and women's evaluations of communication skills in personal relationships: When sex differences make a difference—and when they don't. *Journal of Social and Personal Relationships*, *13*, 201–224. doi:10.1177/0265407596132003

Burleson, B. R., & Samter, W. (1990). Effects of cognitive complexity on the perceived importance of communication skills in friends. *Communication Research*, *17*, 165–182. doi:10.1177/009365090017002002

Burleson, B. R., Samter, W., & Lucchetti, A. E. (1992). Similarity in communication values as a predictor of friendship choices: Studies of friends and best friends. *Southern Communication Journal*, *57*, 260–276. doi:10.1080/10417949209372873
Jones, S. M. (2005). Attachment style differences and similarities in evaluations of affective communication skills and person-centered comforting messages. *Western Journal of Communication*, *69*, 233–249. doi:10.1080/10570310500202405
Kunkel, A., & Burleson, B. (2003). Relational implications of communication skill evaluations and love styles. *Southern Communication Journal*, *68*, 181–197.
MacGeorge, E. L., Feng, B., & Butler, G. L. (2003). Gender differences in the communication values of mature adults. *Communication Research Reports*, *20*, 191–199. doi:10.1080/08824090309388817
Mortenson, S. T. (2002). Sex, communication values, and cultural values: Individualism-collectivism as a mediator of sex differences in communication values in two cultures. *Communication Reports*, *15*, 57–70. doi:10.1080/08934210209367752
Myers, S. A., Knox, R. L., Pawlowski, D. R., & Roog, B. L. (1999). Perceived communication openness and functional communication skills among organizational peers. *Communication Reports*, *12*, 71–83. doi:10.1080/08934219909367712
Rubin, Z. (1970). Measurement of romantic love. *Journal of Personality and Social Psychology*, *16*, 265–273. doi:10.1037/h0029841
Samter, W. (1992). Communicative characteristics of the lonely person's friendship circle. *Communication Research*, *19*, 212–239. doi:10.1177/009365092019002005
Samter, W., & Burleson, B. R. (1990). Evaluations of communication skills as predictors of peer acceptance in a group living situation. *Communication Studies*, *41*, 311–326. doi:10.1080/10510979009368313
Westmyer, S. A., & Myers, S. A. (1996). Communication skills and social support messages across friendship levels. *Communication Research Reports*, *13*, 191–197. doi:10.1080/08824099609362086

Scale

Communication Functions Questionnaire (CFQ-30) Very Close Friend Version (Samter & Burleson, 1990)

<u>*Instructions*</u>: Below are descriptions of several different kinds of communication skills. Please read through the description carefully. Then, fill in the circle on your answer sheet for the response that best represents your feelings regarding ***how important it would be for a very close friend to possess the communication skill***. Make sure the numbers on the answer sheet and this questionnaire correspond. The items refer to how important you <u>*generally*</u> think these communication behaviors and outcomes are in very close friendships. For each item please use the following scale:

Somewhat Important	1	2	3	4	5	Extremely Important

For example, consider the following item:

> "Has the ability to make me believe I have the qualities people will like."

If you think this would be an important or extremely important skill for a very close friend to possess, you would choose "4" or "5." If you think this would be a moderately important skill for a very close friend to possess, you would choose “3.” Finally, if you think this would be only a somewhat important skill for a very close friend to possess, you would choose “1” or “2.” Please make certain that you read and rate each item.

Items Measuring Affective Skill Orientation

Comforting

1) Can help me work through my emotions when I’m feeling upset or depressed.
2) Comforts me when I am feeling sad or depressed.
3) Helps make me feel better when I’m hurt or depressed about something.

Conflict Management

4) Shows me it’s possible to resolve our disagreements in a way that won’t hurt or embarrass each other.
5) Makes me realize that it is better to deal with conflicts we have than to keep things bottled up inside.
6) Can work through our relational problems by addressing the issues rather than engaging in personal attacks.

Ego Support

7) Makes me feel like I’m a good person.
8) Encourages me to believe in myself.
9) Helps me feel proud of my accomplishments.

Regulative

10) Shows me that I have the ability to fix my own mistakes.
11) Encourages me to feel like I can learn from my mistakes by working through things with me.
12) Helps me see how I can improve myself by learning from my mistakes.

Expressiveness

13) Is open in expressing her/his thoughts and feelings to me.
14) Lets me know what’s going on in his/her world.
15) Shares his/her joys, as well as sorrows, with me.

Listening

16) Listens carefully when I am speaking.
17) Is an attentive listener when I need to talk to someone.
18) Gives me her/his full attention when I need to talk.

Items Measuring Instrumental Skill Orientation

Referential

19) Explains things clearly.
20) Makes me understand exactly what he/she is referring to.
21) Can express complicated ideas in a direct, clear way.

<u>Conversation</u>

22) Is a good conversationalist.
23) Is able to start up a conversation easily.
24) Can make conversation easy and fun.

<u>Narrative</u>

25) Can get me laughing because he/she is so good at telling a joke or story.
26) Is able to tell a story in a way that captures my attention.
27) Can make even everyday events seem funny or exciting when telling a story.

<u>Persuasion</u>

28) Makes me feel like I've made my own decision even though I do mostly what he/she wants.
29) Persuades me that doing things his/her way is the best.
30) Can convince me to do just about anything.

Note: Labels should be removed and items randomly ordered prior to administration. Instructions can be modified to apply to an online survey or a pencil-and-paper survey that does not use an answer sheet. Many of the above items were originally printed in Samter and Burleson (1990), Burleson and Samter (1990), and Burleson *et al.* (1996). The version of the CFQ reprinted here was obtained by the profile author from Brant Burleson prior to his passing in December 2011. The profile was constructed from Brant's notes, all of which are available upon request from gbodie@gmail.com.

Profile 12

Conversational Listening Span (CLS)

(Janusik, 2004, 2005, 2007)

Profiled by: Debra L. Worthington, PhD and Courtney Edwards, MA

Auburn University

Construct

The Conversational Listening Span (CLS) measures conversational listening capacity by estimating "the number of items that one can hold active, can paraphrase, and can respond to in the course of a conversation" (Janusik, 2007, p. 144). It qualitatively differs from the Listening Span Test (see Profile 35).

Instrument Type

Cognitive Assessment

Description

CLS is similar in construction to other traditional span instruments, the goal of which is to ascertain the amount of information an individual can hold in working memory. Grounded in working memory theory and simulating a relational, transactional model of communication, the CLS is designed to provide an observable, real-time behavioral indicator of short-term listening capacity. Participants interact with an interviewer who asks questions on a specified topic (e.g., politics, television, or sports). After a group of questions is asked by the interviewer, the participant is asked to provide a response, which includes paraphrasing the questions. A participant's conversational listening span is measured by assessing whether he or she incorporated question content into the response.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

The first step to collecting CLS data is to ask participants to rate their interest on several topics. Past work has used politics, television, and sports as topic areas. Participants can be randomly assigned to topics based on interest scores, or these scores can be used as covariates in subsequent analyses.

Participants are then seated facing an interviewer, who explains that the upcoming task will involve hearing a series of questions to which they should create paraphrased responses (see instructions below). To familiarize participants with the procedure, the dyad works through a practice session. After answering any remaining questions, the interviewer then starts the experimental task.

The experimental task begins by the interviewer introducing a new topic (e.g., “Let’s talk about TV”). The interviewer then asks a set of six questions in groups of two. After the first two questions about this topic, the participant is allowed to respond. Two more questions are asked with another pause for the response. A final two questions are asked to complete the first round of the experimental task. If the participant’s response included an accurate paraphrase for two out of three of the two-question groups, the interviewer introduces another topic. For this topic, questions are asked in groups of three. Three groups of three questions comprise the second round of the experimental task. This continues until the question group contains seven questions, assuming that a participant accurately paraphrases two of the three groups of questions in each round.

Researchers should also train interviewers to deliver the questions as consistently as possible (e.g., verbally disagree with something that was said, or lean back). Ideally, multiple interviewers will be used in any given study to assess any effects.

Scoring

Determining a participant’s CLS score is in line with that of other span tasks, with two exceptions. First, reflecting the natural flow of a conversation, participants are allowed to recall and paraphrase information in any order. Second, the “correctness” of an answer is only a function of whether the question content is paraphrased and responded to. If a participant does not know the answer to a question, saying “I don’t know” is allowed. For example, if asked, “What year did Scottie Pippen join the Chicago Bulls?”, a participant can respond, “I do not know what year Scottie Pippen joined the Bulls.” This response would be evaluated the same as “Scottie Pippen joined the Bulls in 1988.” Thus, individual span tallies are based on lenient rules (Daneman, 1991) and procedures described by Whitney, Ritchie, and Clark (1991).

Span designations range from 0.0 to 7.0. A designation of zero is assigned to participants who do not paraphrase the content of any presented questions in the first round of the experimental task (three two-question groups). A fuller explanation of evaluation procedures is presented in Figure P12.1.

As illustrated in Figure P12.1, participants receive a CLS score when they correctly paraphrase all questions within two of the three question groups at each level. If a participant does not reach this threshold (i.e., they score a zero or only correctly paraphrase one question group), the experimental task stops. Whole scores represent the last set for which a respondent accurately reproduced two of the three questions in the

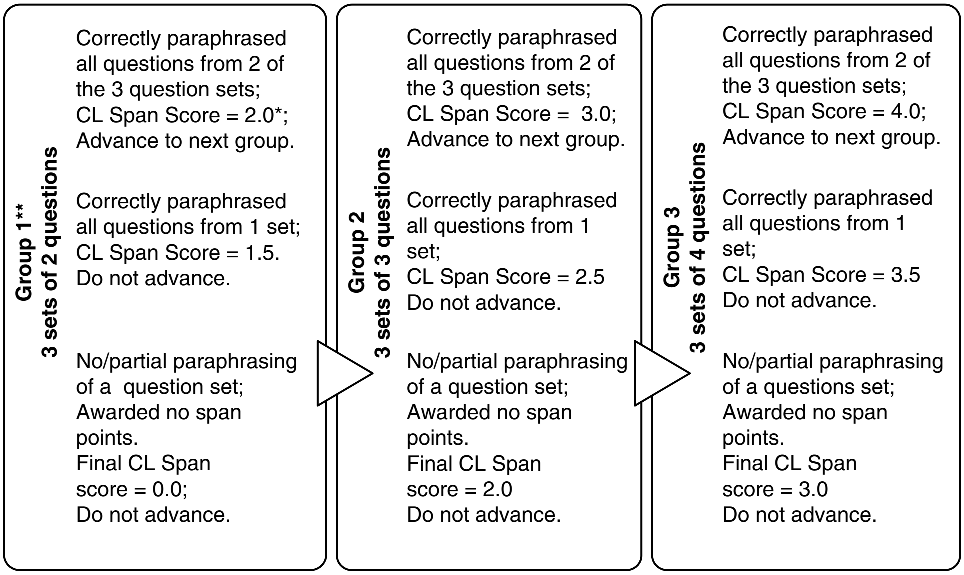


Figure P12.1 CL span scoring designations: experimental study with three group design.
*This number denotes an individual's conversational listening span, or the number of ideas the individual can hold active and respond to in the course of the conversation.
**Although possible, it is assumed that few studies would utilize an initial group consisting of three sets of one question.

group. Participants can earn an additional ½ "point" if they correctly complete one set of three questions from the next group. Thus, a score of 2.5 indicates the successful paraphrasing of the three sets of questions from Group 1 and the successful paraphrasing of one of the three sets of questions from Group 2 (see Figure P12.1). This process continues until the participants fail to meet the threshold to advance, or they reach the seventh group.

Development

Janusik's interest in working memory theory inspired the development of the CLS. Working memory capacity affects the ability to process, interpret, recall, and retain messages (Daneman & Carpenter, 1983). Working memory capacity varies with the individual (Just, Carpenter, & Keller, 1996). Span tests are the preferred means of assessing working memory capacity (Daneman & Merikle, 1996).

Janusik (2004, 2005, 2007) noted that listening and speaking span tasks are inappropriate measures of conversational listening for several reasons: Neither was designed to assess communication processes; both are linear in presentation, whereas conversations are transactional; and they do not account for prior knowledge and experience. The CLS addresses these important differences. Although the CLS shares commonalities with other span tasks, it differs in three significant ways: (a) CLS sentences and questions will be related; (b) participants recall ideas, not specific words; and (c) researchers act as conversational partners and are directed to respond to participants (Janusik, 2004) (see also review of the Listening Span Tests, Profile 35).

Reliability

Janusik (2007) reported score stability across several studies. However, the listed studies are not readily available for review (Janusik, 2006; Janusik & Zhang, 2003; Valikoski, Ilomaki, Maki, & Janusik, 2005). Because interviewers are responsible for scoring (i.e., they determine if a participant can move on to the next level), assessments of interrater reliability should be assessed and continually monitored (see Chapter 6).

Validity

Although the CLS reflects the underlying theory, methodology, and assessment methods of listening and speaking span measures, it differs in that it measures capacity in the specialized context of a communication transaction. Janusik (2007) investigated the criterion and construct validity of the CLS. In terms of construct validity, the CLS correlated with other, related span tasks (Listening, $r = .20$; Speaking, $r = .21$). Results for the speaking span task have been replicated, $r = .23$ (Janusik, 2009). Janusik noted that criterion-related validity was supported; the CLS scores were normally distributed using a sample of 360 undergraduate students.

Availability

A full description of the measure and sample stimulus question sets are provided in Janusik's (2004) dissertation and in multiple articles (e.g., Janusik, 2005, 2007). Examples of these instructions and descriptions are provided below, with permission. Researchers are allowed to use these question sets in studies with appropriate citation.

Sample Studies

At this time, most studies utilizing the CLS have focused on building and refining the measure or building its validity portfolio. These studies are outlined in the Validity section here. In one additional study, Janusik (2009) tested the divergent validity of the CLS, comparing the measure to subscale scores of the Watson-Barker Listening Test (WBLT) (Watson & Barker, 1988; Watson, Barker, Roberts, & Roberts, 2001). Findings showed the CLS was significantly correlated with all subsections of the WBLT, with correlations ranging from a low of .10 for the conversational meaning subscale to .25 for the overall scale score. However, researchers have raised significant questions about the validity of the WBLT (see Profile 64). Thus, these findings are suspect.

Critique

The CLS adapts the traditional span task to better fit the transactional nature of conversations, allowing for a greater range of responses as well as allowing participants, as part of a simulated conversation, to indicate when they do not know the answer to a question. This method allows the researcher to move from an emphasis on the listening product to assessing the listening process (Janusik, 2007). Of course, the CLS requires greater time to develop question sets, train data collectors, and code participant interactions. However, depending on the research goals, it may be well worth the effort. Such research also would provide verification of Janusik's findings and provide additional insight into the reliability of the CLS.

However, two caveats should be noted. First, in some reported studies the nature of the topic and question level are conflated. If an interviewer asks three sets of two questions about television, then moves to ask three sets of three questions about sports, any variability could be due to CLS or to the topic. One way to circumvent this concern is to represent all topics at all levels and randomly assign participants to a topic (or stratified randomly based on interest scores). Alternatively, question sets could be pulled randomly at each level for each participant.

Second, although the CLS attempts to capture the feel of a natural conversation, in reality individuals rarely paraphrase in the manner participants are required for this task. For instance, when we are asked, "What is your name?", we simply respond with our name rather than "My name is X." But, assumedly, the listener had to hold the question in short-term memory long enough to respond. As a result, the proposed operationalization may be of an ability to paraphrase or remember presented instructions more than an accurate depiction of short-term memory. If the task more accurately represents a person's ability to paraphrase, then one would expect rather low correlations between the CLS and other span tasks, which is what past work has found. Thus, more work is encouraged to show convergent validity for the CLS with other ways of measuring short-term memory capacity.

References

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research*, *20*, 445–464. doi:10.1007/BF01067637

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 561–584. doi:10.1037/0278-7393.9.4.561

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review*, *3*, 422–433. doi:10.3758/BF03214546

Janusik, L. A. (2004). Researching listening from the inside out: The relationship between conversational listening span and perceived communicative competence. UMI Proquest: Digital dissertations. Retrieved from http://drum.lib.umd.edu/handle/1903/1417

Janusik, L. A. (2005). Conversational Listening Span: A proposed measure of conversational listening. *International Journal of Listening*, *19*, 12–28. doi:10.1080/10904018.2005.10499070

Janusik, L. A. (2006). [The CLS and divergent validity]. Unpublished raw data.
Janusik, L. A. (2007). Building listening theory: The validation of the Conversational Listening Span. *Communication Studies*, *58*, 139–156. doi 10.1080/10510970701341089.
Janusik, L. A. (2009, November). The Conversational Listening Span Task and the WBLT: Convergence and divergence of listening comprehension. Paper presented at the meeting of the National Communication Association, Chicago, IL.
Janusik, L. A., & Zhang, X. (2003). [Second language processing: Differences in working memory capacity of Chinese men in English and Chinese]. Unpublished raw data.
Just, M. A., Carpenter, P. A., & Keller, T. A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review*, *103*, 773–780. doi:10.1037/0033-295X.103.4.773
Valikoski, T. R., Ilomaki, I., Maki, E., & Janusik, L. A. (2005). Conversational listening span: A comparative study of American and Finnish students. *Prologos*, *1*, 88–108.
Watson, K. W., & Barker, L. L. (1988). Listening assessment: The Watson-Barker listening test. *International Journal of Listening*, *2*, 20–31. doi:10.1080/10904018.1988.10499095
Watson, K. W., Barker, L. L., Roberts, C. V., & Roberts, J. D. (2001). *Watson-Barker listening test: Video version: Facilitator's guide*. Sautee, GA: SPECTRA.
Whitney, P., Ritchie, B. G., & Clark, M. B. (1991). Working-memory capacity and the use of elaborative inferences in text comprehension. *Discourse Processes*, *14*, 133–145. doi:10.1080/01638539109544779

Scale

Measure: Sample CLS Script (Janusik, 2007)

Source: Jaunsik (2005). Reproduced with permission of Taylor & Francis.

Below is a sample script on the general topic of Television. Subtopics are identified, as are verbal and nonverbal instructions, to data collectors (DC).

The instructions are given as follows:

Interviewer: "This part of the study is called the Conversational Listening Span. You and I are going to have a conversation where I ask you questions and you reply. I'll ask you anywhere from two to five questions in a row. When I'm finished, I'd like you to answer each question by first paraphrasing the question and then answering it. You may answer them in any order you wish. What you may not do, though, is paraphrase all of them first and then answer them. Your answers should be how you really think or feel, and if you cannot answer a question, simply paraphrase and say something like, 'I don't know the answer.' In order to get full credit, you must both paraphrase and answer each question. I may not repeat any of the questions nor confirm any questions that you have once we begin for real, so it's very important that you listen carefully. To make certain that you understand what is being asked of you, I'd like you to paraphrase your understanding now, and I'll correct any misperceptions you might have."

(Listen to them paraphrase, and correct any misunderstandings they have).

"It sounds to me like you understand, so let's begin with a practice session."

TELEVISION (General Topic Area)

(Practice) Let's start off talking about *Friends*.
Explain who your favorite character on *Friends* is and why.
What is Phoebe's twin sisters' name? (If the participant answered "I don't know Phoebe's twin sister's name" or "The name of Phoebe's twin sister was Ursula, I think," then the response is a valid indicator of the participant's ability to understand and answer the gist of the question.)

The DC would say "good" or remind them what would be correct.

(Practice continued) I'd like to ask you some questions about TV Violence.
If it were up to you, explain why you would have more, less, or the same amount of violence on TV.
Explain why you would say that the fights on *Jerry* are staged or real.
What is one of the most violent TV shows airing today?

DC: "We've just finished the practice session. Do you have any further questions before we begin for real? I'd like to remind you to do the best that you can. However, I will not be able to answer any more questions after this."

2 Let's talk about watching TV.
DC: "I know I watch more than I sometimes care to admit!"
How many hours of television would you say that you watch daily?
What's the name of the last educational program that you watched?

2 Now I'd like to ask you about TV shows that you like. *DC: Lean forward.*
Tell me what your favorite television show of all time is and why.
Detail which family television show your personal family most resembles.

DC: Brief personal feedback

2 Let's talk about Reality TV shows. *DC: Move back from forward lean.*
How real do you believe reality shows are?
A recent article in the *Washington Post* said that the reality shows are pushing people to consume a lot of liquor so they'll act out. How would you respond to that statement?

<u>Scoring/Span Designation Note</u>: If the participant fully paraphrased and responded to two of the three groups of questions above, then she has earned the right to move to the third level. This pattern is repeated for up to seven levels (or fewer if experimental goals are met). See Figure P12.1 for more scoring details.

DC: "Now we'll move onto the next level. You'll hear three questions at a time, and you'll do just what you did, paraphrase and answer as many of them as you'd like, in any order."

3 I'd like to ask some questions about the News. *DC: Elbows on table.*
Who is your favorite weatherman?
Explain for what season you believe weather reports are more accurate.
If you could only get your news from one news station for the rest of your life, which station would it be and why?
DC: Brief personal response & remove elbows.

3 Let's switch to the *Sopranos.* *DC: Laugh.*
Where is the *Sopranos* filmed?
Explain who your favorite character is on the *Sopranos* and why?
Who did Carmela secretly love on the *Sopranos*?

3 Now let's talk about *The Simpsons.*
Who are the five main characters of *The Simpsons* that make up the Simpson family?
Where are the Simpsons from?
How would you account for the popularity of *The Simpsons*?

Scoring/Span Designation: If participants achieve two out of three sets at the second level, and only one out of three sets at the third level, then their score was 2.5. If they achieved two out of three sets at the third level, then they earn the right to go onto the fourth level. See Figure P12.1 for an illustrated example of CLS scoring.

Profile 13

Conversational Sensitivity Scale (CSS)

(Daly, Vangelisti, & Daughton, 1987)

Profiled by: Debra L. Worthington, PhD

Auburn University

Construct

Conversational Sensitivity (CS) refers to "the propensity of people to attend to and interpret what occurs during conversation" (Daly, Vangelisti, & Daughton, 1987, p. 169).

Instrument Type

Self-Report

Description

Daly *et al.* (1987) developed the 36-item Conversational Sensitivity Scale (CSS) to assess an individual's level of attention to and understanding of underlying meanings during conversations. It assesses eight dimensions believed to be associated with a person's skill at evaluating and responding to common challenges in conversational interactions:

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

CS Construct	Conceptual definition
Detecting meaning	Ability to identify underlying and/or multiple meanings based on what others say
Conversational memory	Ability to remember the content of a conversation
Conversational alternatives	Level of conversational flexibility when selecting from possible words and phrasings
Conversational imagination	Tendency to imagine conversations
Conversational enjoyment	Level of enjoyment in participating in/listening to conversations
Interpretation	Ability to both paraphrase and identify nuances in conversations (i.e., underlying meaning, sarcasm, irony, etc.)
Perceiving affinity	Ability to assess the level of liking, attraction, or affiliations between conversational members
Detecting power	Ability to identify power relationships between conversational members

Administration

Using 5-point Likert scaling, participants are typically able to respond to the 36-item measure in less than 15 minutes. The factor structure allows researchers to utilize only the portions of the measure salient to the research question under study.

Scoring

After reverse coding relevant items, subscale scores are obtained by computing the average of item responses, producing seven scores that each range between 1 and 5. To obtain a total CS score, all items are summed, producing a score that ranges between 36 and 180. In both cases, higher scores indicate greater sensitivity. No known normative data have been reported in the published literature.

Development

Drawing on the theoretical framework of social affordance (Gibson, 1966, 1979), Daly *et al.* (1987) explored the concept of conversational sensitivity as a means of identifying the elements distinguishing individuals who are attentive and responsive in conversations from those who are not. The initial scale was developed and refined through a series of six studies. In the initial study, 150 items were generated by graduate students based on a conceptual definition of conversational sensitivity. Their first study reduced the initial items from 150 to 73 by removing duplicate items and those that did not clearly fit the theoretical description of CS. The resulting 73 items were presented to

149 undergraduate students. The authors noted that the responses were factor analyzed; items were added, deleted, and refined, eventually leading to a revised 58-item scale. This revised scale was administered to 443 undergraduate students. Their responses were submitted to a principal component analysis (orthogonal rotation), which resulted in seven components (see description above) and 36 total items. In addition to further refinement of the retained items, Daly *et al.* added an eighth dimension addressing sensitivity to power relationships in social interactions. Notably, this dimension and other new items were not submitted to further item analysis. The follow-up studies by Daly *et al.* focused on building a validity portfolio for the 36-item measure (see below).

Stacks and Murphy (1993) addressed the content validity of final 36-item CSS. Responses from 263 undergraduate students were submitted to principal component analysis with varimax rotation. The between item correlation matrix suggested a 27-item scale and five components, which accounted for 52.4% of the item variance (MSA = .85, Bartlett test of sphericity = 2470.47, $p < .0001$). However, other studies have confirmed the original eight-factor solution suggested by Daly *et al.* (1987; e.g., Bodie, 2011; Honeycutt, Zagacki, & Edwards, 1992–1993).

Reliability

Results of multiple studies generally support the reliability of the overall scale and its subscales (Bodie, 2011; Chesebro, 1999; Daly *et al.*, 1987; Stacks & Murphy, 1993). Chesebro and Martin (2003), Honeycutt *et al.* (1992–1993), and Hosman (1991) reported overall reliability estimates ranging between .80 and .90. Bodie (2011) reported subscale reliability estimates ranging between .56 and .86, with an average Cronbach's alpha of the eight CSS subscales of .74 and an estimate of .89 for the total scale. Salisbury and Chen (2007) reported estimates of less than .40 for some dimensions, including perceiving affinity and conversational memory.

Validity

In terms of construct validity, Honeycutt *et al.* (1992–1993) reported factor loadings from a confirmatory factor analysis but did not report model fit statistics. Bodie (2011) reported results from a principal axis analysis with varimax rotation that produced the expected eight-factor solution (60.25% of item variance). As noted, one study (Stacks & Murphy, 1993) found a different factor structure than that proposed by the scale developers.

Daly *et al.* (1987) explored convergent validity of the CSS. They found higher CS positively associated with a variety of personality constructs, including empathy, self-esteem, self-monitoring, and private self-consciousness, and Stacks and Martin (1993) found it related to cognitive complexity (see RCQ profile, Profile 56). Chesebro and Martin (2003) found a positive association between conversational sensitivity and cognitive flexibility, $r = .50$; however, the hypothesized inverse relationship between CS and verbal aggressiveness and indirect interpersonal aggressiveness was not supported ($rs = .01$ and $.06$, respectively).

CS has been positively related to verbally praising others (Wigley, Pohl, & Watt, 1989, $r = .44$) and dimensions of interpersonal communication motives: communicating for pleasure – affection ($r = .22$) and relaxation ($r = .28$; Hosman, 1991). Daly *et al.* (1987) examined the relationship between overall CS and Interaction Involvement (II) (see

Profile 25). Overall, CS was positively related to II ($r = .28$), with a stronger relationship reported for the perceptiveness subscale ($r = .55$).

Bodie (2011) utilized the CSS as part of a validation study of the Active-Empathic Listening Scale (AELS; see Profile 2). His findings provide further support for convergent validity; the overall CS and seven of its subscales (the exception being CSS-imagination) were positively associated with the overall AELS and its three subscales: Sensing, Processing, and Responding ($.11 < r < .45$).

Predictive validity was tested by Daly and colleagues (1987). They reported that more sensitive individuals had better recall of and inferred more from their conversations with others than their less sensitive counterparts. They also found that individual CS can be heightened in certain situations, such as when the conversation involves a personal issue, is with someone held in high regard, or is novel in some way.

Availability

The CS scale, along with factor loadings for the initial items, is presented in Daly *et al.*'s (1987) article in *Human Communication Research*. The measure is also presented at the end of this profile and is free to use for research purposes with appropriate citation.

Sample Studies

The relationship between CS, as a general global measure, and individual listening style has been explored. Chesebro (1999) reported a moderate relationship between People Listening ($r = .43$) and a weak association with Content Listening ($r = .18$) of the Listening Styles Profile (LSP-16; Watson, Barker, & Weaver, 1995). Salisbury and Chen (2007) extended this research, examining the relationship between the dimensions of CS and the four listening styles of the LSP-16. Among their findings, they found positive associations between Conversational Alternatives and Conversational Enjoyment, Detecting Meaning, and People Listening ($rs = .14$, .25, and .21, respectively). They also reported positive associations between Conversational Alternatives and Detecting Meaning, Conversational Imagination, and Action Listening ($r = .19$, .22, and .14, respectively). However, a caveat to these findings should be noted; the validity and reliability of the LSP-16 have been questioned (see LSP-R, Profile 36).

Using path analysis, Honeycutt *et al.* (1992–1993) studied the impact of II on conversational sensitivity and its effect on perceived communication competence. They found that elements of imagined interactions (specificity, retroactive, and variety) enhance conversational sensitivity, which then affects perceptions of communication competence.

Finally, Merolla (2006) reported a positive relationship between CS (as a decoding ability) and humor orientation, $r = .47$.

Critique

Introduced almost 30 years ago, the CSS remains an underutilized scale, despite calls for additional research (e.g., Chesebro & Martin, 2003; Stacks & Murphy, 1993). Daly *et al.* (1987), in their introductory article, noted that their original goal was to develop a measure of the general construct, not a multifactor instrument. Although several dimensions were identified, Daly and colleagues acknowledged that other untapped dimensions likely exist. They saw their instrument as the first step in scale development. Their suggestion that researchers further refine the construct by more fully investigating the dimensions they initially identified has not been realized. Few studies have done so, with most treating CS as a global construct. Such examination is warranted. Further review of the dimensions finds that the wording of some items may affect participant responses. For example, all items from the interpretation facet are worded in the reverse, meaning that responses to these items must all be reverse-coded. With regard to listening, several studies suggest CS is related to listening behavior. Intuitively, this makes sense, particularly the dimensions of Detecting Meaning and Interpretation. Finally, like many communication measures, cultural differences in CS have not been fully accessed.

References

Bodie, G. D. (2011). The Active-Empathic Listening Scale (AELS): Conceptualization and evidence of validity with the interpersonal domain. *Communication Quarterly, 59,* 277–295. doi:10.1080/01463373.2011.583495

Chesebro, J. L. (1999). The relationship between listening styles and conversational sensitivity. *Communication Research Reports, 16,* 233–238. doi:10.1080/08824099909388720

Chesebro, J. L., & Martin, M. M. (2003). The relationship between conversational sensitivity, cognitive flexibility, verbal aggressiveness and indirect interpersonal aggressiveness. *Communication Research Reports, 20,* 143–150. doi:0.1080/08824090309388810.

Daly, J. A., Vangelisti, A. L., & Daughton, S. M. (1987). The nature and correlates of conversational sensitivity. *Human Communication Research, 14*(2), 167–202. doi:10.1111/j.1468-2958.1987.tb00126.x

Gibson, E. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Honeycutt, J. M., Zagacki, K. S., & Edwards, R. (1992–1993). Imagined interaction, conversational sensitivity and communication competence. *Imagination, Cognition, and Personality, 12,* 139–157. doi:10.2190/B9PC-51RJ-1D7N-4 M94

Hosman, L. A. (1991). The relationships among need for privacy, loneliness, conversational sensitivity, and interpersonal communication motives. *Communication Reports, 4,* 73–80. doi:10.1080/08934219109367527

Merolla, A. J. (2006). Decoding ability and humor production. *Communication Quarterly, 54,* 175–189. doi:10.1080/01463370600650886

Salisbury, J. R., & Chen, G.-M. (2007). An examination of the relationship between conversational sensitivity and listening styles. *Intercultural Communication Studies, 16,* 251–262. Retrieved from http://web.uri.edu/iaics/files/24-Jennife-Salisbury-Guo-Ming-Chen.pdf

Stacks, D. W., & Murphy, M. A. (1993). Conversational sensitivity: Further validation and extension. *Communication Reports, 6,* 18–25. doi:10.1080/08934219309367557

Watson, K. W., Barker, L. L., & Weaver, J. B., III. (1995). The listening styles profile (LSP-16): Development and validation of an instrument to assess four listening styles. *International Journal of Listening*, *9*, 1–13. doi:10.1080/10904018.1995.10499138
Wigley, C. J., Pohl, G. H., & Watt, M. G. S. (1989). Conversational sensitivity as a correlate of trait verbal aggressiveness and the predisposition to verbally praise. *Communication Reports*, *2*, 92–95. doi:10.1080/08934218909367488

Scale

Conversational Sensitivity Scale (CSS) (Daly, Vangelisti, & Daughton, 1987)

Source: Daly (1987). Reproduced with permission of Taylor & Francis.

How much would you say each of the following statements reflects you and the way you communicate? Indicate your level of agreement using the following scale.

1	2	3	4	5
Strongly Disagree				Strongly Agree

Detecting Meaning

1) I often find myself detecting the purposes or goals of what people are saying in their conversations.
2) Many times, I pick up from conversations little bits of information that people don't mean to disclose.
3) I can often understand why someone said something even though others don't see that intent.
4) In conversations I seem to be able to often predict what another person is going to say even before he or she says it.
5) I often hear things in what people are saying that others don't seem to notice.
6) I often find hidden meanings in what people are saying during conversations.
7) I often notice double meanings in conversations.
8) I often have a sense that I can forecast where people are going in conversations.

Conversational Memory

9) I think I can remember conversations I participate in more than the average person.
10) I'm terrible at recalling conversations I had in the past. (R)
11) If you gave me a few moments I could probably easily recall a conversation I had a few days ago.
12) I have a good memory for conversations.
13) I can often remember specific words or phrases that were said in past conversations.

Conversational Alternatives

14) I have the ability to say the right thing at the right time.
15) If people ask me how to say something I can come up with a number of different ways of saying it.
16) I'm very good at coming up with neat ways of saying things in conversations.

17) I am good at wording the same thought in different ways.
18) In virtually any situation, I can think of tactful ways to say something.

Conversational Imagination

19) I like to think up imaginary conversations in my head.
20) I often make up conversations in my mind.
21) Compared to most people, I don't spend much time inventing "make-believe" conversation. (R)

Conversational Enjoyment

22) I would enjoy being a fly on the wall listening in on other people's conversations.
23) Conversations are fascinating to listen to.
24) I really enjoy overhearing conversations.
25) I'm less interested in listening in on others' conversations than most people. (R)

Interpretation

26) I'm usually the last person in a conversation to catch hidden meanings in puns and riddles. (R)
27) I often have difficulty paraphrasing what another person said in a conversation. (R)
28) I'm not very good at detecting irony or sarcasm in conversations. (R)

Perceiving Affinity

29) Often in conversations, I can tell whether the people involved in the conversation like or dislike one another.
30) I can tell in in conversations whether people are on good terms with one another.
31) I can often tell how long people have known each other just by listening to their conversation.
32) I'm not very good at figuring out who likes whom in social conversations. (R)

Detecting Power

33) I can often to tell when someone is trying to get the upper hand in a conversation.
34) I'm often able to figure out who's in charge in conversations.
35) Most of the time, I'm able to identify the dominant person in a conversation.
36) In group interactions, I'm not good at determining who the leader is in the conversation. (R)

Note: Labels should be removed along with indicators of reverse coding (R), and items should be randomly ordered prior to administration.

Profile 14

The Couples Helping Exercise

(Barker & Lemle, 1984, 1987)

Profiled By: Andrew C. High[1], Jenny L. Crowley[2], and Emily M. Buehler[1]

[1] *The University of Iowa*
[2] *University of Tennessee*

Construct

The Couples Helping Exercise (CHE) was designed to engage two people in a supportive or help-intended conversation about issues experienced by one or more of the participants.

Instrument Type

Interaction Analysis; Behavioral; Experimental Method

Description

Coping with personal stressors sometimes requires formal helping, or the assistance of professional counselors. More often, however, people gain comfort and support via informal helping by communicating with family, friends, or acquaintances. The CHE, originally developed by Barker and Lemle (1984, 1987), is a research method that embodies and instigates informal helping. It is a semistructured communication task that has been employed in numerous contexts. This method commonly falls under what Burleson and MacGeorge (2002) described as the interaction analysis paradigm or, less typically, the experimental paradigm.

Administration

The CHE involves two roles that are described using several terms, including discloser/listener, discloser/helper, and support provider/support receiver, depending on the focus

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

of study. The discloser reveals a personal problem to the listener, who, in turn, listens and attempts to help the discloser. After the interaction, one or both participants complete a questionnaire evaluating the interaction or their partner, and the conversations are often recorded so they can be coded for future analysis.

Whereas some researchers assign participants to the roles of discloser and listener (e.g., High & Solomon, 2014; Holmstrom, Burleson, & Jones, 2005; Trees, 2000), other studies let assignment happen naturally (e.g., Lawrence *et al.*, 2008; Pistrang, Picciotto, & Barker, 2001; Priem, Solomon, & Steuber, 2009). Participants sometimes switch roles after an initial exchange to provide perspective from both roles (Barker & Lemle, 1984; Lawrence *et al.*, 2008). Researchers typically time these interactions, and conversations range from 5 (Holmstrom *et al.*, 2005) to 30 minutes (Verhofstadt, Buysse, & Ickes, 2007). Most interactions, however, last 10 minutes or less (Barker & Lemle, 1984; Priem *et al.*, 2009), although participants can conclude their conversations at any time.

In some studies, participants are given no instructions or are asked to communicate naturally (Pistrang & Barker, 1998). For example, the listener is instructed to "try to be helpful in whatever way feels natural to you" (Pistrang *et al.*, 2001, p. 619). In other studies, researchers train listeners to provide helping behaviors, such as certain levels of verbal person-centered support (High & Solomon, 2014; Holmstrom *et al.*, 2005), nonverbal immediacy (Burgoon & Hale, 1988; Jones, 2004), or active listening (Bodie, Vickery, Cannava, & Jones, 2015). In the CHE, participants typically discuss personal problems that do not involve their study partner. The discussion of relational stressors involving an interaction partner represents conflict communication, increases the likelihood of disagreement, and is excluded from this procedure.

Scoring

The nature of this methodology precludes scoring like that typically done with self-report measures. Scoring is fully a function of either the rubric used to code relevant behaviors from video- or audiotaped recordings of the interaction or the self-report scales used to gauge participants' perceptions of the interactions.

Development

Barker and Lemle based the CHE on two earlier procedures. In particular, the Verbal Interaction Task (Guerney, 1977) prompted participants to discuss personal characteristics, and the Group Assessment of Interpersonal Traits (Goodman, 1972) was used to measure communication behavior. Researchers have used the CHE for over three decades and have added to the basic method in several ways.

The samples collected in research using the CHE vary in both size and composition. When the dyad is the unit of analysis, sample sizes range from approximately 30 to 275 dyads. When the discloser or the listener is of interest, sample sizes range from about 70 to 300 individuals. The CHE commonly incorporates romantic partners as participants, which allows researchers to explore informal helping behaviors within intact relationships and investigate relational outcomes. Sampling heterosexual romantic dyads also allows researchers to create distinguishable dyad members and examine sex differences and similarities in these interactions (Holmstrom *et al.*, 2005; Pasch, Bradbury, & Davila, 1997; Verhofstadt *et al.*, 2007).

In other variations of the CHE, friends serve as participants (Burgoon & Hale, 1988; MacGeorge, Guntzviller, Hanasono, & Feng, 2013). Still other studies use samples of strangers (e.g., Bodie *et al.*, 2015; Burleson & Samter, 1985; High & Solomon, 2014). Studying strangers enables researchers to focus on helping behaviors that are not colored by relational history or typical behavior patterns. Whereas some studies use both the listener and discloser as naïve participants, in other studies, one participant serves as a confederate who is trained to enact specific behaviors, including the provision of certain forms of support.

Some scholars manipulate the topic of conversation, and others allow participants to self-select their topic. One way that researchers manipulate the topic of conversation is by introducing hypothetical scenarios. Participants read and place themselves in hypothetical situations that describe social moral problems or stressful events, such as getting dumped, receiving a parking ticket, or not earning a coveted scholarship. Another way the topic is manipulated is by having a confederate disclose a problem, such as being dumped unexpectedly, which then becomes the topic of conversation (Burleson & Samter, 1985). A third way that researchers manipulate the topic of the CHE is by arousing stress within the participants. Priem and Solomon (2015), for example, used negative feedback following a task to induce feelings of stress in disclosers. Disclosers' stress and the negative feedback they received served as the basis for subsequent conversation.

Alternatively, participants can identify their own topics. When self-selecting topics, participants are typically instructed to choose an aspect of themselves that they wish to change or identify a personal stressor. Participants list several problems they feel comfortable disclosing and have not previously discussed with their partner. From these problems, either the discloser or the researcher determines the topic of conversation. When the researcher chooses the topic, the most severe problem often is selected (High & Solomon, 2014; Jones, 2004; MacGeorge *et al.*, 2013).

Validity

Prior research has provided validity evidence for the CHE in several ways. Some studies measure both participants' perceptions of an interaction to gauge how their evaluations compare. For example, husbands and wives generally report similar evaluations of helping interactions, but even small discrepancies produce distinct outcomes (Lawrence *et al.*, 2008; Priem *et al.*, 2009). In another variation, participants enact the role of both listener and discloser, with one interaction including a friend and a second conversation involving a stranger (Barker & Lemle, 1987; Burgoon & Hale, 1988). The second conversation allows researchers to compare the content of help-intended interactions between strangers and known partners. Some studies ask participants to indicate the realism or typicality of their interaction to confirm that it resembled a normal conversation (High & Solomon, 2014; Jones, 2004). Researchers also employ coders and use their ratings of a conversation as a comparison point against participants' evaluations (Bodie, Jones, Vickery, Hatcher, & Cannava, 2014; High & Solomon, 2014; Jones, 2004). In other cases, researchers have participants act as checks against their initial ratings or a coder's ratings of an interaction by using video- or tape-assisted recall procedures (Pistrang *et al.*, 2001; Priem *et al.*, 2009).

Reliability

Because it is a method, rather than a scale or rubric, reliability is difficult to determine. The CHE was originally intended to examine the informal helping behaviors of relational partners (Barker & Lemle, 1984, 1987), but it also has been used extensively to research marital (Lawrence *et al.*, 2008), nonverbal (Burgoon & Hale, 1988), health (Pistrang & Barker, 1998), supportive (Burleson & Samter, 1985; High & Solomon, 2014), and listening communication (Bodie *et al.*, 2015). The CHE has been applied to a variety of stressors, including the transition to parenthood (Pistrang *et al.*, 2001), personal stressors (High & Solomon, 2014), and medical traumas (Pistrang & Barker, 1998; Pistrang, Clare, & Barker, 1999), and when people require different types of support, such as emotional comfort (Priem *et al.*, 2009) and advice (MacGeorge *et al.*, 2013). These different contexts do not explicitly provide evidence of reliability; however, they demonstrate that the CHE has been used consistently in a number of interpersonal situations. Lastly, some studies measure participants' reactions to their interactions both immediately after completion and after time elapses (Girme, Overall, & Simpson, 2013; High & Solomon, 2014). To the extent that assessments of interactions are consistent across time, the reliability of the CHE is demonstrated.

Availability

The most detailed instructions for the CHE are available in Barker and Lemle (1984, 1987). As discussed in the Development section, however, this procedure has undergone several variations since its development. Accordingly, specifics in terms of its sample, research context, and instructions vary throughout its use.

Sample Studies

The first work using the CHE examined how self-reported relationship satisfaction and experience with therapy impacted the helping experience (Barker & Lemle, 1984). Although later studies have often used self-report measures as independent variables (e.g., MacGeorge *et al.*, 2013; Pasch *et al.*, 1997; Priem & Solomon, 2011), there are other types of predictor variables used within this method. For example, researchers train independent coders to assess both the discloser's behavior (Verhofstadt *et al.*, 2007) and the listener's behavior (Bodie *et al.*, 2015) during the interaction. These coded behaviors and message characteristics are then used as independent variables to predict participants' evaluations of the conversations. Researchers also have manipulated features of the helping interaction to act as independent variables (Barker & Lemle, 1987; Burgoon & Hale, 1988). For example, High and Solomon (2014) manipulated the communication channel and level of verbal person-centeredness of the conversation to examine their effects on listeners' and disclosers' perceptions of helping interactions.

The first studies using the CHE examined both self-report variables and coded behavior as dependent variables (Barker & Lemle, 1984, 1987). Since then, the outcome variables considered within this procedure differ in three main ways. First, research varies in whether it incorporates the perspective of the discloser or the listener. Whereas most studies measure a discloser's perceptions of an interaction (e.g., Holmstrom *et al.*, 2005; MacGeorge *et al.*, 2013; Pistrang & Barker, 1998), some early research coded the listener's behavior as an outcome measure (Barker & Lemle, 1984, 1987). Still other research measures or codes both the disclosers' and the listeners' evaluations of the same interaction (Bodie *et al.*, 2015; High & Solomon, 2014; Pasch *et al.*, 1997; Priem *et al.*, 2009).

A second variation concerns the timing of the dependent variables. Most research assesses outcomes immediately after completing the exercise (e.g., Burleson & Samter, 1985; Pistrang *et al.*, 1999; Priem & Solomon, 2011). In contrast, High and Solomon (2014) assessed disclosers' perceptions of the conversation and their stressor 3 weeks after their interaction. Likewise, Girme *et al.* (2013) measured people's goal achievement at 3-month intervals during the following year.

Third, researchers employing the CHE typically assess participants' perceptions of either the messages they exchange or the outcomes they experience. Many studies have measured people's perceptions of the messages exchanged during this exercise, including the helpfulness, sensitivity, credibility, quality, helpfulness, and empathy of the messages (Bodie *et al.*, 2015; Burleson & Samter, 1985; MacGeorge *et al.*, 2013; Pistrang *et al.*, 1999; see Multidimensional Evaluation of Enacted Social Support, Profile 42). Other research considers the cognitive, behavioral, or affective outcomes experienced by participants. These outcomes include the improvement in people's stressors, their cardiovascular reactivity, their cortisol levels, and their marital satisfaction following the exercise (High & Solomon, 2014; Lawrence *et al.*, 2008; Priem & Solomon, 2015). Thus, the CHE is adaptable to several different outcomes.

Critique

The CHE's chief strength is its elicitation of naturalistic interactions in controlled environments. It addresses the limitations of retrospective self-report procedures, which Barker and Lemle (1984) noted introduce "biases due both to lapse of time and to selective distortion of participants about events in their own relationship" (p. 323). Participants who interact with relational partners report that their interactions represent their typical or usual patterns of communication. Furthermore, participants' behaviors are commonly recorded, which supplies researchers with dyadic conversational and behavioral data that can be coded.

Despite these strengths, the laboratory environment used in the CHE introduces a degree of artificiality to the interactions. Any laboratory-based interaction sacrifices external validity for control, so it is unclear how communication in these interactions resembles conversations outside of the lab. Along these lines, Barker and Lemle (1984) noted that the tightly defined roles of listener and discloser might impact participants' behaviors, yielding results that are inconsistent with other methods (see also Goldsmith, 2004). Similarly, it is unclear how people's communication with strangers resembles their behavior within intact relationships that maintain a history of helping. At the same time, a recent study suggested that language style and matching patterns are rather similar for stranger and friend interactions (Cannava & Bodie, 2016). Despite these limitations, the CHE captures more realistic and typical interactions than are possible through self-report and recall procedures.

References

Barker, C., & Lemle, R. (1984). The helping process in couples. *American Journal of Community Psychology*, *12*, 321–336. doi:10.1007/bf00896753

Barker, C., & Lemle, R. (1987). Informal helping in partner and stranger dyads. *Journal of Marriage and the Family*, *49*, 541–547. doi:10.2307/352199

Bodie, G. D., Jones, S. M., Vickery, A. J., Hatcher, L., & Cannava, K. (2014). Examining the construct validity of enacted support: A multitrait-multimethod analysis of three perspectives for judging immediacy and listening behaviors. *Communication Monographs*, *81*, 495–523. doi:10.1080/03637751.2014.957223

Bodie, G. D., Vickery, A. J., Cannava, K., & Jones, S. M. (2015). The role of "active listening" in informal helping conversations: Impact on perceptions of listener helpfulness, sensitivity, and supportiveness and discloser emotional improvement. *Western Journal of Communication*, *79*, 151–173. doi:10.1080/10570314.2014.943429

Burgoon, J. K., & Hale, J. L. (1988). Nonverbal expectancy violations: Model elaboration and application to immediacy behaviors. *Communication Monographs*, *55*, 58–79. doi:10.1080/03637758809376158

Burleson, B. R., & MacGeorge, E. L. (2002). Supportive communication. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 374–424). Thousand Oaks, CA: Sage.

Burleson, B. R., & Samter, W. (1985). Consistencies in theoretical and naive evaluations of comforting messages. *Communication Monographs*, *52*, 103–123. doi:10.1080/03637758509376099

Cannava, K., & Bodie, G. D. (2016). Language use and style matching in supportive conversations among strangers and friends. *Journal of Social and Personal Relationships*. doi: 10.1177/0265407516641222

Girme, Y. U., Overall, N. C., & Simpson, J. A. (2013). When visibility matters: Short-term versus long-term costs and benefits of visible and invisible support. *Personality and Social Psychology Bulletin*, *39*, 1441–1454. doi:10.1177/0146167213497802

Goldsmith, D. J. (2004). *Communicating social support*. New York: Cambridge University Press.

Goodman, G. (1972). *Companionship therapy: Studies in structured intimacy*. San Francisco, CA: Jossey-Bass.

Guerney, B. G., Jr. (1977). *Relationship enhancement: Skill training programs for therapy, problem prevention, and enrichment*. San Francisco, CA: Jossey-Bass.

High, A. C., & Solomon, D. H. (2014). Communication channel, sex, and the immediate and longitudinal outcomes of verbal person-centered support. *Communication Monographs*, *81*, 439–468. doi:10.1080/03637751.2014.933245

Holmstrom, A. J., Burleson, B. R., & Jones, S. M. (2005). Some consequences for helpers who deliver "cold comfort": Why it's worse for women than men to be inept when providing emotional support. *Sex Roles*, *53*, 153–172. doi:10.1007/s11199-005-5676-4

Jones, S. (2004). Putting the person into person-centered and immediate emotional support: Emotional change and perceived helper competence as outcomes of comforting in helping situations. *Communication Research*, *31*, 338–360. doi:10.1177/0093650204263436

Lawrence, E., Bunde, M., Barry, R. A., Brock, R. L., Sullivan, K. T., Pasch, L. A., & Adams, E. E. (2008). Partner support and marital satisfaction: Support amount, adequacy, provision, and solicitation. *Personal Relationships*, *15*, 445–463. doi:10.1111/j.1475-6811.2008.00209.x

MacGeorge, E. L., Guntzviller, L. M., Hanasono, L., & Feng, B. (2013). Testing Advice Response Theory in interactions with friends. *Communication Research*. Published Online First. doi:10.1177/0093650213510938

Pasch, L. A., Bradbury, T. N., & Davila, J. (1997). Gender, negative affectivity, and observed social support behavior in marital interaction. *Personal Relationships*, *4*, 361–378. doi:10.1111/j.1475-6811.1997.tb00151.x

Pistrang, N., & Barker, C. (1998). Partners and fellow patients: Two sources of emotional support for women with breast cancer. *American Journal of Community Psychology*, *26*, 439–456. doi:10.1023/a: 1022163205450

Pistrang, N., Clare, L., & Barker, C. (1999). The helping process in couples during recovery from heart attack: A single case study. *British Journal of Medical Psychology*, *72*, 227–237. doi: 10.1348/000711299159970

Pistrang, N., Picciotto, A., & Barker, C. (2001). The communication of empathy in couples during the transition to parenthood. *Journal of Community Psychology*, *29*, 615–636. doi:10.1002/jcop.2000.abs

Priem, J. S., & Solomon, D. H. (2011). Relational uncertainty and cortisol responses to hurtful and supportive messages from a dating partner. *Personal Relationships*, *18*, 198–223. doi:10.1111/j.1475-6811.2011.01353.x

Priem, J. S., & Solomon, D. H. (2015). Emotional support and physiological stress recovery: The role of support matching, adequacy, and invisibility. *Communication Monographs*, *82*, 88–112. doi:10.1080/03637751.2014.971416

Priem, J. S., Solomon, D. H., & Steuber, K. R. (2009). Accuracy and bias in perceptions of emotionally supportive communication in marriage. *Personal Relationships*, *16*, 531–552. doi:10.1111/j.1475-6811.2009.01238.x

Trees, A. R. (2000). Nonverbal communication and the support process: Interactional sensitivity in interactions between mothers and young adult children. *Communication Monographs*, *67*, 239–261. doi:10.1080/03637750009376509

Verhofstadt, L. L., Buysse, A., & Ickes, W. (2007). Social support in couples: An examination of gender differences using self-report and observational methods. *Sex Roles*, *57*, 267–282. doi:10.1007/s11199-007-9257-6

Profile 15

Doctors' Interpersonal Skills Questionnaire (DISQ)

(Greco, Cavanagh, Brownlea, & McGovern, 1999)

Profiled by: Helen Meldrum, EdD

Bentley University

Construct

The Doctors' Interpersonal Skills Questionnaire (DISQ), also available as The Interpersonal Skills Questionnaire for allied health professionals (ISQ), was designed to elicit patient views of a clinician's behaviors and provide evidence of the quality of their interactions.

Instrument Type

Other-Report

Description

Michael Greco and colleagues developed the first version of the DISQ to focus on physician behaviors, and it was conceived as a method to quantify patients' impressions of the quality of their relationship with their doctor. The DISQ was originally a 12-item instrument including prompts such as "On this visit I would rate the doctor's ability to really listen to me as…"[1] The designers of the DISQ typically use it as part of a "360 feedback process" intended to provide information for professional development (see Chapter 5). The DISQ is often packaged with the ISQ as components of a holistic

1 The assessment is included as part of the Client-Focused Evaluations Programme (CFEP). The scale authors have obtained endorsements from medical regulatory agencies in both the United Kingdom and Australia (e.g., the General Medical Council and the Royal Australian College of General Practitioners).

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

assessment to profile the quality of the important relationships in a clinician's professional role. Colleagues, support staff, and patients provide information that can be documented and used as supporting materials in portfolios for continuing education credits, reaccreditation, and licensing renewals.

Administration

The 12-item DISQ takes approximately 2–3 minutes to complete. Patients rate each item using the following semantic designators: 1 = *poor*, 2 = *fair*, 3 = *good*, 4 = *very good*, and 5 = *excellent*. Exact instructions are unclear, but patients are generally asked to think about their recently completed consultation when responding. Patients also are reminded that the clinician has volunteered to undertake this review and that their honesty is appreciated. Test instructions reinforce the anonymity and confidentiality of responses. The authors suggest collecting a random sample of 40–50 "consecutive post-consultation patients" to avoid any tendency to handpick participants who are known to be favorably inclined toward their practitioners. Most questionnaires are paper and pencil, completed at the practice site, and slipped into a ballot box. Mail-in and online options also are available.

Scoring

Responses to the 12 items are averaged to calculate an index score that is notated as a percentage of a theoretically perfect score of 100% (Greco, Brownlea, & McGovern, 2001). This calculation represents a measure of interpersonal skill level.[2]

Development

The DISQ was first introduced by Greco and his research colleagues in the mid-1990s. The process of refining the test in general medical practice is described in two articles (Greco, Brownlea, McGovern, & Cavanagh, 2000; Greco *et al.*, 1999). The developers drew upon existing research and commentary from patient and physician focus groups to develop the initial measure. The first version of the DISQ incorporated a 6-point numerical response scale, but a pilot study indicated that scores were too highly slanted toward favorable reviews. A second version of the DISQ was then designed using a 5-point evaluation scale. Greco's team has since organized as a commercial enterprise, administering the DISQ/ISQ now more frequently as part of a package of assessments providing feedback from both peers and patients.

2 When used as suggested by the scale's authors, scoring is done in-house at CFEP. The director of the agency stated that attempts to use copies of the questionnaire by individuals would not only violate copyright laws but also be lacking in benchmark data, thereby invalidating any meaningful final summation (M. Greco, personal communication, June 3, 2015).

Reliability

The assessment appears to yield consistent outcomes from a variety of samples over time. Early in the development of the questionnaire, the team looked at more than 8000 patient evaluations and reported a strong internal reliability estimate (Cronbach's alpha) of .96 and test–retest reliability, $r = 0.75$ (for further details, see Greco *et al.*, 1999).

Validity

The DISQ's content validity was derived through a comparison of patient and doctor focus group discussion data and existing published research on the interpersonal skills of doctors. Greco *et al.* (2000) reported a number of validity-related findings. Construct validity was reflected in the relation between the aggregate of DISQ items with patients' overall satisfaction with that medical consultation, $r = 0.79$. Criterion validity was demonstrated by a significant correlation, $r = 0.77$, between DISQ and Falvo and Smith's (1983) Interaction Scale, which measures patients' satisfaction with physicians' interpersonal skills. Concurrent validity was observed in a moderate relationship between DISQ mean scores and observation-based ratings of external supervising clinical teachers, $r = 0.48$. This indicates a significant relationship between experienced general practitioners who used the DISQ to assess younger physicians and the ratings made by patients of those same physicians.

Availability

The ISQ/DISQ is available for purchase from Client-Focused Evaluations Programme (CFEP; http://www.cfepsurveys.com.au). The CFEP provides a step-by-step administration manual and facilitator guide. The items that comprise the scale are presented at the end of this profile and are reprinted with permission. Readers are advised that permission must be sought to use items for any purpose.

Sample Studies

Although the DISQ and ISQ appear to be used frequently in professional contexts, limited research using the instruments is available beyond that provided by its developers. Orton, Orton, and Gray (2012) employed the DISQ in a study of the effect of physician burnout on physician–patient interactions. Based on a sample of more than 500 doctors, they reported that personal burnout did not adversely affect patient satisfaction with their physician's interactions.

In an Italian study, the DISQ was used as the basis of a customized assessment on satisfaction with communication of a multiple sclerosis diagnosis disclosure (Solari *et al.*, 2010). Questions from the DISQ, in addition to other sources, were used to develop a new measure of patient experience at the point they are first told they have MS.

Researchers affiliated with the original development team have had an active program of research examining ways in which a wide variety of medical practices respond to patient data. Several studies have utilized pre-and posttest designs, including research showing that instruction in active listening shows significant improvement in patient ratings for the interpersonal skill–oriented items on the scale (e.g., Greco, Francis, Buckley, Brownlea, & McGovern, 1998). Although the active listening instruction did not affect the items related to warmth, reassurance, and confidence, the training did improve patients' ratings of their understanding of the explanations they were given of their medical conditions. Patients also reported an improvement in being allowed to tell their own story. Recent studies also tend to include feedback from colleagues (e.g., a study with a sample size of over 30,000 patients was reported on in Wright *et al.*, 2012).

An article outlining how the questionnaire can be used as part of an interpersonal skills training program provides insight into ways that physicians can learn to incorporate this type of patient feedback. Baker, Greco, O'Brien, and Squire (n.d.) reported an approximately 10% increase in the number of physicians receiving *excellent* ratings from their patients across all the DISQ items. Notably, the area seeing the greatest improvement was related to the physician's ability to listen to their patients.

Critique

Healthcare providers often face the dilemma of choosing between doing the right thing for a patient's health and just giving the patient what he or she wants. This daily reality indicates how patient ratings can swing against even the most diligent practitioner. There also are some well-documented issues with satisfaction questionnaires in general. For example, older patients and those with lower socioeconomic status are less likely to report dissatisfaction with their healthcare than are the young and well educated (Sitzia & Wood, 1997). Perhaps patients who already feel disenfranchised are hesitant to be critical because of fear of being denied good treatment. Satisfaction may often be influenced by factors beyond the control of the provider. On the other hand, accountability of healthcare practitioners must be continuously pursued.

Although many provider–patient questionnaires lack the theoretical underpinnings and advanced analysis of the DISQ/ISQ, there is a dearth of replication studies by researchers who are not affiliated with the academics that originated the assessment. Although the questionnaire does not include items directly assessing key issues like patient adherence or a felt sense of empowerment, a short test simply cannot cover all potential areas of interest. Given the considerable resources invested and a long period of testing this instrument in the field, it is likely that the DISQ/ISQ assessment is a viable option for medical practices and academics willing and able to pay for the consultation services from CFEP. It should be noted, however, that the studies reported here show large correlations between the DISQ and similar constructs. The degree to which the DISQ measures something unique from general patient satisfaction, for instance, should be tested in more robust studies.

References

Baker, L. H., Greco, M., O'Brien, M. L., & Squire, S. (n.d.). *Improving doctors' communication skills: Teaching what is measured*. Retrieved August 10, 2015 from www.calquality.org/storage/documents/patientexp/08-DIAQ**Baker.**pdf

Falvo, D. R., & Smith, J. K. (1983). Assessing residents' behavioral science skills: Patients' views of physician–patient interaction. *The Journal of Family Practice*, *17*, 479–483. PMID:7354293

Greco, M., Brownlea, A., & McGovern, J. (2001). Impact of patient feedback on the interpersonal skills of GP Registrars: Results of a longitudinal study. *Medical Education*, *35*, 748–756. doi:10.1046/j.1365-2923.2001.00976.x

Greco, M., Brownlea, A., McGovern, J., & Cavanagh, M. (2000). Consumers as educators: Implementation of patient feedback in general practice training. *Health Communication*, *12*, 173–193. doi:10.1207/S15327027HC1202_4

Greco, M., Cavanagh, M., Brownlea, A., & McGovern, J. (1999). Validation studies of the Doctors' Interpersonal Skills Questionnaire. *Education for General Practice*, *10*, 256–264.

Greco, M., Francis, W., Buckley, J., Brownlea, A., & McGovern, J. (1998). Real-patient evaluation of communication skills teaching for GP registrars. *Family Practice*, *15*, 51–57. doi:10.1093/fampra/15.1.51

Orton, P., Orton, C., & Gray, D. P. (2012). Depersonalized doctors: A cross-sectional study of 564 doctors, 760 consultations and 1876 patient reports in UK general practice. *BMJ Open*, *2*(1), e000274. doi:10.1136/bmjopen-2011-000274

Sitzia, J., & Wood, N. (1997). Patient satisfaction: A review of issues and concepts. *Social Science & Medicine*, *45*, 1829–1843. doi:10.1016/S0277-9536(97)00128-7

Solari, A., Mattarozzi, K., Vignatelli, L., Giordano, A., Russo, P. M., Uccelli, M. M., D'alessandro, R., SIMS-Trial group, & GERONIMUS group. (2010). Development and validation of a patient self-assessed questionnaire on satisfaction with communication of the multiple sclerosis diagnosis. *Multiple Sclerosis*, *16*, 1237–1247. doi:10.1177/1352458510376178

Wright, C., Richards, S. H., Hill, J. J., Roberts, M. J., Norman, G. R., Greco, M., & Campbell, J. L. (2012). Multisource feedback in evaluating the performance of doctors: The example of the UK General Medical Council patient and colleague questionnaires. *Academic Medicine*, *87*, 1668–1678. doi:10.1097/ACM.0b013e3182724cc0

Scale

The Doctors' Interpersonal Skills Questionnaire (DISQ)*
(Greco, Cavanagh, Brownlea, & McGovern, 1999)

Source: Greco *et al.* (1999). Reproduced with permission of Taylor & Francis.

Using the following scale, please respond to each of the following items:

5 = Excellent
4 = Very Good
3 = Good
2 = Fair
1 = Poor

1) My overall satisfaction with this visit to the doctor is ….
2) The warmth of the doctor's greeting to me was ….
3) On this visit I would rate the doctor's ability to really listen to me as ….
4) The doctor's explanations of things to me were ….

5) The extent to which I felt reassured by this doctor was ….
6) My confidence in this doctor's ability is ….
7) The opportunity the doctor gave me to express my concerns or fears was ….
8) The respect shown to me by this doctor was ….
9) The amount of time given to me for this visit was ….
10) This doctor's consideration of my personal situation in deciding a treatment or advising me was ….
11) The doctor's concern for me as a person in this visit was ….
12) The recommendation I would give to my friends about this doctor would be ….

*Responses to the 12 items are averaged to calculate an index score that is notated as a percentage of a theoretically perfect score of 100%.

Note: Permission from the authors is needed to use this scale for research or training purposes.

Profile 16

Effective Listening and Interactive Communication Scale (ELICS)

(King, Servais, Bolack, Shepherd, & Willoughby, 2012)

Profiled by: Gillian King, PhD[1] and Michelle Servais, PhD[2]

[1] *Bloorview Research Institute and Department of Occupational Science and Occupational Therapy, University of Toronto*
[2] *Thames Valley Children's Centre*

Construct

The Effective Listening and Interactive Communication Scale (ELICS) was designed to measure four theory-based listening/communication skills relevant to pediatric rehabilitation practice.

Instrument Type

Self-Report

Description

Therapists' listening and communication skills are fundamental to the delivery of children's rehabilitation services, but few measures comprehensively assess these skills. The Effective Listening and Interactive Communication Scale (ELICS) was developed to reflect a multifaceted conceptualization of listening. The ELICS portrays listening as a purposeful, goal-oriented, and relational activity (Bavelas, Coates, & Johnson, 2002; King, 2009b). The 24 items of this self-assessment capture four types of listening skills relevant to pediatric rehabilitation practice: *Receptive Listening* (seven items reflecting mindful attention to understand the client's situation), *Exploratory Listening* (seven items reflecting dialogue to elicit information and establish clarity about issues), *Consensus-oriented Listening* (six items reflecting brainstorming and explanation of rationales to establish shared understanding and jointly determined goals), and *Action-oriented Listening* (four items reflecting supporting and enabling clients to establish actions toward desired outcomes).

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

The ELICS is self-administered. Respondents are asked to complete the scale based on their own views of their listening and communication behavior and skills, in the context of their professional practice within the past 6 months. They are asked to indicate to what extent they actually show the behaviors or skills described, using a 7-point scale with all points labeled, including *not at all* (1), to a moderate extent (4), and *to a very great extent* (7). The measure takes approximately 10 to 15 minutes to complete.

Scoring

Mean scores are calculated for the four types of listening skills. Numeric responses for items on each subscale are averaged, producing four scores.

Development

The ELICS was developed by a team of clinicians, managers, and researchers, each with more than 10 years of experience in children's service organizations. The measure is based on a conceptualization of clinical listening/communication as a goal-oriented, relational activity consisting of multiple strategies (Boudreau, Cassell, & Fuks, 2009). The authors adopted a skill-based approach, reflecting the assumption that listening and communication skills can be improved through resources that provide feedback and prompt self-reflection, and a developmental perspective, reflecting interest in the dialogue, relationship-building, and negotiation aspects that characterize the client–practitioner communication process over time (King, 2009a). The conceptualization was functional and relational in nature, reflecting the belief that clinicians have multiple purposes in mind when they listen, which evolve over the course of the intervention relationship. Core listening and face-to-face communication skills were identified through a comprehensive review of the literature on communication skills in the fields of medicine and allied health (King *et al.*, 2012).

Measure development followed a construct approach to test development and consisted of item generation, piloting with clinicians, and psychometric testing (King *et al.*, 2012). Data from 41 pediatric rehabilitation clinicians were analyzed using principal component analysis with a varimax rotation. Inspection of eigenvalues along with interpretation of item loadings revealed four components. The eigenvalues and percentage of variance accounted for by each component were as follows: component 1 (9.0; 37.6%), component 2 (3.0; 12.5%), component 3 (2.1; 8.8%), and component 4 (1.7; 6.9%). Final items were selected through a series of item reduction steps involving inspection of item loadings and subscale reliabilities and consideration of item meaning. The criteria for retaining items were (a) a loading of at least 0.55 and (b) if an item loaded on two components, then a minimum difference of 0.2 was needed to retain the item. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy score was 0.67, indicating that it was appropriate to perform such an analysis on these data (Tabachnik & Fidell, 1989).

Reliability

The coefficient alphas for the four ELICS scales have ranged from 0.78 to 0.90, indicating very good to excellent internal consistency (King *et al.*, 2012).

Validity

There is preliminary evidence in support of the ELICS's validity. Face and content validity were ensured by the development process, which involved a review of the literature of communication skills in health service delivery and piloting with groups of clinicians (King *et al.*, 2012). Construct validity was demonstrated through principal component analysis, although future work should replicate this structure using confirmatory techniques. To provide evidence of predictive validity, a number of factors were examined, including the ELICS's ability to discriminate among novice, emerging expert, and expert pediatric rehabilitation clinicians. Clinical expertise was defined as a composite measure comprising family-centered behavior, peer- and self-rated levels of listening skill or general expertise, and years in practice. As predicted, expert clinicians had significantly higher scores than novices on Receptive, Consensus-oriented, and Action-oriented Listening (King *et al.*, 2012). For Receptive Listening, the means and standard deviations were 6.06 (0.7) and 5.29 (0.5) for experts and novices, respectively, $F(2,38) = 4.78$, $p < 0.01$, $\eta^2 = 0.2$; for Consensus-oriented Listening, 6.17 (0.5) and 5.21 (0.6), respectively, $F(2,38) = 5.73$, $p < 0.01$, $\eta^2 = 0.2$; and for Action-oriented Listening, 6.19 (0.6) and 5.04 (0.6), respectively, $F(2,38) = 10.93$, $p < 0.001$, $\eta^2 = 0.4$.

Although developed as a summative, retrospective self-assessment tool to differentiate excellent from good and novice listeners, the ELICS has been used as a self-report outcome measure (administered pretest and post test) to assess the effects of interventions targeting listening skills. Clinical responsiveness was demonstrated in a pilot study of the effects of an educational intervention on the listening skills of pediatric rehabilitation clinicians. Significant change on all four ELICS scales was found from pretest to post test and/or follow-up (2 weeks later), with large effect sizes (King *et al.*, 2017). For Receptive Listening, the means and standard deviations were 5.64 (0.76) and 6.31 (0.66) at pretest and follow-up, respectively, $F(2,10) = 5.74$, $p < 0.02$, $\eta^2 = 0.53$; for Exploratory Listening, 4.33 (0.61) and 5.29 (1.05), respectively, $F(2,10) = 12.70$, $p < 0.002$, $\eta^2 = 0.72$; for Consensus-oriented Listening, 5.17 (0.60) and 6.02 (0.64), respectively, $F(2,10) = 5.23$, $p < 0.03$, $\eta^2 = 0.51$; and for Action-oriented Listening, 5.29 (0.95) and 6.17 (0.74), respectively, $F(2,10) = 7.23$, $p < 0.01$, $\eta^2 = 0.59$.

Availability

The ELICS is included in the American Psychological Association PsycTESTS searchable electronic database for tests and measures. Participant instructions and items, organized by factor, are included below, with permission. The scale is free to use for research purposes with appropriate citation.

Sample Studies

The ELICS is a relatively new measure, published in 2012. At the time of preparation of this summary (August 2015), the measure had only been used by the authors. The ELICS was used to assess self-reported change (pre to post) in a study of an occupational therapy mentoring intervention (King *et al.*, 2011) and in a study of an educational listening skills intervention combining interprofessional discussion of digital video listening simulations with solution-focused coaching (King *et al.*, 2017).

The ELICS also has been used as a measure of clinicians' skill level. Because expert clinicians are typically excellent listeners, we have used the ELICS as part of a battery of measures to differentiate clinicians in terms of expertise level. In a study of clinicians delivering a transition program targeting life skills development for youth with disabilities (King *et al.*, 2015), novice, intermediate, and expert status was determined based on years in practice, the ELICS, the Practice Skills Inventory (O'Hare, Tran, & Collins, 2002), and a self-nomination measure of expertise. In contrast to novices, experts displayed a more holistic perspective and paid more attention to higher order issues, such as providing opportunities and enabling youth (King *et al.*, 2015).

Critique

The ELICS is a comprehensive self-report measure of listening and communication skills relevant to a specific context—pediatric rehabilitation practice. The measure may be useful in other areas of clinical practice, but this has not been explored. A review of research scales assessing listening competency (Fontana, Cohen, & Wolvin, 2015) indicated that the ELICS was dissimilar to other listening scales, which generally assess listening traits (e.g., asking questions); the ELICS assesses listening as a purposeful functional skill in a specific context (e.g., Exploratory Listening). Development work involved a relatively small sample of 41 practicing clinicians. Further psychometric information is required using a larger sample, including evidence of test–retest reliability (King *et al.*, 2012). Evidence of construct validity through confirmatory factor analytic procedures is also needed.

The primary limitation of the ELICS is its self-report nature. Bias may be reduced by using the ELICS in a guided reflection format or in conjunction with other methods of assessment and skill development. Assessing relationships with observational measures of displayed listening skills is an important direction for future research, although the correspondence may be low, as self-ratings are based on a holistic understanding of one's listening skills over multiple situations, whereas observer ratings are situation specific.

Recently, an ELICS assessment rubric (the ELICS-AR) was developed. It is an observational measure of the extent to which clinicians display listening skill behaviors, using the same 24 items as in the ELICS, rated on a 5-point performance scale (0 = *not at all*, 1 = *rarely*, 2 = *occasionally*, 3 = *frequently*, and 4 = *consistently*). Preliminary work suggests low associations with clinicians' self-perceptions.

References

Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. *Journal of Communication*, *52*, 566–580. doi:10.1111/j.1460-2466.2002.tb02562.x

Boudreau, J. D., Cassell, E., & Fuks, A. (2009). Preparing medical students to become attentive listeners. *Medical Teacher, 31*, 22–29. doi:10.1080/01421590802350776
Fontana, P. C., Cohen, S. D., & Wolvin, A. D. (2015). Understanding listening competency: A systematic review of research scales. *International Journal of Listening, 29*, 148–176. doi:10.1080/10904018.2015.1015226
King, G. (2009a). A framework of personal and environmental learning-based strategies to foster therapist expertise. *Learning in Health and Social Care, 8*, 185–199. doi:10.1111/j.1473-6861.2008.00210.x
King, G. (2009b). A relational goal-oriented model of optimal service delivery to children and families. *Physical & Occupational Therapy in Pediatrics, 29*, 384–408. doi:10.3109/01942630903222118
King, G., McPherson, A., Kingsnorth, S., Stewart, D., Glencross-Eimantas, T., Jones-Galley, K., Morrison, A., Isihie, A. M., & Gorter, J. W. (2015). Residential immersive life skills programs for youth with disabilities: Service providers' perceptions of change processes. *Disability and Rehabilitation, 37*, 2418–2428. doi:10.3109/09638288.2015.1031285
King, G. A., Servais, M., Bolack, L., Shepherd, T. A., & Willoughby, C. (2012). Development of a measure to assess effective listening and interactive communication skills in the delivery of children's rehabilitation services. *Disability and Rehabilitation, 34*, 459–469. doi:10.3109/09638288.2011.608143
King, G., Servais, M., Shepherd, T., Willoughby, C., Bolack, L., Moodie, S.,…McNaughton, N. (2017). A listening skill educational intervention for pediatric rehabilitation clinicians: A mixed-methods pilot study. *Developmental Neurorehabilitation. 20*(1), 40–42. doi:10.3109/17518423.2015.1063731
King, G., Tam, C., Fay, L., Pilkington, M., Servais, M., & Petrosian, H. (2011). Evaluation of an occupational therapy mentoring program: Effects on therapists' skills and family-centered behavior. *Physical & Occupational Therapy in Pediatrics, 31*, 245–262. doi:10.3109/01942638.2010.523451
O'Hare, T., Tran, T. V., & Collins, P. (2002). Validating the internal structure of the practice skills inventory. *Research on Social Work Practice, 12*, 653–668. doi:10.1177/1049731502012005005
Tabachnik, B. G., & Fidell, L. S. (1989). *Using multivariate statistics.* New York: Harper Collins.

Scale

Effective Listening and Interactive Communication Scale (ELICS) (King, Servais, Bolack, Shepherd, & Willoughby, 2012)

Source: King *et al.* (2012). Reproduced with permission of Taylor & Francis.

A Self-Assessment Scale of Listening and Communication Skills[1]

Please complete the following scale based on your own views of your listening and communication behavior and skills, in the context of your professional practice within the past six months. Please think about your face-to-face meetings or sessions with people to whom you provide services, as a professional in your discipline.

1 This title can be used for the scale when administered, in place of the formal name.

Please base your answers on your first impressions. Each question asks you to indicate <u>to what extent</u> you actually show the behaviors or skills described, on a scale from 1 (Not at All) to 7 (To a Very Great Extent).

For each question, we would like you to think about **the degree to which you displayed each of the behaviors or skills** described in this questionnaire. The rating you select should NOT represent the frequency with which you engaged in each particular behavior or displayed each skill but rather your level of involvement and investment in practicing it.

Important Instructions:

1) We would like you to describe your <u>**actual**</u> behavior and skills, rather than what you feel would be ideal service. We recognize that professionals may be unable to display behavior to the extent they might wish, due to caseload size, policies, and other constraining factors. Please be assured that your **confidential** responses will not be viewed as a judgment of you or how you provide services.
2) Please recognize that just because a behavior is addressed by this measure it DOES NOT mean that it is necessarily an important behavior or skill for all professions or to all professionals. Thus, do not feel that selecting a low number is equivalent to giving yourself a poor evaluation.
3) We would like you to think about <u>**your experiences**</u> as a provider of services <u>**over the past six months**</u>.
4) The terms “the person” or “people” refers to the individuals to whom you provide services (this could be children, youth, and/or parents).
5) The term “encounter” means face-to-face interactions (i.e., meetings or sessions with people).
6) The term “issue” refers to the problem or reason why you are meeting with someone in your professional capacity.

7 = To a Very Great Extent
6 = To a Great Extent
5 = To a Fairly Great Extent
4 = To a Moderate Extent
3 = To a Small Extent
2 = To a Very Small Extent
1 = Not at All

To what extent do you…

<u>Consensus-oriented Listening</u>

1) pay particular attention to nonverbal cues when you first meet people?
2) explain reasons or rationales for the things you propose?
3) clarify agreed-upon goals?
4) try to ensure that the person understands what has been achieved or agreed upon in the encounter?
5) check that the other person has understood what you have said?
6) brainstorm ideas with people?
7) try to reach a shared perspective or jointly agreed upon decision?

Exploratory Listening

1) try to keep people talking about their issues, even when you are having a busy day?
2) encourage people to ask questions?
3) provide information, education, and instruction?
4) challenge people who seem stuck on an issue, to encourage them?
5) explore people's worries and concerns?
6) feel you are able to identify a person's greatest worry or concern about an issue, and the reason why?
7) challenge people when you think this will be helpful in assisting them to take a next step?

Receptive Listening

1) acknowledge that people's concerns are legitimate, to make them feel heard?
2) try to be open to what people are saying to you?
3) listen to what is *not* being said?
4) try to fully understand the person's perspective?
5) try to be present in the moment with the person?
6) try to be aware of when people want to be engaged and when they do not?

Action-oriented Listening

1) encourage people to lead the direction and pace of intervention?
2) engage in action planning to establish the next step?
3) prioritize issues with people?
4) work to create a shared vision of the desired end outcome?

Note: Labels should be removed and items randomized prior to administration. The version of the ELICS provided for this profile has a standard, randomized order. Readers are encouraged to access that order by consulting the APA PsycTESTS database to get the form.

Profile 17

Empathic Accuracy: Standard Stimulus Paradigm (EA-SSP)

(Marangoni, Garcia, Ickes, & Teng, 1995)

Authors: Vivian Ta, MSc and William Ickes, PhD

University of Texas at Arlington

Construct

The construct of empathic accuracy is measured as a performance variable. Specifically, in the standard stimulus paradigm, empathic accuracy is measured as the extent to which individual perceivers accurately infer the thoughts and feelings of the same target person(s) from a "standard stimulus" video recording of the target person(s)' interaction with another person. Accurately inferring others' thoughts and feelings is mentioned as a key to effective listening in several definitions and models of the process (see Chapter 1).

Instrument Type

Cognitive Assessment

Description

To measure empathic accuracy via the standard stimulus paradigm (EA-SSP), a researcher must first obtain one or more video-recorded interactions using the unstructured dyadic interaction paradigm (UDIP) (see UPID profile, Profile 18). For the EA-SSP, participants are asked to view one or more video recordings of other people's dyadic interactions. In most cases, a "standard stimulus" video is made by compiling excerpts of interactions between different sets of interaction partners (see Gleason, Jensen-Campbell, & Ickes, 2009; Hall *et al.*, 2014); however, a single, extended interaction of only two partners also can be used for this purpose (as, e.g., in the video recordings of the individual therapy sessions that were used as "standard stimulus" recordings in the study by Marangoni *et al.*, 1995).

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

The interacting partners can represent a variety of relationship types, depending on the specific goals of the study being conducted. For example, they can be strangers, same-sex friends, dating couples, parent and child, teacher and student, doctor and patient, therapist and client, or salesperson and customer, to name a few.

As noted here, essentially the same methods that are used in the UDIP also are used to develop the video recordings that are used in the standard stimulus paradigm (see UPID profile, Profile 18). Once these standard stimulus videos have been made, they are then shown to participants who are given the task to try to infer, as accurately as possible, the specific content of the thoughts and feelings of one or more of the interaction partners who appear in the recording. Because the actual thoughts and feelings reported by the interaction partners were obtained, comparisons can be made with the inferred thoughts and feelings of EA-SSP participants. Empathic accuracy is the degree to which perceivers correctly infer the thoughts and feelings that were reported by the people in the video-recorded interactions.

Administration

Study participants all view the same standard stimulus recordings of interactions that occurred between other people (i.e., the target persons). The participants are instructed to infer the specific content of the target person's reported thought or feeling at predetermined intervals. The form used by the observing participant is adapted from that used with the UDIP, with the exception of the changing perspective (i.e., a perceiver is attempting to accurately infer the interaction partner[s]' thoughts and feelings in the UPID profile) (see Profile 18).

Scoring

To obtain a measure of empathic accuracy, the perceiver's empathic inferences are compared to the target person's reported thought or feeling. Scoring follows the procedures outlined in the UDIP profile. Specifically, independent raters compare the target person's actual thought or feeling with the inferred thought or feeling and judge how similar they are on a 3-point scale: 0 (*essentially different content*), 1 (*similar, but not the same, content*), and 2 (*essentially the same content*). The similarity ratings of the set of independent raters are averaged for each inference. Then, those averaged ratings are summed across all inferences to compute the *total accuracy points* earned by each perceiver. The total accuracy points will be greater for perceivers who make many inferences than for those who make few inferences. Therefore, each perceiver's total accuracy points are divided by the maximum number of accuracy points possible (number of inferences times the maximum score per inference) and multiplied by 100 to obtain a percent-correct empathic accuracy measure that can range from 0 to 100. This percentage measure of empathic accuracy is conveniently scaled, is easily interpreted, and corrects for differences in the total number of inferences made.

The EA-SSP can be meaningfully compared across all perceivers in a study, because all of the individual perceivers have inferred thoughts and feelings from the same (set of) target person(s). Thus, the standard stimulus paradigm is particularly well suited to study how individual differences in perceiver characteristics (e.g., personality, ability, and attention) are related to individual differences in the perceivers' empathic accuracy scores.

Development

The first study using the EA-SSP to assess the "open-ended inference" measure of empathic accuracy was reported by Marangoni *et al.* (1995). In this study, the researchers asked 80 participants (the "perceivers") to view a standard stimulus video recording that contained edited versions of three interactions involving different female clients who each discussed a real relationship problem with the same client-centered, male therapist. With the clients' consent, each client's therapy session was video recorded "live," without any rehearsal. Immediately after their therapy session, each client was debriefed and asked, while viewing a video recording of their therapy session, to create a written record of all the specific thoughts and feelings they had had during the session and the specific times at which they had occurred (see UPID profile, Profile 18).

Once the standard stimulus tape had been created, 80 perceivers viewed the standard stimulus video recording in individual sessions, and, at each of the points at which a client's thought or feeling had been reported, inferred the specific content of that feeling by writing their inference down on a thought/feeling inference form. The empathic accuracy of each perceiver with respect to each client's thoughts and feelings was then computed using the scoring and computation procedure described here.

In the 20 years since the Marangoni *et al.* (1995) study was published, other investigators have developed their own standard stimulus videos to fit their particular research needs. Some notable examples can be found in the studies by Gleason *et al.* (2009) and Hall *et al.* (2014).

Reliability

The interrater reliability of the perceiver's empathic accuracy, as assessed by independent raters using the scoring procedure described in this profile, generally averages about .90 with as many as 6–8 raters. With fewer raters, this value will decline, but it is usually in the mid-.80s when 4–5 raters are used (Ickes, 2001). Cross-target consistency was .86 (Cronbach's alpha) across the three target tapes that were used in the first standard stimulus study conducted by Marangoni *et al.* (1995). In a standard stimulus study in which perceivers inferred the thoughts and feelings of different teacher–student pairs (Gleason *et al.*, 2009), the perceivers' overall empathic accuracy for the set of eight middle school teachers was strongly correlated ($r = .73$) with their overall accuracy for the set of eight middle school students, again supporting the generality of the empathic accuracy measure across different categories of target persons. The average interrater reliabilities of the empathic accuracy ratings provided by the trained raters in these two studies were .80 and .95, respectively.

Validity

The validity portfolio for the EA-SSP includes findings that are consistent with commonsense expectations about what variables should be related to the ability to accurately infer other people's thoughts and feelings. For example, Marangoni *et al.* (1995) found that perceivers' empathic accuracy scores were significantly greater at the end of the psychotherapy recordings than at the beginning, consistent with the expectation that perceivers would become better acquainted with the clients and their problems over time. In addition, perceivers who received immediate feedback regarding the clients' actual thoughts and feelings during the middle portion of each recording achieved better empathic accuracy scores by the end of the recording than perceivers who did not receive feedback, consistent with the expectation that accurate feedback would help perceivers correct any misconceptions and incorrect cognitive frames that limited their empathic accuracy.

Findings obtained in the research reported by Hall *et al.* (2014) also help to substantiate the construct validity of the EA-SSP. In their study, excerpts from various interactions between internal medicine patients and their physicians during regularly scheduled exams comprised the standard stimulus video.

The results reported by Hall *et al.* (2014) revealed that the Test of Accurate Perception of Patients' Affect (TAPPA), which was used to assess how accurately healthcare professionals could "read" the thoughts and feelings of patients, was related to other aspects of interpersonal accuracy. For example, accuracy on the TAPPA was positively correlated with accurate recall of what an actor-physician said in a scripted interaction. In addition, the TAPPA was related to other tests of emotion recognition, such as the Patient Emotion Cue Test (PECT), with correlation coefficients ranging from .25 to .56.

Availability

Instead of attempting to develop a single standard stimulus video that all researchers can use in studies with widely varying goals and participant samples, this paradigm encourages researchers to develop different videos designed to meet the needs of specific studies having their own unique goals and participant samples. (The thoughts-and-feelings form referenced throughout is found in the UPID profile, Profile 18.)

Sample Studies

The standard stimulus paradigm has been used to compare the differences in empathic accuracy between mildly to moderately autistic individuals and their normally developing counterparts. In one such study, the mildly autistic individuals showed a deficit in their ability to infer the target person's thoughts and feelings (Demurie, DeCorel, & Roeyers, 2011). A study by Ponnet, Buysse, Roeyers, and De Clercq (2008) yielded similar results and reported that this effect was more pronounced when the target video recording was an unstructured "get to know you" conversation between two individuals versus a structured "get to know you" conversation between two people who were provided with a set of questions they could use to learn about each other.

The standard stimulus paradigm also has been used to investigate whether children's empathic accuracy is related to their peer relationships and adjustment. In a study by Gleason *et al.* (2009), 116 children in grades 5 through 8 completed the EA-SSP as well as measures that assessed their peer relationships. The results indicated that children who were less skilled at inferring other people's thoughts and feelings were more likely to experience adjustment problems than children who were more skilled (also see RCQ and Interpersonal Decentering profiles, Profiles 56 and 26, respectively).

In Hall *et al.* (2014), the EA-SSP was used to assess how accurately healthcare professionals could “read” the thoughts and feelings of patients appearing in the Test of Accurate Perception of Patients’ Affect (TAPPA). The TAPPA includes 48 audiovisual clips of patients interacting with their physicians in real medical visits. The healthcare professionals (mostly nurses) who completed this measure were instructed to infer the thoughts and feelings of the patient in each clip. Overall, the findings suggest that the ability to infer patients’ thoughts and feelings is relevant in the clinical practice of healthcare professionals, and that this ability appears to improve over time.

Supplemented by the use of signal detection analyses, the EA-SSP also has been used to document attributional biases that contribute to interpersonal conflicts. For example, a study by Schweinle, Ickes, and Bernstein (2002) found that men who reported abusing their female partners were more likely than nonabusive men to display the biased belief that women harbor critical and rejecting thoughts and feelings about their male partners. Specifically, the abusive men perceived criticism and rejection significantly more often than it was actually reported when they were asked to infer the thoughts and feelings of the target women in the standard stimulus recordings. Investigating a different type of perceptual bias, Schmid Mast, Hall, and Ickes (2006) found that men who seek a dominant role in their romantic relationships are more likely than men who prefer a subordinate position to attribute power-related thoughts and feelings to others.

Critique

Similar to the study of empathic accuracy within the UDIP, studying empathic accuracy within the EA-SSP also requires a large investment in research space, equipment, trained raters, and time. It is important to note that the target persons in the standard stimulus videos must be able to accurately report their thoughts and feelings, because any inaccuracy in their reporting will compromise the validity of any inferences made by perceivers in subsequent studies.

In addition, developing appropriate standard stimulus videos is not an easy task. To create the respective videos used in the studies by Marangoni *et al.* (1995), Gleason *et al.* (2009), and Hall *et al.* (2014), the researchers put a great deal of creative thought and planning into the types of videos needed. In the two more recent studies, they also recruited the help of media professionals to film and edit the interactions from which the final, standard stimulus video recording would be compiled. However, the effort required is justified because the diversity of these studies helps to establish the generalizability of research findings.

The major advantage of the EA-SSP is that all participants see exactly the same video and make the same set of thought–feeling inferences. As a result, empathic accuracy scores can be meaningfully compared across all participants and related to personality traits and other individual difference factors. The EA-SSP is also well suited to studies assessing the differential impact of varying information channels on perceivers’ empathic accuracy scores (see, e.g., Gesn & Ickes, 1999; Hall & Schmid Mast, 2007; Zaki, Bolger, & Ochsner, 2009).

References

Demurie, E., DeCorel, M., & Roeyers, H. (2011). Empathic accuracy in adolescents with autism spectrum disorders and adolescents with attention-deficit/hyperactivity disorder. *Research in Autism Spectrum Disorders*, *5*, 126–134. doi:10.1016/j.rasd.2010.03.002

Gesn, P., & Ickes, W. (1999). The development of meaning contexts for empathic accuracy: Cannel and sequence effects. *Journal of Personality and Social Psychology*, *77*, 746–761. doi:10.1037/0022-3514.77.4.746

Gleason, K. A., Jensen-Campbell, L. A., & Ickes, W. (2009). The role of empathic accuracy in adolescents' peer relations and adjustment. *Personality and Social Psychology Bulletin*, *35*, 997–1011. doi:10.1177/0146167209336605

Hall, J. A., & Schmid Mast, M. (2007). Sources of accuracy in the empathic accuracy paradigm. *Emotion*, *7*, 438–446. doi:10.1037/1528-3542.7.2.438

Hall, J. A., Ship, A. N., Ruben, M. A., Curtin, E. M., Roter, D. L., Clever, S. L., Smith, C. C., & Pounds, K. (2014). The Test of Accurate Perception of Patients' Affect (TAPPA): An ecologically valid tool for assessing interpersonal perception accuracy in clinicians. *Patient Education and Counseling*, *94*, 218–223. doi:10.1016/j.pec.2013.10.004

Ickes, W. (2001). Measuring empathic accuracy. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: theory and measurement (LEA Series in Personality and Clinical Psychology*, pp. 219–241). Mahwah, NJ: Lawrence Erlbaum Associates.

Marangoni, C., Garcia, S., Ickes, W., & Teng, G. (1995). Empathic accuracy in a clinically relevant setting. *Journal of Personality and Social Psychology*, *68*, 854–869. doi:10.1037/0022-3514.68.5.854

Ponnet, K., Buysse, A., Roeyers, H., & De Clercq, A. (2008). Mind reading in young adults with ASD: Does structure matter? *Journal of Autism and Developmental Disorders*, *38*, 905–918. doi:10.1007/s10803-007-0462-5

Schmid Mast, M., Hall, J., & Ickes W. (2006). Inferring power-related thoughts and feelings in others: a signal detection analysis. *European Journal of Social Psychology*, *36*, 1–10. doi:10.1002/ejsp.335

Schweinle, W., Ickes, W., & Bernstein, I. (2002). Empathic inaccuracy in husband to wife aggression: The overattribution bias. *Personal Relationships*, *9*, 141–158. doi:10.1111/1475-6811.00009

Zaki, J., Bolger, N., & Ochsner, K. (2009). Unpacking the informational bases of empathic accuracy. *Emotion*, *9*, 478–487. doi:10.1037/a0016551

Profile 18

Empathic Accuracy: Unstructured Dyadic Interaction Paradigm (EA-UDIP)

(Ickes, Stinson, Bissonnette, & Garcia, 1990)

Authors: Vivian Ta, MSc, and William Ickes, PhD

University of Texas at Arlington

Construct

Empathic accuracy is measured as a performance variable. In the unstructured dyadic interaction paradigm (UDIP), empathic accuracy is measured as the extent to which a perceiver accurately infers a target person's thoughts or feelings from a video recording of their spontaneous interaction together. Inferring others' thoughts and feelings during interaction is a key component of effective listening, according to several definitions and models (see Chapter 1).

Instrument Type

Cognitive Assessment

Description

The UDIP involves unobtrusively video recording an unstructured interaction between two individuals. After the interaction, the participants are asked to view the recording twice. Following the first viewing, they are asked to record the thoughts and feelings they had during the interaction using a standard form (see below). Following the second viewing, they are asked to infer the thoughts and feelings of their partner. The measure of empathic accuracy is derived from these data using standard procedures.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

When the UDIP is used to measure empathic accuracy, pairs of research participants (which can represent any type of pre-existing relationship, from strangers to friends or marriage partners) are recruited for a given session. The experimenter escorts the two participants into an observation room and asks them to take a seat on a sofa. The room is equipped with a wireless microphone and video camera, both concealed in a way that enables the dyad members to be unobtrusively recorded. Once both participants have been seated in the observation room, the experimenter "discovers" a reason for having to run a quick errand (either to retrieve additional consent forms or to replace a slide projector bulb that has apparently just burned out), and leaves the participants alone together. At that point, a research assistant in the control room activates the video equipment to begin recording the dyad members' unstructured interaction for a predetermined duration.

At the end of the observation and recording period, the experimenter returns to the observation room, and the video recording is terminated. The experimenter probes for any evidence of suspicion and then conducts a partial debriefing. The participants are told that they have been video recorded for the purpose of studying their naturally occurring interaction behavior. In addition, they are informed that if either of them objects to having been recorded without their permission, they may exercise their right to have the recording erased immediately. If both participants agree to release their recorded interaction as a source of data, they are asked to sign a consent form in which they also agree to view the video-recorded interaction.

To view the interaction, the participants are seated in separate but identical cubicles. Using a start/pause control, they independently pause the video recording at the points where they remember having had thoughts or feelings. The entire interaction is viewed so that a complete list of all their thoughts and feelings is created. Participants cannot reverse or fast-forward the recording, as they have access only to a start/pause control button.

Participants are asked to report all of the thoughts and feelings they distinctly remember having had during the interaction, but not to report any thoughts or feelings that they experience for the first time while viewing the recording. At each of these recording stops, the participants use a coding form to record: (a) the time the thought or feeling occurred (available from a time counter that is displayed on the video image), (b) whether they were experiencing a thought or a feeling at that time, and (c) the specific content of the thought or feeling. This procedure is repeated until both dyad members have independently logged all of their actual thoughts and feelings during the recorded interaction.

The participants are then asked to view the recording a second time in order to infer the specific thoughts and feelings that their interaction partner reported having had at each of the partner's recording stops. To do this, a research assistant pauses the recording at each of the times the participant's interaction partner reported having had a specific thought or feeling; participants write down their thought or feeling inferences at each of these pauses.

Scoring

To obtain a measure of empathic accuracy, one must assess the degree to which the content of each of the perceiver's empathic inferences matches the content of the corresponding thought or feeling that the target person actually reported. To accomplish this, independent raters compare each actual thought or feeling with the inferred thought or feeling and judge how similar they are on a 3-point scale: 0 (*essentially different content*), 1 (*similar, but not the same, content*), and 2 (*essentially the same content*). The similarity ratings are averaged across the set of independent raters for each inference. Then, those averaged ratings are summed across all inferences to compute the *total accuracy points* earned by each perceiver. Because total accuracy points will be greater for perceivers who make many inferences than for those who make few inferences, each perceiver's total accuracy points must be divided by the maximum number of accuracy points possible (number of inferences times the maximum score per inference) and multiplied by 100 to obtain a percent-correct empathic accuracy measure that can range from 0 to 100. This percentage measure of empathic accuracy is conveniently scaled, is easily interpreted, and corrects for differences in the total number of inferences made. For example, imagine the following scenario:

Person A reported three thoughts and feelings during the interaction with Person B. Person B then inferred each of Person A's thoughts and feelings. Two independent raters then scored Person B's accuracy:

	Inference #1	Inference #2	Inference #3
Rater 1	1	1	2
Rater 2	1	0	1

The similarity ratings (i.e., the "accuracy points") were then averaged across raters for each inference:

	Inference #1	Inference #2	Inference #3
Rater 1	1	1	2
Rater 2	1	0	1
Average	1	0.5	1.5

Adding all of the averaged ratings gives us Person B's "total accuracy points" earned, which is 3 ($1 + 0.5 + 1.5 = 3$). This number is then divided by the maximum number of accuracy points possible (a total of three inferences and a maximum score of 2 per inference; $3 \times 2 = 6$), which gives us 0.5. Finally, this number is multiplied by 100 to obtain a percent-correct empathic accuracy score, which gives us 50.

Development

The method for assessing empathic accuracy through the UDIP was first introduced by Ickes, Stinson, Bissonnette, and Garcia (1990). Their goal was to develop a method for measuring empathic accuracy that would meet all of the criteria implied by Carl Rogers's (1957) definition of "accurate empathy": (a) the method should allow one person (the perceiver) to generate his or her own inferences about the content of the thoughts and feelings of another person (the target person), (b) the method should allow an assessment of the degree to which the perceiver's inferences matches the content of the target person's actual thought or feeling, and (c) the method should provide a way to track the perceiver's empathic accuracy over time.

Reliability

Using Cronbach's alpha, with raters treated as the column variable, interrater reliability averages about .90 with as many as 6–8 raters. With fewer raters, this value will decline, but is typically in the mid-.80s when 4–5 raters are used (Ickes, 2001).

Validity

The validity portfolio of this performance measure of empathic accuracy can be found in the large body of research findings accumulated for more than two decades (for reviews, see Hodges, Lewis, & Ickes, 2014; Ickes, 2009; Ickes & Hodges, 2013). Predictions have been derived from both common sense (Graham, 1994; Stinson & Ickes, 1992) and formal theory (Simpson, Ickes, & Grich, 1999; Simpson, Oriña, & Ickes, 2003). In both cases, supportive evidence has been obtained. On the other hand, the validity of the measure also rests on the assumption that target persons can accurately report the content of their own thoughts and feelings, an assumption that has received initial support (Ickes, Robertson, Tooke, & Teng, 1986) but needs to be examined more extensively.

Availability

To assess empathic accuracy using the unstructured dyadic interaction paradigm, experimenters must have access to the necessary equipment and lab space (see Ickes, 2001, for guidelines). In most cases, the setting must allow interaction partners to be unobtrusively video- and audio-recorded. Researchers must also have access to equipment and coding forms that enable participants to view the interaction in which they have just participated and to record their own thoughts and feelings as well as their inferences about the specific content of their partner's thoughts and feelings. A standard coding form is shown at the end of this profile.

Sample Studies

The UDIP has been used to study empathic accuracy between strangers, friends, dating partners, and marriage partners. For example, in a study of strangers' interactions, Ickes *et al.* (1990b) found that perceivers' empathic accuracy increased as the physical attractiveness of their opposite-sex interaction partner increased, suggesting that physical attractiveness motivates the perceiver to want to get to know the partner better and that increased motivation results in greater empathic accuracy. In a study involving heterosexual dating partners (Simpson, Ickes, & Blackstone, 1995), participants rated the attractiveness potential of available partners of the opposite sex while sitting next to their current dating partner. The results revealed that dating partners who felt highly threatened while doing this task displayed evidence of *motivated empathic inaccuracy*—that is, not wanting to know what their dating partner was thinking and feeling. In a study of marriage partners, Simpson *et al.* (2003) videotaped married couples as they attempted to resolve a problem in their marriage. The results showed that when the partner's thoughts and feelings were relationship threatening, greater empathic accuracy was related to a pretest-to-posttest decline in the perceiver's feeling of closeness to the partner. However, when the partner's thoughts and feelings were nonthreatening, greater empathic accuracy was related to a pretest-to-posttest increase in the perceiver's feeling of closeness to the partner.

The unstructured dyadic interaction paradigm also has been used to study asymmetries in empathic accuracy within relationships. For example, Clements, Holtzworth-Munroe, Schweinle, and Ickes (2007) found that the empathic accuracy of abusive men was low in regard to their own wives' thoughts and feelings but not in regard to the thoughts and feelings of other women (in other words, the abusive men were selectively inaccurate). There also are differences in empathic accuracy across relationships that possess various degrees of intimacy and acquaintanceship. For example, Thomas and Fletcher (2003) compared empathic accuracy between strangers, friends, and dating partners and found that dating partners were significantly more accurate than both friends and strangers.

Because this method depends on participants verbalizing their inferences about the other person's thoughts from cues provided by the target person's verbal (and nonverbal) messages, verbal intelligence is expected to predict empathic accuracy. Ickes *et al.* (1990b found that grade point average (GPA) predicted empathic accuracy scores in American university students. A separate study conducted 18 years later also reported a significant positive correlation between IQ and empathic accuracy (Ponnet, Buysse, Roeyers, & De Clercq, 2008).

Critique

The UDIP assesses empathic accuracy in spontaneous, naturalistic interaction, so the level of mundane realism and applicability to real-life interaction is high. The dyadic interaction paradigm is useful for studying empathic accuracy in pairs of individuals whose level of acquaintance can vary widely, depending on the purposes of the study (e.g., strangers, acquaintances, close friends, or couples who are dating, married, or cohabiting). It is also useful for making comparisons between certain types of dyads (e.g., strangers versus friends) and within certain types of dyads (e.g., a person who scores high in extraversion paired with a partner who scores low in extraversion).

Validity evidence for the empathic accuracy measure, as well as for the ability of raters to be consistent, is good in studies using the unstructured dyadic interaction paradigm. However, the paradigm itself is quite demanding to use, requiring a large investment in space, equipment, time, and trained raters. Because data are collected in dyads, special statistical techniques are required to deal with the issue of dyadic interdependence, such as multilevel modeling (see Gonzalez & Griffin, 1997; Kenny, 1988). And, as noted above, the participants must accurately report their thoughts and feelings while in the role of target persons for the subsequent measure of their partners' empathic accuracy to be valid. However, extensive evidence that target persons can—and do—accurately report their thoughts and feelings is available in Ickes *et al.* (1986).

References

Clements, K., Holtzworth-Munroe, A., Schweinle, W., & Ickes, W. (2007). Empathic accuracy of intimate partners in violent versus nonviolent relationships. *Personal Relationships*, *14*, 369–388. doi:10.1111/j.1475-6811.2007.00161.x

Gonzalez, R., & Griffin, D. (1997). On the statistics of interdependence: Treating dyadic data with respect. In S. Duck, K. Dindia, W. Ickes, R. Milardo, R. Mills, & B. Sarason (Eds.), *Handbook of personal relationships: Theory, research, and interventions* (2nd ed., pp. 271–302). Chichester, UK: Wiley.

Graham, T. (1994). *Gender, relationship, and target difference in empathic accuracy*. Unpublished master's thesis. University of Texas at Arlington.

Hodges, S. D., Lewis, K. L., & Ickes, W. (2014). The matter of other minds: Empathic accuracy and the factors that influence it. In P. Shaver, M. Mikulincer (Eds.), J. A. Simpson, & J. Dovidio (Assoc. Eds.), *APA handbook of personality and social psychology: Vol 2. Interpersonal relations and group processes* (pp. 319–348). Washington, DC: APA.

Ickes, W. (2001). Measuring empathic accuracy. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement (The LEA Series in Personality and Clinical Psychology*, pp. 219–241). Mahwah, NJ: Lawrence Erlbaum Associates.

Ickes, W. (2009). *Strangers in a strange lab: How personality shapes our initial encounters with others*. New York: Oxford University Press.

Ickes, W., & Hodges, S. D. (2013). Empathic accuracy. In J. Simpson & L. Campbell (Eds.), *Handbook of close relationships* (pp. 348–373). Oxford: Oxford University Press.

Ickes, W., Robertson, E., Tooke, W., & Teng, G. (1986). Naturalistic social cognition: Methodology, assessment, and validation. *Journal of Personality and Social Psychology*, *51*, 66–82. doi:10.1037/0022-3514.51.1.66

Ickes, W., Stinson, L., Bissonnette, V., and Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology*, *59*, 730–742. doi:10.1037/0022-3514.59.4.730

Kenny, D. A. (1988). The analysis of data from two-person relationships. In S. Duck, D. Hay, S. Hobfoll, W. Ickes, & B. Montgomery (Eds.), *Handbook of personal relationships: Theory, research, and interventions* (pp. 57–77). Chichester, UK: Wiley.

Ponnet, K., Buysse, A., Roeyers, H., & De Clercq, A. (2008). Mind reading in young adults with ASD: Does structure matter? *Journal of Autism and Developmental Disorder*, *38*, 905–918. doi:10.1007/s10803-007-0462-5

Rogers, C. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, *21*, 95–103. 10.1037/h0045357

Simpson, J., Ickes, W., & Blackstone, T. (1995). When the head protect the heart: Empathic accuracy in dating relationships. *Journal of Personality and Social Psychology*, *69*, 629–641. doi:10.1037/0022-3514.69.4.629

Simpson, J. A., Ickes, W., & Grich, J. (1999). When accuracy hurts: Reactions of anxious-uncertain individuals to a relationship-threatening situation. *Journal of Personality and Social Psychology*, *76*, 754–769. doi:10.1037/0022-3514.76.5.754

Simpson, J. A., Oriña, M. M., & Ickes, W. (2003). When accuracy hurts, and when it helps: A test of the empathic accuracy model in marital interactions. *Journal of Personality and Social Psychology*, *85*, 881–893. doi:10.1037/0022-3514.85.5.881
Stinson, L., & Ickes, W. (1992). Empathic accuracy in the interactions of male friends versus male strangers. *Journal of Personality and Social Psychology*, *62*, 787–797. doi:10.1037/0022-3514.62.5.787
Thomas, G., & Fletcher, G. J. O. (2003). Mind-reading accuracy in intimate relationships: Assessing the roles of the relationship, the target, and the judge. *Journal of Personality and Social Psychology*, *85*, 1079–1094. doi:10.1037/0022-3514.85.6.1079

Scale

Standard Form for Recording Thoughts and Feelings (Ikes, Stinson, Brisonnette, & Garcia, 1990)

DATE ________

NUMBER ________

M F

TIME	THOUGHT OR FEELING	+, 0, –	
	☐ He/she was thinking: ☐ He/she was feeling:	+ 0 –	
	☐ He/she was thinking: ☐ He/she was feeling:	+ 0 –	
	☐ He/she was thinking: ☐ He/she was feeling:	+ 0 –	
	☐ He/she was thinking: ☐ He/she was feeling:	+ 0 –	
	☐ He/she was thinking: ☐ He/she was feeling:	+ 0 –	
	☐ He/she was thinking: ☐ He/she was feeling:	+ 0 –	

Profile 19

Facilitating Listening Scale (FLS)

(Bouskila-Yam & Kluger, 2011)[1]

Profiled by: Avraham N. Kluger[1] and Osnat Bouskila-Yam[2]

[1] The Hebrew University of Jerusalem
[2] Osnat Bouskila-Yam Management Consultancy

Construct

The Facilitating Listening Scale (FLS) assesses speaker perceptions of the listening behaviors of their interlocutors and the consequences of these behaviors for the speakers.

Instrument Type

Self-Report, Other-Report

Description

Developed by Bouskila-Yam and Kluger (2011), the FLS contains nine subscales that measure perceptions of both the target's constructive and the destructive listening behaviors, as well as perceptions of positive and negative consequences for the speaker. Four of the subscales are general in nature: The *Constructive Listening Behavior* and the *Destructive Listening Behavior* subscales reflect perceptions that the listener both pays attention and creates a facilitating atmosphere for talking (or is judgmental and dismissive). The *Positive Listening Consequences* and the *Negative Listening Consequences* subscales reflect attributions that the listener makes the speaker feel comfortable and liked (or worried). Of these subscales, three are measured with 10 items each; the Negative Listening Consequences subscale contains three items.

1 The preparation of this report was supported by grants from the Recanati Fund at the School of Business Administration, and by The Israel Science Foundation (145/12) to the first author.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

The remaining subscales are specific and measured with two to four items each: *reframing* (rephrasing the speaker's message), *domineering* (interrupting the speaker), *escaping* (the listener is busy with one's phone, computer, etc.), *no time* (indicating time limits to listening), and *changing the subject*. It is possible to use the subscales as standalone measures.

Administration

The FLS is a self-administered questionnaire that takes about 10 minutes to complete. Respondents complete the FLS with respect to a particular target (e.g., one's supervisor) and are invited to consider "your experience of being listened to by your ____," where the blank could be any type of target, like a supervisor or partner.

Scoring

Respondents can receive a total of nine scores from items on the FLS. To obtain scores, items that constitute that subscale are averaged. No normative data have been presented in published work to date.

Development

The FLS was designed to capture behaviors that have the potential to facilitate growth and change of the speaker (Bavelas, Coates, & Johnson, 2000; Pasupathi, 2001; Rogers & Roethlisberger, 1991/1952). Bouskila-Yam and Kluger (2011) adopted items from 10 published listening instruments (see Table P19.1) and developed 11 new items to reflect the consequences of listening based on theories pertaining to listening effects on growth and change (Bavelas *et al.*, 2000; Pasupathi, 2001). This process resulted in a 138-item pool that was administered to subordinates, recruited via snowball sampling, in various organizations. Out of 1030 volunteers who completed the survey, 977 provided useable responses. The respondents took either the English version ($n = 173$) or the Hebrew version ($n = 804$) of the FLS. The 138 items were subjected to a principal component analysis (PCA) with a promax rotation that yielded 19 components with an eigenvalue exceeding one. After inspecting the factor–structure matrix, only the first nine components were interpretable. Items were assigned to these nine components (and then to a respective subscale) if they had their highest loading on that component and that loading was .40 or higher. When a component had more than 10 items with a high loading, only the 10 strongest items were retained for the final scale.

Both the 19 components (component scores) and the retained 9 components (forced solution) were subjected to a second-order PCA (promax rotation). These two analyses suggested the existence of a single component, based on a screen test, which suggests that people have a general perception of listening (on a continuum from bad to good).

The FLS was originally developed to generate subordinates' perceptions regarding the listening of their supervisor, but it can be used in various contexts.

Table P19.1 Scales from which FLS items were selected.

Scale name	No. of items	Author(s)
Listening Style Profile (LSP-16)	16	Watson, Barker, & Weaver (1995)
Interpersonal Listening in the Personal Selling Context (ILSP)	14	Castleberry & Shepherd (1993)
(Unnamed)	24	Castro (2010)
Active Empathetic Listening (AEL)	47	Drollinger, Comer, & Warrington (2006)
Parent Confirmation Behavior Indicator (PCBI)	28	Ellis (2002)
Active Listening Observation Scale (ALOS-global)	7	Fassaert, Van Dulmen, Schellevis, & Bensing (2007)
Active Listening (AL)	27	Kubota, Mishima, Ikemi, & Nagata (1997)
The Active Listening Attitude Scale (ALAS)	47	Mishima, Kubota, & Nagata (2000)
Listening Style Inventory (LSI)	10	Pearce, Johnson, & Barker (2003)
Salesperson Listening	13	Ramsey & Sohi (1997)

Reliability

The Constructive and Destructive Listening Behavior scales and the Positive Listening Consequences scale, each of which is measured with 10 items, have generated score reliabilities above .90. The shorter subscales have generated reliability estimates between .69 and .84.

Validity

Kluger and Zaidel (2013) provided some support for the existence of separate constructive and destructive listening behavior item clusters. Specifically, they subjected 48 listening items ($N = 238$) to a PCA with a promax rotation. These items included items taken from three sources: the Active Listening Attitude Scale (Mishima *et al.*, 2000; see Profile 3), FLS items that loaded on the constructive and destructive listening behaviors components, and items taken from the Listening Styles Profile (LSP; items were taken from table 1 in Bodie & Worthington, 2010; see Profile 36). This analysis yielded eight components, and the first two components were constructive and destructive listening behaviors. Other components reflected constructs taken from the LSP (e.g., a preference to listen to technical information) and specific FLS behaviors (e.g., escaping; see table 1 in Kluger & Zaidel, 2013). Items that loaded highly on the constructive and destructive listening behavior components in Kluger and Zaidel's work were worded similarly to the items reported by Bouskila-Yam and Kluger (2011).

Kluger and Zaidel (2013) additionally showed that, despite the high correlations of constructive listening with destructive listening ($r = -.65$), a path analysis suggested that the constructive listening scale was the sole predictor of leadership consideration (people-oriented leadership), whereas both constructive and destructive listening scales played a role in predicting leadership facets of inconsideration and initiating structure (task-oriented leadership). These results provide initial support to the convergent and divergent validity of these FLS scales—that listening has two highly correlated, yet not isomorphic, facets: constructive and destructive.

The construct validity of the constructive listening scale received additional support in a study that employed six items taken from the constructive listening scale along with additional theory-based items written by subject matter experts (Schroeder & Bergeron, 2015). An exploratory factor analysis (principle axis extraction) of a final set of 18 items ($N = 567$) yielded a single factor ($\alpha = .97$). The contents of the additional items were very similar to the constructive listening items from the original FLS (e.g., "Demonstrates an understanding of my view" and "Makes an attempt to understand my intention for talking"). Moreover, the Schroeder and Bergeron (2015) listening scale was correlated with perceptions of organizational support, organizational-citizenship behavior toward the organization, helping, employee voice, vitality, and meaning, where correlations ranged between .32 (meaning) and .63 (perceptions of organizational support).

Availability

The FLS is provided at the end of this profile and is free to use for research purposes with appropriate citation.[2]

Sample Studies

The constructive listening scale was given to disputants in an out-of-court mediation (Freres, 2014), who rated the listening of the mediators ($N = 50$). Constructive listening positively correlated with trusting the mediator ($r = .82$), satisfaction with the process ($r = .61$), willingness to recommend that mediator ($r = .56$), fairness ($r = .61$), and disputant voice ($r = .74$); the correlation with reaching an out-of-court settlement (yes/no; $r = .29$) was not statistically significant.

The seven items measuring constructive listening behavior that loaded most highly in Bouskila-Yam and Kluger (2011) were administered to employees ($N = 172$ embedded in 75 teams) who rated the listening of their supervisors, along with their own social anxiety and attitudes toward their supervisor (Itzchakov, 2015). Using hierarchical linear modeling, the constructive listening behavior scale was inversely related to social anxiety ($B = -0.75$, 95% CI [−0.90, −0.60]) and attitude extremity ($B = -1.91$, 95% CI [−2.47, −1.35]) and positively associated with objective-attitude ambivalence ($B = 4.11$, 95% CI [2.93, 5.29]; a measure of awareness of pros and cons in the attitude toward the supervisor). Importantly, these correlational results were replicated when listening was manipulated experimentally. For example, in one experiment, business students were randomly paired either with a trained listener (i.e., a business coach or a social worker) or with a colleague, spoke for

2 The full set of 138 items is available at http://media.wix.com/ugd/b878f6_5fa3afe60f2d42699476c4eb5176203f.docx?dn=FLS%20IOBC%202011%20Final.docx.

12 minutes about their attitude toward becoming a manager, and rated their attitudes after being listened to. Although the context of the attitudes measured in the experiment was different from the context of attitudes toward the supervisor, the results were similar, conferring credibility to the constructive listening behavior scale as a measure that captures listening aspects found among good (i.e., trained) listeners.

Ten items of the constructive listening behavior scale and 16 items from the original Listening Style Profile (LSP-16) (Watson *et al.*, 1995) were adapted to assess participants' views of their ideal listener (Itzchakov, Kluger, Emanuel-Tor, & Koren Gizbar, 2014). PCA with promax rotation ($N = 195$) yielded five components, with all the constructive listening items loading on the first component. Items of the LSP-16 loaded, mostly as expected, on other components. The constructive-listening behavior scale, $\alpha = .93$, correlated most strongly with the People Listening facet of the LSP-16, $r = .70$. It also was correlated with scales of the Big-5 personality inventory: agreeableness ($r = .30$), openness ($r = .25$), conscientiousness ($r = .24$), and extroversion ($r = .18$) (but not with neuroticism, $r = .00$). Moreover, the constructive listening behavior scale showed even a higher correlation ($r = .38$) with the General Factor of Personality (the average of the Big-5 subscales with neuroticism reversed; Musek, 2007).

Critique

The FLS shows promise to capture the essence of the perception of being listened to and offers a distinction between constructive and destructive listening behaviors (as well as consequences). More research is needed, however, (a) to confirm its factor structure using confirmatory factor analysis on new samples; (b) to establish its generalizability beyond the business context and the languages and cultures in which it was tested (Hebrew in Israel and English in the United States); (c) to establish its predictive validity and, especially, the divergent validity of the constructive versus destructive facets of listening; (d) to test its incremental validity relative to other existing listening scales; and (e) to test whether a different method of sampling the universe of perceived listening will yield a similar structure. For example, a new pool of listening items could be created by asking laypeople to offer defining attributes of listening (Bodie, St. Cyr, Pence, Rold, & Honeycutt, 2012) and use these definitions to build a scale (Frei & Shaver, 2002) and compare it to the FLS.

Given that validity information has been reported largely for the constructive listening behavior scale, the nomological network of the other scales is unknown. Thus, the labeling of the FLS scales could be disputed. For example, one could relabel the constructive listening behavior with the label *cognitive listening*, and the positive consequences scale with the label *affective listening* or *supportive listening*.

Moreover, much of the work on the constructive scale has not used the full scale but has measured this listening facet with different items. The degree to which these different versions of the scale are isomorphic and result in similar relationships to important outcomes is an empirical question that needs to be addressed in future work.

Users of the measure should be careful in deciding whether to measure listening behaviors or listening consequences. It is possible to use both scales in the same study, but researchers might be tempted to create a short scale by including items from several subscales. This is likely to yield highly reliable scores that are not necessarily homogeneous.

References

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology, 79*, 941–952. doi:10.1037//0022-3514.79.6.941

Bodie, G. D., St. Cyr, K., Pence, M., Rold, M., & Honeycutt, J. (2012). Listening competence in initial interactions I: Distinguishing between what listening is and what listeners do. *International Journal of Listening, 26*, 1–28. doi:10.1080/10904018.2012.639645

Bodie, G. D., & Worthington, D. (2010). Revisiting the Listening Styles Profile (LSP-16): A confirmatory factor analytic approach to scale validation and reliability estimation. *International Journal of Listening, 24*, 69–88. doi:10.1080/10904011003744516

Bouskila-Yam, O., & Kluger, A. N. (2011, December). The Facilitating Listening Scale (FLS). In A. N. Kluger (Chair), *Listening,* a symposium presented at the 1st Israel Organizational Behavior Conference, Tel Aviv, Israel.

Castleberry, S. B., & Shepherd, C. D. (1993). Effective interpersonal listening and personal selling. *Journal of Personal Selling & Sales Management, 13*, 35–49. doi:10.1080/08853134.1993.10753935

Castro, D. R. (2010). *Listening measures*. Jerusalem: School of Business, The Hebrew University of Jerusalem.

Drollinger, T., Comer, L. B., & Warrington, P. T. (2006). Development and validation of the active empathetic listening scale. *Psychology & Marketing, 23*, 161–180. doi:10.1002/mar.20105

Ellis, K. (2002). Perceived parental confirmation: Development and validation of an instrument. *Southern Communication Journal, 67*, 319–334. doi:10.1080/10417940209373242

Fassaert, T., Van Dulmen, S., Schellevis, F., & Bensing, J. (2007). Active listening in medical consultations: Development of the active listening observation scale (ALOS-global). *Patient Education and Counseling, 68*, 258–264. doi:10.1016/j.pec.2007.06.011

Frei, J. R., & Shaver, P. R. (2002). Respect in close relationships: Prototype definition, self-report assessment, and initial correlates. *Personal Relationships, 9*, 121–139. doi:10.1111/1475-6811.00008

Freres, M. (2014). *Empathetic mediator listening and assertive mediator techniques affect the probability of settlement in small claims mediation* (Unpublished master's thesis). The Chicago School of Professional Psychology, Los Angeles, CA.

Itzchakov, G. (2015, May). If you listen to me, I will change my attitude. In A. N. Kluger (Chair), *Listening: Why should you and why should you not?* Symposium presented at the 30th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Itzchakov, G., Kluger, A. N., Emanuel-Tor, M., & Koren Gizbar, H. (2014). How do you like me to listen to you? *International Journal of Listening, 28*, 177–185. doi:10.1080/10904018.2014.917929

Kluger, A. N., & Zaidel, K. (2013). Are listeners perceived as leaders? *International Journal of Listening, 27*, 73–84. doi:10.1080/10904018.2013.754283

Kubota, S., Mishima, N., Ikemi, A., & Nagata, S. (1997). A research in the effects of active listening on corporate mental health training. *Journal of Occupational Health, 39*, 274–279. doi:10.1539/joh.39.274

Mishima, N., Kubota, S., & Nagata, S. (2000). The development of a questionnaire to assess the attitude of active listening. *Journal of Occupational Health, 42*, 111–118. doi:10.1539/joh.42.111

Musek, J. (2007). A general factor of personality: Evidence for the Big One in the five-factor model. *Journal of Research in Personality*, *41*, 1213–1233. http:// dx.doi.org/10.1016/ j.jrp.2007.02.003
Pasupathi, M. (2001). The social construction of the personal past and its implications for adult development. *Psychological Bulletin*, *127*, 651–672. doi:10.1037//0033-2909.127.5.651
Pearce, C. G., Johnson, I. W., & Barker, R. T. (2003). Assessment of the listening styles inventory: Progress in establishing reliability and validity. *Journal of Business and Technical Communication*, *17*, 84–113. doi:10.1177/1050651902238546
Ramsey, R. P., & Sohi, R. S. (1997). Listening to your customers: The impact of perceived salesperson listening behavior on relationship outcomes. *Journal of the Academy of Marketing Science*, *25*, 127–137. doi:10.1007/BF02894348
Rogers, C. R., & Roethlisberger, F. J. (1991/1952). HBR Classic: Barriers and gateways to communication (Reprinted from *Harvard Business Review*, July–August 1952). *Harvard Business Review*, *69*, 105–111.
Schroeder, T. D., & Bergeron, D. M. (In preparation). *Lending an ear: The impact of listening on employee outcomes*. Department of Organizational Behavior, Weatherhead School of Management. Case Western Reserve University.
Watson, K. W., Barker, L. L., & Weaver, J. B. (1995). The Listening Styles Profile (LSP-16): Development and validation of an instrument to assess four listening styles. *International Journal of Listening*, *9*(1), 1–13. doi:10.1080/10904018.1995.10499138

Scale

The Facilitating Listening Scale (FLS) (Bouskila-Yam & Kluger, 2011)

Source: Bouskila-Yam and Kluger (2011). Reproduced with permission of Kluger.

For each item, we want you to reflect on your experience of being listened to by your supervisor. For each item, indicate your level of agreement/disagreement using the following scale:

1 = Strongly Disagree
2 = Disagree
3 = Somewhat Disagree
4 = Neither Agree nor Disagree
5 = Somewhat Agree
6 = Agree
7 = Strongly Agree

The following items measure perceptions of listening behaviors and should be introduced by the prompt, "When my current supervisor listens to me, most of the time, s/he..."

Constructive listening behaviors

1) Tries hard to understand what I am saying
2) Asks questions that show his/her understanding of my opinions
3) Encourages me to clarify a problem
4) Expresses interest in my stories

5) Listens to me attentively
6) Pays close attention to what I say
7) Gives me time and space to talk
8) Gives me his/her undivided attention
9) Creates a positive atmosphere for me to talk
10) Allows me to express myself fully

<u>Destructive listening behaviors</u>

11) Talks offensively
12) Criticizes my feelings
13) Frowns (showing disapproving facial expressions)
14) Discounts or explains away my feelings
15) Is not willing to listen to me
16) Does not pay attention to things I say
17) Talks back to me aggressively
18) Becomes irritated
19) Is impatient
20) Gets tense

<u>Constructive listening behaviors – reframing</u>

21) Restates what I say
22) Gives me a brief summary of what I have said
23) Completes my sentences to help me clarify what I am saying
24) Asks continuing questions like "Could you tell me more?"

<u>Destructive listening behaviors – domineering</u>

25) Often interrupts me while I am talking
26) Begins to talk before I finish talking
27) Talks more than me
28) Imposes his/her own views
29) Listens to me calmly (reverse scored)
30) Hurries me into talking faster

<u>Destructive listening behaviors – escaping</u>

31) Stares at the computer screen while I'm talking to him/her
32) Uses the telephone while I'm talking to him/her
33) Is distracted while I'm talking

<u>Destructive listening behaviors – no time</u>

34) Begins a discussion by telling me how long s/he has for me
35) Looks at his/her watch or clocks in the room when s/he has limited time to listen to me
36) Hurries me and lets me know that s/he has a limited amount of time to listen

<u>Destructive listening behaviors – changing the subject</u>

37) Starts talking about unrelated issues
38) Changes the subject too frequently

The following items measure perceptions of the consequences of being listened to by one's supervisor (or other relational target). These items should be introduced by the prompt, "When my current supervisor listens to me, most of the time, it makes me..."

<u>Positive listening consequences</u>

1) Enjoy being listened to
2) Feel that s/he cares about me
3) Feel that it is easy for me to open my heart
4) Feel that I am a unique and valuable human being
5) Feel that s/he is interested in me
6) Feel close to him/her
7) Feel comfortable
8) Feel that s/he accepts me for who I am
9) Feel confident

<u>Negative listening consequences</u>

10) Concerned about what s/he thinks of me
11) Worry about myself
12) Aware of my shortcomings (disadvantages)

Note: Subscale identifying labels should be removed and items randomized prior to administration.

Profile 20

Five Facet Mindfulness Questionnaire (FFMQ)

(Baer, Smith, Hopkins, Krietemeyer, & Toney, 2006)

Profiled by: Carly M. Danielson, MA and Susanne M. Jones, PhD

University of Minnesot, Twin Cities

Construct

The Five Facet Mindfulness Questionnaire was designed to measure five components that are theorized to comprise mindfulness: observing, describing, acting with awareness, nonjudging, and nonreacting (Baer, Smith, Hopkins, Krietemeyer, & Toney, 2006).

Instrument Type

Self-Report

Description

The Five Facet Mindfulness Questionnaire (FFMQ) measures five facets of mindfulness: *observing* (attending to internal and external stimuli), *describing* (assigning internal experiences with language), *acting with awareness* (attending to activities in the present moment), *nonjudging of inner experiences* (having nonevaluative positions to thoughts and feelings), and *nonreacting to inner experiences* (not getting caught up or carried away in thoughts and feelings) (Baer *et al.*, 2006). Each facet is captured with seven to eight items, resulting in a 39-item scale.

Administration

The FFMQ is a self-administered questionnaire that takes less than 10 minutes to complete. Participants are prompted to rate 39 statements on the basis of how true each

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

response is for them. Each item is rated on a scale that ranges from 1 (*never or very rarely true*) to 5 (*very often or always true*).

Scoring

After reverse coding 19 items, a mean score is calculated for responses to items of each of the five facets. As discussed further in this profile, we do not recommend calculating a composite score.

Development

The FFMQ was developed to test whether mindfulness is a multifaceted construct and, if so, to determine the characteristics that make up mindfulness (Baer *et al.*, 2006). The items that comprise the scale emerged through factor-analytic distillation of five mindfulness measures, namely the Mindful Attention Awareness Scale (MAAS; Brown & Ryan, 2003), the Freiburg Mindfulness Inventory (FMI; Bucheld, Grossman, & Walach, 2001), the Kentucky Inventory of Mindfulness Skills (KIMS; Baer, Smith, & Allen, 2004), the Cognitive and Affective Mindfulness Scale (CAMS; Feldman, Hayes, Kumar, Greeson, & Laurenceau, 2007), and the Mindfulness Questionnaire (Chadwick, Hember, Mead, Lilley, & Dagnan, 2005). The factor-analytic procedures resulted in five mindfulness facets that converge on an overarching construct, which is theorized to be mindfulness. The FFMQ can be used to measure either dispositional mindfulness or mindful skill differences in response to interventions (e.g., meditation interventions, therapy interventions). The FFMQ has been translated into several languages, including Dutch (Veehof, ten Klooster, Taal, Westerhof, & Bohlmeijer, 2011), German (Tran, Glück, & Nader, 2013), French (Heeren, Douilliez, Peschard, Debrauwere, & Philippot, 2011), Portuguese (Barros, Kozasa, Souza, & Ronzani, 2014), and Chinese (Hou, Wong, Lo, Mak, & Ma, 2014).

Reliability

Generally, the FFMQ generates highly reliable scores among US nonmeditators (.86) and meditators (.95) (Van Dam, Earleywine, & Danoff-Burg, 2009). Other studies report acceptable reliabilities of the FFMQ factors among nonmeditators and meditators in the United States (αs ranging from .75 to .93; Christopher, Neuser, Michael, & Baitmangalkar, 2012; Curtiss & Klemanski, 2014; Greason & Cashwell, 2009) and Dutch samples (αs ranging from .70 to .89; De Bruin, Topper, Muskens, Bogels, & Kamphuis, 2012). Test–retest reliabilities of the FFMQ are also quite good (αs ranging from .71 to .88; Giovannini *et al.*, 2014; Hou *et al.*, 2014).

Three of the five facets of mindfulness have shown good internal consistencies, ranging from .87 to .92 across some studies (Baer *et al.*, 2006, 2008; Veehof *et al.*, 2011). The observe facet has generated reliability estimates of .69 in nonmeditating samples (Veehof *et al.*, 2011). In addition, some studies report that the observe facet and its relation to other facets, as well as outcomes of interest, tends to be moderated by meditation experience, such that meditators report higher observe mean scores than nonmeditators

(Baer *et al.*, 2006, 2008). The observe facet may represent characteristics of other constructs, such as rumination, but further research is needed to test that speculation (Tran *et al.*, 2013). Some studies found lower nonreacting reliability estimates in student ($\alpha = .67$; Baer *et al.*, 2008) and community samples ($\alpha = .69$; Tran *et al.*, 2013), and lower nonreacting factor loadings (<.40) for nonmeditating samples (Veehof *et al.*, 2011). The somewhat lower reliabilities for the observing and nonreacting facets might be due to sample and cultural differences, and further research also needs to examine this claim.

Between-factor correlations tend to be modest, with *r*s ranging from .15 to .40 (Christopher *et al.*, 2012; Van Dam, Hobkirk, Danoff-Burg, & Earleywine, 2012). The FFMQ has been adapted to several contexts and subpopulations (e.g., anxiety disorders, substance abuse, depression). For example, the FFMQ-S is a modified FFMQ that measures sexual difficulties and has generated stable reliabilities for all five facets with αs ranging from .78 to .88 (Adam, Heeren, Day, & De Sutter, 2014).

Validity

The FFMQ has demonstrated strong convergent and divergent validity (Baer *et al.*, 2008, 2011; Christopher *et al.*, 2012). Mindfulness levels, as assessed with this scale, fluctuate substantially with meditation experience; meditators consistently score higher on FFMQ facets than nonmeditators (e.g., Christopher *et al.*, 2012; Van Dam *et al.*, 2009, 2012). At least three studies examined whether these results are truly a function of mindfulness differences that characterize these two groups (Baer, Samuel, & Lykins, 2011; Van Dam *et al.*, 2009, 2012). Two studies in particular tested potential method effects by examining the influence of positive and negative item wording on mean mindfulness levels in meditating and nonmeditating samples. Using differential item functioning (DIF), two studies found that several FFMQ items function differently in meditating and nonmeditating samples (Baer *et al.*, 2011; Van Dam *et al.*, 2009). Specifically, Van Dam *et al.* (2009) found that meditators were more likely to endorse the positively worded FFMQ items, whereas nonmeditators were more likely to reject negatively worded items. Baer *et al.* (2011) obtained similar results, but with a smaller set of FFMQ items; these researchers also used less stringent probability levels than in Van Dam *et al.*

Two additional validity issues concern the construct structure of mindfulness and the actual length of the FFMQ. First, concern exists as to whether the five FFMQ facets truly converge on an overarching mindfulness construct. This issue emerged because Baer *et al.* used item parceling in their original 2006 study to obtain a properly fitting hierarchical model. Parceling is problematic because it can hide the true relationship among scale items and potentially lead to misspecified models (for a full discussion of item parceling, see Little, Cunningham, Shahar, & Widaman, 2002). Some studies have since verified the hierarchical facet structure (e.g., Christopher *et al.*, 2012), whereas other studies suggest that the latent FFMQ construct may consist of a set of intercorrelated subscales, rather than a single hierarchical 5-factor model (e.g., Van Dam *et al.*, 2012).

Second, some studies have found problems of item fit and redundancy, and have tested shortened FFMQ versions. Bohlmeijer, ten Klooster, Fledderus, Veehof, and Baer's (2011) 24-item FFMQ-SF (short form) shows an acceptable model fit for Dutch people with depression (αs ranging from .75 to .87). Tran *et al.* (2013) also tested a

20-item FFMQ-SF using a community sample (αs ranging from .62 to .81) and a student sample (αs ranging from .48 to .79). Nonreacting featured weak psychometric properties in both samples. Additional studies have reported acceptable reliabilities using a short form for diabetics (αs ranging from .71 to .82; Tak, Hendrieckx, Nefs, Nyklíček, Speight, & Pouwer, 2015) and chronic pain sufferers (α = .82, only the total FFMQ scale was reported; Trompetter, Bohlmeijer, Veehof, & Schreurs, 2015).

Availability

The 39-item FFMQ was first published in Baer *et al.* (2006) and is also available at the end of this profile. It is reproduced with permission of Sage Publications, Inc. and free to use for research purposes.

Sample Studies

The FFMQ is widely used to assess individual differences in mindfulness dispositions (e.g., Christopher *et al.*, 2012), as well as differences after mindfulness interventions (Baer, Carmody, & Hunsinger, 2012). The FFMQ has been adapted as a useful measure to assess mindfulness skills in sexual functioning (FFMQ-S; Adam *et al.*, 2014) and is used to predict emotional regulation (Baer *et al.*, 2012; Bohlmeijer *et al.*, 2011) and well-being (Deng, Liu, Rodriguez, & Xia, 2011). The FFMQ has been successfully used to study the outcomes of mindfulness-based interventions for people with generalized anxiety disorder (Fisak & von Lehe, 2012) and substance abuse (Fernandez, Wood, Stein, & Rossi, 2010) or craving (Witkiewitz, Bowen, Douglas, & Hsu, 2013).

Researchers studying outcomes of mindfulness-based interventions have found enhanced listening skills (Schure, Christopher, & Christopher, 2008). Having a heightened ability to attend to and be aware of the present facilitates deeper listening and attunement capabilities (Keane, 2014). Lower scores on the observe facets are associated with more negative emotional regulation strategies, including worrying and rumination (Desrosiers, Vine, Curtiss, & Klemanski, 2014). Mindfulness interventions that cultivate observing and attending skills (the core of effective listening) have been successfully applied in counseling settings (Keane, 2014; Greason & Welfare, 2013), among people suffering from depression and/or anxiety (Desrosiers *et al.*, 2014), as well as among cancer patients (Bränström, Kvillemo, Brandberg, & Moskowitz, 2010). Lastly, mindfulness training assessed by the FFMQ found enhanced abilities for teachers to mindfully listen and attend to their bodily sensations, leading to reduced stress (Frank, Reibel, Broderick, Cantrell, & Metz, 2015).

Critique

The FFMQ is a popular scale with an impressive validity portfolio. Its psychometric properties are routinely updated and explored, and the interested researcher is encouraged to check for new information on this scale before using it. The observing and nonreacting facets have thus far generated the most variable psychometric properties, particularly in nonmeditating samples. Because there are issues with differential item functioning for those who meditate compared to those who do not, and because the five facets are not universally acknowledged to tap an overall mindfulness factor, we recommend utilizing the FFMQ subscales rather than a composite score in particular research situations. If researchers are interested in a unidimensional mindfulness scale, they are referred to the MAAS (Brown & Ryan, 2003). The FFMQ also appears to be influenced by language, culture, and meditation experience (Christopher *et al.*, 2012; Tran *et al.*, 2013). Further examination of the FFMQ facets is needed to tease out the moderating impact of these variables.

References

Adam, F., Heeren, A., Day, J., & De Sutter, P. (2014). Development of the Sexual Five-Facet Mindfulness Questionnaire (FFMQ-S): Validation among a community sample of French-speaking women. *The Journal of Sex Research*. Advance online publication. doi:10.1080/00224499.2014.894490

Baer, R. A., Carmody, J., & Hunsinger, M. (2012). Weekly change in mindfulness and perceived stress in a mindfulness-based stress reduction program. *Journal of Clinical Psychology*, *68*, 755–765. doi:10.1002/jclp.21865

Baer, R. A., Samuel, D. B., & Lykins, E. (2011). Differential item functioning on the Five Facet Mindfulness Questionnaire is minimal in demographically matched meditators and nonmeditators. *Assessment*, *18*, 3–10. doi:10.1177/1073191110392498

Baer, R. A., Smith, G. T., & Allen, K. B. (2004). Assessment of mindfulness by self-report: The Kentucky Inventory of Mindfulness Skills. *Assessment*, *11*, 191–206. doi:10.1177/1073191104268029

Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment*, *13*, 27–45. doi:10.1177/1073191105283504

Baer, R. A., Smith, G. T., Lykins, E., Button, D., Krietemeyer, J., Sauer, S., & Williams, J. M. G. (2008). Construct validity of the Five Facet Mindfulness Questionnaire in meditating and nonmeditating samples. *Assessment*, *15*, 329–342. doi:10.1177/1073191107313003

Barros, V. V., Kozasa, E. H., Souza, I. C. W., & Ronzani, T. M. (2014). Validity evidence of the Brazilian version of the Five Facet Mindfulness Questionnaire (FFMQ). *Psicologia: Teoria e Pesquisa*, *30*, 317–327. doi:10.1590/S0102-37722014000300009

Bohlmeijer, E., ten Klooster, P. M., Fledderus, M., Veehof, M., & Baer, R. (2011). Psychometric properties of the Five Facet Mindfulness Questionnaire in depressed adults and development of a short form. *Assessment*, *18*, 308–320. doi:10.1177/1073191111408231

Bränström, R., Kvillemo, P., Brandberg, Y., & Moskowitz, J. (2010). Self-report mindfulness as a mediator of psychological well-being in a stress reduction intervention for cancer patients—A randomized study. *Annals of Behavioral Medicine*, *39*, 151–161. doi:10.1007/s12160-010-9168-6

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, *84*, 822–848. doi:10.1037/0022-3514.84.4.822

Bucheld, N., Grossman, P., & Walach, H. (2001). Measuring mindfulness in insight meditation (Vipassana) and meditation-based psycho-therapy: The development of the Freiburg Mindfulness Inventory (FMI). *Journal for Meditation and Meditation Research, 1*, 11–34. doi:10.1016/j.paid.2005.11.025

Chadwick, P., Hember, M., Mead, S., Lilley, B., & Dagnan, D. (2005). Responding mindfully to unpleasant thoughts and images: Reliability and validity of the Mindfulness Questionnaire. Unpublished manuscript.

Christopher, M., Neuser, N., Michael, P., & Baitmangalkar, A. (2012). Exploring the psychometric properties of the Five Facet Mindfulness Questionnaire. *Mindfulness, 3*, 124–131. doi:10.1007/s12671-011-0086-x

Curtiss, J., & Klemanski, D. (2014). Factor analysis of the Five Facet Mindfulness Questionnaire in a heterogeneous clinical sample. *Journal of Psychopathology and Behavioral Assessment, 36*, 686–694. doi:10.1007/s10862-014-9429-y

De Bruin, E., Topper, M., Muskens, J. G., Bogles, S. M., & Kamphuis, J. H. (2012). Psychometric properties of the Five Facets Mindfulness Questionnaire (FFMQ) in a meditating and nonmeditating sample. *Assessment, 19*, 187–189. doi:10.1177/1073191112446654

Deng, Y., Liu, X., Rodriguez, M., & Xia, C. (2011). The Five Facet Mindfulness Questionnaire: Psychometric properties of the Chinese version. *Mindfulness, 2*, 123–128. doi:10.1007/s12671-011-0050-9

Desrosiers, A., Vine, V., Curtiss, J., & Klemanski, D. K. (2014). Observing nonreactively: A condition process model linking mindfulness facets, cognitive emotion regulation strategies, and depression and anxiety symptoms. *Journal of Affective Disorders, 165*, 31–37. doi:10.1016/j.jad.2014.04.024

Feldman, G., Hayes, A., Kumar, S., Greeson, J., & Laurenceau, J. (2007). Mindfulness and emotion regulation: The development and initial validation of the cognitive and affective mindfulness scale-revised (CAMS-R). *Journal of Psychopathology and Behavioral Assessment, 29*, 177–190. doi:10.1007/s10862-006-9035-8

Fernandez, A. C., Wood, M. D., Stein, L. A. R., & Rossi, J. S. (2010). Measuring mindfulness and examining its relationship with alcohol use and negative consequences. *Psychology of Addictive Behaviors, 24*, 608–616. doi:10.1037/a0021742

Fisak, B., & von Lehe, A. C. (2012). The relation between the five facets of mindfulness and worry in a non-clinical sample. *Mindfulness, 3*, 15–21. doi:10.1007/s12671-011-0075-0

Frank, J. L., Reibel, D., Broderick, P., Cantrell, T., & Metz, S. (2015). The effectiveness of mindfulness-based stress reduction on educator stress and well-being: Results from a pilot study. *Mindfulness, 6*, 208–216. doi:10.1007/s12671-013-0246-2

Giovannini, C., Giromini, L., Bonalume, L., Tagini, A., Lang, M., & Amadei, G. (2014). The Italian Five Facet Mindfulness Questionnaire: A contribution to its validity and reliability. *Journal of Psychopathology and Behavioral Assessment, 36*, 415–423. doi:10.1007/s10862-013-9403-0

Greason, P. B., & Cashwell, C. S. (2009). Mindfulness and counseling self-efficacy: The mediating role of attention and empathy. *Counselor Education and Supervision, 49*, 2–18. doi:10.1002/j.1556-6978.2009.tb00083.x

Greason, P. B., & Welfare, L. E. (2013). The impact of mindfulness and meditation practice on client perceptions of common therapeutic factors. *Journal of Humanistic Counseling, 52*, 235–254. doi:10.1002/j.2161-1939.2013.00045.x

Herren, A., Douilliez, C., Peschard, V., Debrauwere, L., & Philippot, P. (2011). Cross-cultural validity of the Five Facets Mindfulness Questionnaire: Adaptation and validation in

a French-speaking sample. *European Review of Applied Psychology*, *61*, 147–151. doi:10.1016/j.erap.2011.02.001
Hou, J., Wong, S. Y., Lo, H. H., Mak, W. W., & Ma, H. S. (2014). Validation of a Chinese version of the Five Facet Mindfulness Questionnaire in Hong Kong and development of short form. *Assessment*, *21*, 363–371. doi:10.1177/1073191113485121
Keane, A. (2014). The influence of therapist mindfulness practice on psychotherapeutic work: A mixed-methods study. *Mindfulness*, *5*, 689–703. doi:10.1007/s12671-013-0223-9
Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*, 151–173. doi:10.1207/S15328007SEM0902_1
Schure, M. B., Christopher, J., & Christopher, S. (2008). Mind-body medicine and the art of self-care: Teaching mindfulness to counseling students through yoga, meditation, and qigong. *Journal of Counseling & Development*, *86*, 47–56. doi:10.1002/j.1556-6678.2008.tb00625.x
Tak, S. F., Hendrieckx, C., Nefs, G., Nyklíček, I., Speight, J., & Pouwer, F. (2015). The association between types of eating behavior and dispositional mindfulness in adults with diabetes: Results from Diabetes MILES, The Netherlands. *Appetite*, *87*, 288–295. doi:10.1016/j.appet.2015.01.006
Tran, U. S., Glück, T. M., & Nader, I. W. (2013). Investigating the Five Facet Mindfulness Questionnaire (FFMQ): Construction of a short form and evidence of a two-factor higher order structure of mindfulness. *Journal of Clinical Psychology*, *69*, 951–965. doi:10.1002/jclp.21996
Trompetter, H., Bohlmeijer, E., Veehof, M., & Schreurs, K. (2015). Internet-based guided self-help intervention for chronic pain based on Acceptance and Commitment Therapy: A randomized controlled trial. *Journal of Behavioral Medicine*, *38*, 66–80. doi:10.1007/s10865-014-9579-0
Van Dam, N. T., Earleywine, M., & Danoff-Burg, S. (2009). Differential item function across meditators and non-meditators on the Five Facet Mindfulness Questionnaire. *Personality and Individual Differences*, *47*, 516–521. doi:10.1016/j.paid.2009.05.005
Van Dam, N. T., Hobkirk, A. L., Danoff-Burg, S., & Earleywine, M. (2012). Mind your words: Positive and negative items create method effects on the Five Facet Mindfulness Questionnaire. *Assessment*, *19*, 198–204. doi:10.1177/1073191112438743
Veehof, M. M., ten Klooster, P. M., Taal, E., Westerhof, G. J., & Bohlmeijer, E. T. (2011). Psychometric properties of the Dutch Five Facet Mindfulness Questionnaire (FFMQ) in patients with fibromyalgia. *Clinical Rheumatology*, *30*, 1045–1054. doi:10.1007/s10067-011-1690-9
Witkiewitz, K., Bowen, S., Douglas, H., & Hsu, S. H. (2013). Mindfulness-based relapse prevention for substance craving. *Addictive Behaviors*, *38*, 1563–1571. doi:10.1016/j.addbeh.2012.04.001

Scale

The Five Facet Mindfulness Questionnaire (Baer, Smith, Hopkins, Krietemeyer, & Tooney, 2006)

Source: Baer *et al*. (2006). Reproduced with permission from Sage Publications.

Instructions: Please rate each of the following statements using the scale provided. Mark the number on the scale that best describes your own opinion of what is generally true for you.

1 = Never or very rarely true
2 = Rarely true
3 = Sometimes true
4 = Often true
5 = Very often or always true

<u>Observing</u>

1) When I'm walking, I deliberately notice the sensations of my body moving.
2) When I take a shower or bath, I stay alert to the sensations of water on my body.
3) I notice how foods and drinks affect my thoughts, bodily sensations, and emotions.
4) I pay attention to sensations, such as the wind in my hair or sun on my face.
5) I pay attention to sounds, such as clocks ticking, birds chirping, or cars passing.
6) I notice the smells and aromas of things.
7) I notice visual elements in art or nature, such as colors, shapes, textures, or patterns of light and shadow.
8) I pay attention to how my emotions affect my thoughts and behavior.

<u>Describing</u>

9) I'm good at finding words to describe my feelings.
10) I can easily put my beliefs, opinions, and expectations into words.
11) It's hard for me to find the words to describe what I'm thinking.*
12) I have trouble thinking of the right words to express how I feel about things.*
13) When I have a sensation in my body, it's difficult for me to describe it because I can't find the right words.*
14) Even when I'm feeling terribly upset, I can find a way to put it into words.
15) My natural tendency is to put my experiences into words.
16) I can usually describe how I feel at the moment in considerable detail.

<u>Acting with Awareness</u>

17) When I do things, my mind wanders off and I'm easily distracted.*
18) I don't pay attention to what I'm doing because I'm daydreaming, worrying, or otherwise distracted.*
19) I am easily distracted.*
20) I find it difficult to stay focused on what's happening in the present.*
21) It seems I am "running on automatic" without much awareness of what I'm doing.*
22) I rush through activities without being really attentive to them.*
23) I do jobs or tasks automatically without being aware of what I'm doing.*
24) I find myself doing things without paying attention.*

<u>Nonjudging</u>

25) I criticize myself for having irrational or inappropriate emotions.*
26) I tell myself I shouldn't be feeling the way I'm feeling.*
27) I believe some of my thoughts are abnormal or bad and I shouldn't think that way.*

28) I make judgments about whether my thoughts are good or bad.*
29) I tell myself that I shouldn't be thinking the way I'm thinking.*
30) I think some of my emotions are bad or inappropriate and I shouldn't feel them.*
31) When I have distressing thoughts or images, I judge myself as good or bad, depending what the thought/image is about.*
32) I disapprove of myself when I have irrational ideas.*

<u>Nonreacting</u>

33) I perceive my feelings and emotions without having to react to them.
34) I watch my feelings without getting lost in them.
35) When I have distressing thoughts or images, I "step back" and am aware of the thought or image without getting taken over by it.
36) In difficult situations, I can pause without immediately reacting.
37) When I have distressing thoughts or images, I feel calm soon after.
38) When I have distressing thoughts or images I am able just to notice them without reacting.
39) When I have distressing thoughts or images, I just notice them and let them go.

Note: Labels should be removed and items randomized prior to administration. Items marked with an asterisk (*) should be reverse-scored. The scale instructions and items are reprinted, with permission, from Sage Publications, Inc., in line with STM signatory guidelines.

Profile 21

Functional Magnetic Resonance Imaging (fMRI)

(Belliveau *et al.*, 1991)

Profiled by: Iballa Burunat and Elvira Brattico

University of Jyväskylä, Finland
University of Aarhus, Denmark

Construct

Functional magnetic resonance imaging (fMRI) is a noninvasive technique for measuring and mapping brain activity, which provides high-resolution images of the different brain tissues.

Instrument Type

Cognitive Assessment (Brain activity measure)

Description

Functional magnetic resonance imaging (fMRI) is among the main techniques for understanding the human brain in action. FMRI is a noninvasive neuroimaging method for measuring and mapping brain activity by relating changes in local cerebral blood oxygenation across time with behavioral measures. FMRI measurements are performed by neuroscientists and physicians to measure how different parts of the brain are engaged in critical psychological and behavioral functions, such as movement, perception, learning, and thought. Understanding how different brain regions relate to different processes is a central aim of neuroscience and can be of enormous help clinically, for instance by examining the variation of brain function across normal and ill populations.

FMRI works on the principle that changes in blood flow (which provide oxygen and glucose necessary for supplying energy to active neurons) are closely related to changes in brain activity in response to some task (Crosson *et al.*, 2010). More specifically, fMRI

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

measures the relative amount of oxygenated-hemoglobin to deoxygenated-hemoglobin in the blood, which is interpreted as an indirect measure of neural activity. Animal studies have provided strong evidence in supporting this relationship between neural activity (local field potentials) and measured fMRI signal (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001). For instance, when listening to sounds, blood flow in auditory brain areas increases, which is accompanied by an increase in hemoglobin (the iron-rich protein molecule that transports oxygen within red blood cells). To date, however, the coupling between neural activity and blood flow has not been fully understood.

Administration

An fMRI study involves choices regarding experimental design and the statistical methods to analyze resulting data. The experimental design refers to the temporal organization structure in which participants perform cognitive and/or behavioral tasks during the fMRI experiment (Amaro & Barker, 2006). There is a wide spectrum of possibilities when it comes to fMRI experimental design and options for stimulus presentation. The optimal design will depend on many factors, such as the nature of the task, the signal-to-noise ratio over time, and the specific comparisons made. These factors are directly related to the efficacy of the subsequent statistical analysis; therefore, these should be taken into consideration when deciding the experimental design (Lindquist & Wager, 2014).

The two main fMRI designs are *blocked designs* and *event-related designs*, depending on the way stimuli are presented. In blocked designs, experimental conditions are spaced into intervals or blocks; that is, conditions A and B will be performed during alternating experimental blocks (alternating design). In this way, the researcher can determine the differential activity between the two conditions. Alternatively, A and B can be spaced or connected by a third block C, a control condition (controlled blocked design), thus allowing the identification of brain areas active in response to each condition separately and in response to both conditions. It is important to keep in mind possible fatigue or boredom effects if block lengths are too long. Event-related designs consist of different intervals lengths (from 2–3 seconds up to 20 seconds), and stimuli are presented in random order. This design offers more flexibility and aims to target transient changes in brain activity. Because event-related designs include fewer events that are averaged, it possesses weaker statistical power compared to blocked designs.

Attending to a task, in the absence of performing it, however, is sufficient to elicit sustained brain signals. Evidence has suggested the existence of two types of signals: (a) separate transient responses maybe related to the task and (b) sustained responses related to top-down processes (Chawla, Rees, & Friston, 1999). These responses can be captured using a mixed design, which provides a scheme to disentangle these two types of events: sustained activity throughout the whole task versus transient activity evoked by each single task trial. This is achieved by merging blocked and event-related designs into a mixed design, which combines features from both approaches. In a mixed design, participants are presented with control blocks separated by task blocks, during which items are presented at different temporal intervals, where transient responses are likely to occur (Petersen & Dubis, 2012).

More recently, neuroscientists have attempted to capture and investigate how the brain reacts to real-world phenomena using naturalistic or free-behavior tasks, toward further understanding of how the brain perceives complex, continuous multidimensional stimuli

(Hasson, Nir, Levy, Fuhrmann, & Malach, 2004). A transition to mapping brain activity using naturalistic stimuli has been fostered by the recognition that simple, artificial, unimodal stimuli cannot capture the richness of dynamic, natural phenomena (Maguire, 2012). For the analysis of naturalistic fMRI data, participants' behavioral responses can be recorded subsequent (to, e.g., a passive listening fMRI scanning) to retrospectively analyze their data. Alternatively, patterns in the brain activity can be found by employing computational feature extraction (Alluri *et al.*, 2012) or mathematical algorithms such as multivoxel pattern analysis (MVPA; Norman, Polyn, Detre, & Haxby, 2006), Independent Component Analysis (ICA; McKeown & Sejnowski, 1998), or complex network analysis (Bullmore & Sporns, 2009).

The statistical analysis of fMRI data most commonly aims at (a) localizing brain areas activated by the task of interest, which will help predict psychological states or mental disease; or (b) investigating relationships between brain areas. In the first case, the general linear model (GLM) is the most used statistical approach, which is useful to compare different tasks or events or to map brain activity related to the performance of a task of interest. The analysis is *massive univariate*, which means that each separate GLM analysis is performed at each voxel, first assuming independence between voxels but subsequently dealing with the actual dependency between voxels by means of, for example, random field theory. In the case of studying relations between brain areas, connectivity studies have been of increased interest in recent years. Connectivity studies explore how different brain regions interact and to which extent these interactions depend on not only the experimental tasks (Lindquist & Wager, 2014) but also brain activity during rest. During the so-called resting-state fMRI (rsfMRI), participants' brain responses are acquired while participants do not perform any specific task. These studies are aimed at investigating the co-activation between brain areas during rest, which are believed to reflect functional communication between brain regions (van den Heuvel & Pol, 2010). A search in PubMed identified over 3000 rsfMRI studies.

Development

The foundations and further development leading to fMRI span nearly a century. From the 1920s until the 1940s, physics research laid out the possibility to experimentally manipulate the magnetic properties of atomic nuclei. During the 1970s, the first MR images on biological tissue were created, which, by the 1980s, became clinically prevalent (Huettel, Song, & McCarthy, 2009). In 1948, Kety and Schmidt evidenced that blood flow regulation in the brain responds to oxygen metabolism demands of active neurons, but it was not until 1990 that Ogawa, Lee, Kay, and Tank (1990) suggested using paramagnetic deoxyhemoglobin as a naturally occurring contrast agent for MRI. This realization formed the basis of the blood oxygenation level-dependent (BOLD) contrast (i.e., following blood oxygen changes induced by metabolic demand or blood flow). The discovery that changes in blood oxygenation could be measured opened a new avenue of visualizing the functioning brain by using an endogenous contrast mechanism.

Belliveau *et al.* (1991) were the first to successfully conduct an fMRI experiment. Using a visual stimulus paradigm, they localized increases in blood volume in the primary visual cortex.

Due to its noninvasiveness and availability, fMRI has led the neuroimaging field since the 1990s. The first auditory fMRI studies, including those of language processing, were conducted in the early 1990s, demonstrating the viability of fMRI to replicate PET findings (Price, 2012). Since then, over 2000 studies have been conducted using auditory stimuli.[1] Although the BOLD contrast results in a signal change of about 1%, it is still the foundation for the majority of current fMRI research.

Over the past 20 years, fMRI has been more commonly used to identify regions that become active as a result of engaging in specific activities (specific behaviors or cognitive tasks; Crosson *et al.*, 2010). However, it is increasingly used to explore the intrinsic brain connectivity across different areas or even across all voxels, studying the brain as a complex network. Consistent with the PubMed database, there are, to date, more than 4000 fMRI connectivity studies.

Reliability

Several studies have investigated the degree of reliability of fMRI findings. For instance, Specht, Willmes, Shah, and Jäncke (2003) performed reliability assessment of an fMRI experiment by combining three different reliability methods: (a) voxel intraclass correlation coefficient (ICC), (b) correlation of contrast *t*-values for pairs of activation maps, and (c) proportion of overlapping brain volume. By applying these measures to an attention fMRI experiment, where participants had to either ignore a visual stimulus or attend to it, Specht *et al.* (2003) demonstrated that their three reliability measures could provide a consistent evaluation of the reproducibility of the task. Reliability results indicated consistency for the ICC for regions engaged in visual processing. The current state of the reliability in fMRI findings and the factors affecting it was examined by Bennett and Miller (2010). They provided ways to improve fMRI reliability, such as (a) increasing the signal-to-noise (SNR) ratio and contrast-to-noise (CNR) ratio of the acquisition, (b) reducing intersubject differences in cognitive state, and (c) increasing the statistical power of the experiment. Research by Burunat *et al.* (2016) supported the reliability of previous naturalistic fMRI findings on neural processing of auditory features obtained while participants listened continuously to music. The study stressed the importance to consider replication as a mean to assess the reliability of previous fMRI results, and it exposed the need to carefully adjust the study's methodology, particularly when investigating phenomena with high intersubject variability.

Validity

The fMRI experimental setup is complex, and its in-built limitations constrain ecological validity. For instance, acoustic scanner noise may bias the measured neural responses, thus confounding the results of auditory fMRI experiments, particularly in studies targeting cognitive processes (Novitski *et al.*, 2003). There are ways to improve ecological validity by means of current noise attenuation techniques or by even modifying the configuration of the MR hardware, although there is always an associated cost in terms of resolution, SNR, or motion sensitivity (Moelker & Pattynama, 2003). Studies examining affective cognitive processes have been affected by the loud bursts of scanner noise

1 Based on a search of the PubMed database.

(Skouras, Gray, Critchley, & Koelsch, 2013). In order to overcome the effect of scanner noise on stimulus presentation during the acquisition of fMRI responses, single brain volumes can be acquired following a silent stimulus presentation period. This important methodological advancement is called *sparse-sampling* and allows auditory stimuli to be presented without contamination by acoustic scanner noise (Perrachione & Ghosh, 2013); however, it comes at the cost of ecological validity.

It is also important to note that validity is not guaranteed by reliability. A controlled experiment yields more reliable brain responses than one with increased ecological validity, but this is trading off validity for reliability. Thus, at the expense of decreasing reliability, by using naturalistic fMRI paradigms the researcher may increase validity (Hasson & Honey, 2012; Hasson, Malach, & Heeger, 2010). In the field of auditory neuroscience, validity of fMRI findings of increased ecological value (naturalistic fMRI) was supported by Burunat *et al.* (2016). In summary, it is recommended to construct an experimental design with an appropriate balance between validity and reliability.

Sample Studies

Research using fMRI has made several contributions in the field of listening. FMRI was used by Jäncke, Specht, Shah, and Hugdahl (2003) to measure the brain responses in the context of dichotic listening, where participants were instructed to focus attention to stimuli in the left ear, right ear, or both ears. They observed that various cognitive functions within the dorsal and ventral stream of auditory information processing were used by the dichotic listening task. Their results also supported the modulation of the planum temporale by attentional strategies.[2] Moreover, an association between speech perception and motor system has been evidenced by fMRI experiments. Wilson, Saygin, Sereno, and Iacoboni (2004) investigated the role of motor regions in speech perception in an fMRI study in which participants were requested to passively listen to monosyllables and to produce the same speech sounds. They found a premotor cortical area activated by merely listening to the speech sounds, which supported the view that the motor system is involved in speech perception. This may contribute also to language comprehension, as psycholinguistic theories have proposed. Similarly, Menenti, Gierhan, Segaert and Hagoort (2011) studied whether the speech-production system overlaps with speech comprehension using fMRI. Their results revealed a considerable overlap between brain areas engaged in semantic, lexical, and syntactic processing and those for speaking and listening.

Bartel-Friedrich, Broecker, Knoergen, and Koesling (2010) developed an fMRI test set to improve both diagnostic performance and the monitoring of treatment outcomes of complex central auditory function disorders. Their results on healthy controls constitute a reference for future examinations in children with central auditory processing disorders. Silbert *et al.* (2014) studied language production and comprehension using complex, real-life speech and measured functional connectivity across systems by means of new methodological and analytical tools. Their results showed that the production of a real-life narrative is not localized to the left hemisphere, but it overlaps extensively with the comprehension system, challenging the view of language

2 The planum temporale is the cortical area located just behind the auditory cortex.

production lateralization to the left hemisphere. They found that the production of real-world speech recruited bilateral motor and language brain systems as well as non-linguistic areas. Their findings also challenged a strong version of the motor theory of speech perception (Liberman & Mattingly, 1985) because they observed that speech comprehension did not rely on the articulatory system.

Alluri *et al.* (2012) attempted to map a set of musical features (timbral, tonal, and rhythmic) from a rich music stimulus onto the brain anatomy during continuous, naturalistic listening. They did this by correlating the temporal course of each of the features against the brain responses of the participants and found significant activations, which in part overlapped with previous reported controlled experiments but also extended to other areas, revealing cognitive, motor, and limbic brain regions engaged in acoustic feature processing. More recently, Burunat *et al.* (2015) studied the functional connectivity between brain hemispheres of musicians and nonmusicians as they listened to music. Their findings showed that musicians exhibited more mirror-like brain responses compared to nonmusicians, especially in visual and motor brain networks, and this interhemispheric symmetry was more pronounced in keyboardists than in string players. They concluded that a dependency existed between musical training and functional symmetry: Motor training seems to affect music perception.

Critique

Common criticisms of fMRI come from its limited temporal resolution, which is directly related to the slow blood flow response upon which it depends. Nonetheless, more powerful field strengths and improved surface coils, in addition to advanced statistical methods, are gradually allowing the improvement of both temporal and spatial resolution. Another criticism focuses on the fact that fMRI does not quantitatively measure brain activity but instead hemodynamics, an indirect measure of neuronal activity. How the fMRI signal and the underpinning neural activity relate is still under investigation.

It is also important to note that fMRI constitutes only one among many functional neuroimaging techniques, such as positron emission tomography (PET) and transcranial magnetic stimulation (TMS). FMRI does have specific advantages and disadvantages compared to these other techniques, however. Among its main advantages, fMRI is noninvasive: it does not require surgery or exposure to ionizing radiation. FMRI provides high-resolution images of the different brain tissues, which means high accuracy to locate brain activity with a voxel (i.e., 3D pixels in the brain image) size of around 2 mm^3 on a conventional 1.5–3 Tesla scanner. The stronger the magnet, the higher the possibility for improving the spatial resolution. This allows researchers to measure with high precision the functional activation of subcortical areas that are instead almost invisible to other noninvasive techniques such as magnetoencephalography (MEG), electroencephalography (EEG) or functional near-infrared spectroscopy (fNIRS).

The main disadvantage of fMRI is its lower temporal resolution, usually in the order of 1–3 seconds, compared to techniques like EEG, or MEG, which operate in the order of milliseconds. The *temporal resolution* refers to the temporal accuracy for describing the changes in brain activity. This low temporal resolution comes as a result of the trade-off between spatial and temporal accuracy. FMRI scanners and their maintenance are expensive relative, for instance, to EEG.

Another disadvantage of fMRI of particular relevance to the study of auditory function in the brain is the acoustic noise produced by the scanner. The most recent scanner models produce noise at around 80 dB of loudness. Measures can be taken to reduce the noise, such as noise-reducing headphones and foam around patients' heads. Participants taking part in an fMRI experiment must meet standard fMRI eligibility criteria, including having no ferromagnetic material in their body (even large tattoos of circular shape might cause discomfort due to warming up of the skin area under the tattoo), not being pregnant or breastfeeding (to discard any minimal effects on fetus and newborn), and not being susceptible to claustrophobia.

The fMRI measure, together with other high-temporal-resolution techniques (e.g., EEG or MEG), can yield interesting and meaningful conclusions. For instance, an increasingly widespread approach is combined EEG–fMRI, where both EEG and fMRI measurements are recorded simultaneously. It is important to stress that any fMRI experiment is as good as its hypothesis, design, and analysis, and to always keep in mind what research questions can be addressed by fMRI (Aue, Lavelle, & Cacioppo, 2009).

References

Alluri, V., Toiviainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M., & Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage*, *59*, 3677–3689. doi:10.1016/j.neuroimage.2011.11.019

Amaro, E., & Barker, G. J. (2006). Study design in fMRI: Basic principles. *Brain and Cognition*, *60*, 220–232. doi:10.1016/j.bandc.2005.11.009

Aue, T., Lavelle, L. A., & Cacioppo, J. T. (2009). Great expectations: What can fMRI research tell us about psychological phenomena? *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *73*, 10–16. doi:10.1016/j.ijpsycho.2008.12.017

Bartel-Friedrich, S., Broecker, Y., Knoergen, M., & Koesling, S. (2010). Development of fMRI tests for children with central auditory processing disorders. *In Vivo*, *24*, 201–210. doi:24/2/201 [pii]

Belliveau, J. W., Kennedy, D. N., McKinstry, R. C., Buchbinder, B. R., Weisskoff, R. M., Cohen, M. S., Vevea, J. M., Brady, T. J., & Rosen, B. R. (1991). Functional mapping of the human visual cortex by magnetic resonance imaging. *Science*, *254*, 716–719. doi:10.1126/science.1948051

Bennett, C. M., & Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? *Annals of the New York Academy of Sciences*, *1191*, 133–155. doi:10.1111/j.1749-6632.2010.05446.x

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews: Neuroscience*, *10*, 186–198. doi:10.1038/nrn2575

Burunat, I., Brattico, E., Puoliväli, T., Ristaniemi, T., Sams, M., & Toiviainen, P. (2015). Action in perception: Prominent visuo-motor functional symmetry in musicians during music listening. *PLoS ONE*, *10*, e0138238. doi:10.1371/journal.pone.0138238

Burunat, I., Toiviainen, P., Alluri, V., Bogert, B., Ristaniemi, T., Sams, M., & Brattico, E. (2016). The reliability of continuous brain responses during naturalistic listening to music. *NeuroImage*, *124*, 224–231. doi:10.1016/j.neuroimage.2015.09.005

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience*, *2*, 671–676. doi:10.1038/10230

Crosson, B., Ford, A., McGregor, K. M., Meinzer, M., Cheshkov, S., Li, X., Walker-Bateson, D., & Briggs, R. W. (2010). Functional imaging and related techniques: An introduction for rehabilitation researchers. *Journal of Rehabilitation Research and Development*, *47*, vii–xxxiv. doi:10.1682/JRRD.2010.02.0017

Hasson, U., & Honey, C. J. (2012). Future trends in neuroimaging: Neural processes as expressed within real-life contexts. *NeuroImage*, *62*, 1272–1278. doi:10.1016/j.neuroimage.2012.02.004

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences*, *14*, 40–48. doi:10.1016/j.tics.2009.10.011.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science (*New York, N.Y.*), 303*, 1634–1640. doi:10.1126/science.1089506

Huettel, S. A., Song, A. W., & McCarthy, G. (2009). *Functional magnetic resonance imaging* (2nd ed.). Sunderland, MA: Sinauer Associates.

Jäncke, L., Specht, K., Shah, J. N., & Hugdahl, K. (2003). Focused attention in a simple dichotic listening task: An fMRI experiment. *Cognitive Brain Research*, *16*, 257–266. doi:10.1016/S0926-6410(02)00281-1

Kety, S. S., & Schmidt, C. F. (1948). The nitrous oxide method for the quantitative determination of cerebral blood flow in man: Theory, procedure and normal values. *The Journal of Clinical Investigation*, *27*, 476–483. doi:10.1172/JCI101994

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36. doi:10.1016/0010-0277(85)90021-6

Lindquist, M. A., & Wager, T. D. (2014). Principles of functional magnetic resonance imaging. In *Handbook of neuroimaging data analysis* (pp. 3–48). London: Chapman & Hall, CRC Press.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*, 150–157. doi:10.1038/35084005

Maguire, E. A. (2012). Studying the freely-behaving brain with fMRI. *NeuroImage*, *62*, 1170–1176. doi:10.1016/j.neuroimage.2012.01.009

McKeown, M. J., & Sejnowski, T. J. (1998). Independent component analysis of fMRI data: Examining the assumptions. *Human Brain Mapping*, *6*, 368–372. doi:10.1002/(SICI)1097-0193(1998)6:5/6<368::AID-HBM7>3.0.CO;2-E

Menenti, L., Gierhan, S. M. E., Segaert, K., & Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science: A Journal of the American Psychological Society/APS*, *22*, 1173–1182. doi:10.1177/0956797611418347

Moelker, A., & Pattynama, P. M. T. (2003). Acoustic noise concerns in functional magnetic resonance imaging. *Human Brain Mapping*, *20*, 123–141. doi:10.1002/hbm.10134

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*, 424–430. doi:10.1016/j.tics.2006.07.005

Novitski, N., Anourova, I., Martinkauppi, S., Aronen, H. J., Näätänen, R., & Carlson, S. (2003). Effects of noise from functional magnetic resonance imaging on auditory event-related potentials in working memory task. *NeuroImage*, *20*, 1320–1328. doi:10.1016/S1053-8119(03)00390-2

Ogawa, S., Lee, T., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Science*, *87*, 9868–9872. doi:10.1073/pnas.87.24.9868

Perrachione, T. K., & Ghosh, S. S. (2013). Optimized design and analysis of sparse-sampling fMRI experiments. *Frontiers in Neuroscience*, *7*, 1–18. doi:10.3389/fnins.2013.00055

Petersen, S., & Dubis, J. (2012). The mixed block/event-related design. *NeuroImage*, *62*, 1177–1184. doi:10.1016/j.neuroimage.2011.09.084.

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, *62*, 816–847. doi:10.1016/j.neuroimage.2012.04.062

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., and Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences of the USA*, *111*, E4687–E4696.

Skouras, S., Gray, M., Critchley, H., & Koelsch, S. (2013). fMRI scanner noise interaction with affective neural processes. *PLoS ONE*, *8*, 1–13. doi:10.1371/journal.pone.0080564

Specht, K., Willmes, K., Shah, N. J., & Jäncke, L. (2003). Assessment of reliability in functional imaging studies. *Journal of Magnetic Resonance Imaging: JMRI*, *17*, 463–471. doi:10.1002/jmri.10277

van den Heuvel, M. P., & Pol, H. E. H. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, *20*, 519–534. doi:10.1016/j.euroneuro.2010.03.008

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*, 701–702. doi:10.1038/nn1263

Profile 22

Highly Sensitive Person Scale (HSPS)

(Aron & Aron, 1997)

Profiled by: Christopher C. Gearhart, PhD

Tarleton State University

Construct

The Highly Sensitive Person Scale (HSPS) was designed to measure individual differences in sensory-processing sensitivity (SPS) or the sensitivity to physical, emotional, and social stimuli.

Instrument Type

Self-Report

Description

The HSPS measures individual differences in temperament associated with sensitivity to emotional, physical, and social stimuli (Aron & Aron, 1997). The scale operationalizes the construct of sensory-processing sensitivity (SPS), which reflects an individual's tendency to more thoroughly process sensory data in the environment. Subsequent investigation has identified a three-factor model of the HSPS that includes subscales of Ease of Excitation (EOE), or becoming mentally overwhelmed by external and internal demands; Aesthetic Sensitivity (AES), which captures aesthetic awareness and appreciation; and Low Sensory Threshold (LST), related to unpleasant sensory arousal to external stimuli (Smolewska, McCabe, & Woody, 2006).

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

The HSPS is a self-report questionnaire that takes approximately 5 minutes to complete. Respondents answer a series of questions, indicating how much the situation described in each applies to them, using a 7-point scale ranging from 1 (*not at all*) to 7 (*extremely*).

Scoring

For unidimensional scoring, numerical responses to all items are summed and then divided by 27 to create a composite HSPS score. For the multidimensional model, there are 25 items that measure three factors. Scale items and score calculations associated with each subfactor are presented at the end of this profile.

There is a lack of clarity with regard to exactly how to determine whether a respondent has scored as a highly sensitive person or not; sample distributions of HSPS scores are often nonnormal (Aron, Aron, & Jagiellowicz, 2012). Because it is measured as a continuous variable but individuals are thought to either be highly sensitive or not, scores from the HSPS must be dichotomized. Aron and Aron (1997), treating the HSPS as unidimensional, initially defined highly sensitive persons as those within the highest scoring quarter of the sample. Aron *et al.* (2012) indicated that scores are distributed such that there is a "break point" whereby the "curve" is flattened. They suggested that between 10% and 35% of respondents fall into the highly sensitive category. Using this procedure, for instance, Gearhart (2014) dichotomized scores at approximately +1 SD, or the highest scoring 15% of his sample of US undergraduate students.

Development

Elaine Aron defined SPS as "sensitivity to both internal and external stimuli, including social and emotional cues" (Aron *et al.*, 2010, p. 220), reflecting deeper cognitive processing of sensory information. SPS does not reflect a difference in the abilities of sensory organs themselves but a neurological difference in how individuals process stimuli in their environments. To wit, higher self-report scores on a measure of SPS have been associated with greater neurological activity and functioning, as measured via functional magnetic resonance imaging (Jagiellowicz *et al.*, 2011; see fMRI, Profile 21).

In the development of the scale, Aron and colleagues recruited people who were "easily overwhelmed by stimuli (such as noisy places or evocative or shocking entertainment)" (Aron *et al.*, 2012, p. 272) to participate in 3-hour interviews about their sensitivity attributes. From these interviews, a 60-item questionnaire was created, which ranged far beyond being easily overwhelmed by stimulation. A narrowed set of 27 items was administered to 604 undergraduate psychology students at different US universities and an American community sample of 301 obtained using random digit dialing. Reliability estimates (Cronbach's α) from these data ranged from .64 to .87.

Although the 27 items that comprise the scale were originally considered as a single factor, the scale authors conceded that several subfactors may exist in the scale.

Smolewska *et al.* (2006) proposed a three-factor model of the HSPS: 12 items measuring EOE, 7 items measuring AES, and 6 items measuring LST. In this conceptualization, two items were removed from the original scale (see Scale section). The three-factor model has demonstrated adequate fit using confirmatory factor analyses (Evans & Rothbart, 2008; Liss, Mailloux, & Erchull, 2008; Smolewska *et al.*, 2006).

Reliability

During the construction of the HSPS, two studies with independent samples were conducted on the final set of 27 items to estimate score reliability. Studies 6 and 7 demonstrated reliability that was .87 and .85, respectively. Subsequent studies (e.g., Benham, 2006; Evans & Rothbart, 2008; Hofmann & Bitran, 2007; Meyer, Ajchenbrenner, & Bowles, 2005; Meyer & Carver, 2000) have reported alphas of .85 or higher when using all items.

Reliability estimates measured via Cronbach's alpha for the three factors have ranged across studies (Ahadi & Basharpoor, 2010; Evers, Rasche, & Schabracq, 2008; Gerstenberg, 2012; Liss *et al.*, 2008): EOE (.74–.87), LST (.73–.83), and AES (.60–.81). AES tends to demonstrate lower internal consistency scores, which may be considered a limitation.

Validity

Much of the research investigating the validity profile of the HSPS has utilized the unidimensional structure and provided evidence that SPS is related to theoretically relevant constructs, such as introversion, neuroticism, and Behavioral Inhibition System (BIS) functioning (Aron & Aron, 1997; Smolewska *et al.*, 2006). In their initial conceptualization of the HSPS, Aron and Aron (1997) provided evidence of discriminant, convergent, and overall construct validity supported by a set of seven studies. They indicated that scores on the unidimensional HSPS were strongly associated with social introversion (Mdn $r = .29$) and emotionality (Mdn $r = .54$). As expected, the HSPS was correlated, $r = .62$, with Mehrabian's (1976) measure of low sensory screening. Similar to introversion, Gearhart and Bodie (2012) explored the associations of the 27-item HSPS with the four factors generated by the Personal Report of Communication Apprehension (PRCA-24). Highly sensitive persons reported higher levels of apprehension across all four domains (public speaking, group, dyadic, and meeting), $.23 < r < .33$.

Correlations between the HSPS and measures of Big Five personality characteristics were also identified at the magnitude of $r = .45$ (neuroticism) and $r = -.09$ (extraversion) by Smolewska *et al.* (2006). Additionally, Smolewska *et al.* identified a positive, moderate relationship, $r = .32$, between SPS and temperament as measured by BIS scores, and regression analysis identified BIS scores as a significant predictor of self-perceived SPS. Motivation to behave in a cautious manner in order to prevent negative consequences and unpleasant states (BIS reactivity) appears to have the clearest link with SPS.

Concerning the three-dimensional model, several studies have found adequate fit statistics using confirmatory factor analysis (Evans & Rothbart, 2008; Liss *et al.*, 2008; Smolewska *et al.*, 2006), supporting construct validity. Smolewska *et al.* (2006) provided evidence that subfactors of the HSPS also relate to various individual personality and temperament features. All three subfactors (EOE, LST, and AES) were found to have similar associations as the unidimensional model with the criterion variables of interest: positive relationships with neuroticism, negative or no relationships with extraversion, and positive relationships with BIS functioning. Similarly, Gerstenberg (2012) measured personality using the NEO inventory and identified negative relations between all three HSPS subfactors and extraversion as well as positive relations between all subfactors and neuroticism. Liss *et al.* (2008) identified positive associations between LST and EOE (but not AES) and poor social skills and poor communication. Results of these studies suggest that the three factors of the HSPS relate similarly as the unidimensional model to personality variables, thereby offering evidence of validity for the three-factor model.

Availability

The original, 27-item version can be found in Aron and Aron (1997) and is provided at the end of this profile, with permission. The items comprising the three-factor model proposed by Smolewska *et al.* (2006) are identified as well. The scale is free to use for research purposes.

Sample Studies

For a thorough review of research implementing the HSPS, including neurological studies utilizing fMRI scans, see Aron *et al.* (2012). Regarding social and physiological outcomes, studies have found consistent support for the influence of SPS on anxiety and stress. Gearhart and Bodie (2012) found that American college students reporting higher levels of SPS were more likely to experience higher levels of communication apprehension and college stress. Interestingly, items regarding a tendency to become overwhelmed by sensory stimuli were most significantly related to self-reported stress. These results are similar to those identified by Evers *et al.* (2008), who examined associations between SPS and workplace stress in the United States, and are comparable to Benham's (2006) analysis of physical stress and SPS. They also are in line with research by Liss and colleagues who have identified relations among SPS, anxiety, and depression in American samples (Liss *et al.*, 2008; Liss, Timmel, Baxley, & Killingsworth, 2005). Lastly, the measure has been previously used in listening research, which questioned whether individuals with higher SPS were more likely to commit errors in nonverbal decoding (Gearhart, 2014). This belief was not supported by the data.

Critique

Unfortunately, as discussed above, the question factor structure stability makes any research using the HSPS questionable. Given issues regarding the structure of the scale (i.e., uni- vs. multidimensional), its use as an instrument to measure sensory processing is potentially flawed. Indeed, authors of the original scale indicated that a revision of the instrument may be appropriate (Aron *et al.*, 2012). Despite several studies that provide evidence of validity for both conceptualizations, the scale's authors noted that the measure “may not capture enough behaviors directly reflecting depth of processing” (Aron *et al.*, 2012, p. 277). The authors recommended using additional measures of neuroticism along with the HSPS in order to partial out any potential spurious influence of neuroticism (Aron *et al.*, 2012).

Treating a complex, multidimensional temperament characteristic such as sensitivity as a single construct is problematic. Even today, sensitivity researchers debate what characteristics make up sensitivity (e.g., Evans & Rothbart, 2008). Thus, some suggest that the HSPS measure should be considered as multidimensional (Smolewska *et al.*, 2006). When using the three-factor model, however, reliability estimates for the AES subfactor are regularly borderline acceptable. Based on this profile, researchers, educators, and trainers should be aware of structural and reliability issues when using the HSPS.

References

Ahadi, B., & Basharpoor, S. (2010). Relationship between sensory-processing sensitivity, personality dimensions, and mental health. *Journal of Applied Sciences*, *10*, 570–574. doi:10.3923/jas.2010.570.574

Aron, E. N., & Aron, A. (1997). Sensory-processing sensitivity and its relation to introversion and emotionality. *Journal of Personality and Social Psychology*, *73*, 345–368. doi:10.1037/0022-3514.73.2.345

Aron, E. N., Aron, A., & Jagiellowicz, J. (2012). Sensory processing sensitivity: A review in the light of the evolution of biological responsivity. *Personality and Social Psychology Review*, *16*, 262–282. doi:10.1177/108886831143421

Aron, A., Ketay, S., Hedden, T., Aron, E. N., Markus, H. R., & Gabrieli, J. D. E. (2010). Temperament trait of sensory processing sensitivity moderates cultural differences in neural response. *Social Cognitive and Affective Neuroscience*, *5*, 219–226. doi:10.1093/scan/nsq028

Benham, G. (2006). The highly sensitive person: Stress and physical symptom reports. *Personality and Individual Differences*, *40*, 1433–1440. doi:10.1016/j.paid.2005.11.021 1434.

Evans, D. E., & Rothbart, M. K. (2008). Temperamental sensitivity: Two constructs or one? *Personality and Individual Differences*, *44*, 108–118. doi:0.1016/j.paid.2007.07.016

Evers, A., Rasche, J., & Schabracq, M. J. (2008). High sensory-processing at work. *International Journal of Stress Management*, *15*, 189–198. doi:10.1037/1072-5245.15.2.189

Gearhart, C. C. (2014). Sensory-processing sensitivity and nonverbal decoding: The effect on listening ability and accuracy. *International Journal of Listening*, *28*, 98–111. doi:10.1080/10904018.2014.880867

Gearhart, C. C., & Bodie, G. D. (2012). Sensory-processing sensitivity and communication apprehension: Dual influences on self-reported stress in a college student sample. *Communication Reports*, *25*, 27–39. doi:10.1080/08934215.2012.672216

Gerstenberg, F. X. (2012). Sensory-processing sensitivity predicts performance on a visual search task followed by an increase in perceived stress. *Personality and Individual Differences*, *53*, 496–500. doi: 10.1016/j.paid.2012.04.019

Hofmann, S. G., & Bitran, S. (2007). Sensory-processing sensitivity in social anxiety disorder: Relationship to harm avoidance and diagnostic subtypes. *Journal of Anxiety Disorders*, *21*, 944–954. doi:10.1016/j.janxdis.2006.12.003

Jagiellowicz, J., Xu, X., Aron, A., Aron, E. N., Cao, G., Feng, T., & Weng, X. (2011). Sensory-processing sensitivity and neural responses to changes in visual scenes. *Social Cognitive and Affective Neuroscience*, *6*, 38–47. doi:1093/scan/nsq001

Liss, M., Mailloux, J., & Erchull, M. J. (2008). The relationship between sensory processing sensitivity, alexithymia, autism, depression, and anxiety. *Personality and Individual Differences*, *45*, 255–259. doi:10.1016/j.paid.2008.04.009

Liss, M., Timmel, L., Baxley, K., & Killingsworth, P. (2005). Sensory-processing sensitivity and its relation to parental bonding, anxiety, and depression. *Personality and Individual Differences*, *39*, 1429–1439. doi:10.1016/j.paid.2005.05.007

Mehrabian, A. (1976). *Manual for the questionnaire measure of stimulus screening and arousability*. Los Angeles, CA: UCLA.

Meyer, B., Ajchenbrenner, M., & Bowles, D. P. (2005). Sensory sensitivity, attachment experiences, and rejection responses among adults with borderline and avoidant features. *Journal of Personality Disorders*, *19*, 641–658. doi:10.1521/pedi.2005.19.6.641

Meyer, B., & Carver, C. S. (2000). Negative childhood accounts, sensitivity, and pessimism: A study of avoidant personality disorder features in college students. *Journal of Personality Disorders*,*14*, 233–248. doi:10.1521/pedi.2000.14.3.233

Smolewska, K., McCabe, S., & Woody, E. (2006). A psychometric evaluation of the Highly Sensitive Person Scale: The components of sensory-processing sensitivity and their relation to the BIS/BAS and "Big Five." *Personality and Individual Differences*, *40*, 1269–1279. doi:10.1016/j.paid.2005.09.022

Scale

The Highly Sensitive Person Scale (HSPS) (Aron & Aron, 1997)

Respondent directions: Please indicate the degree to which you agree that each question describes you in general. Respond by using the following scale ranging from 1 (*not at all*) to 7 (*extremely*). Do not think too much about any one item and work quickly by giving your first impression.

1) Are you easily overwhelmed by strong sensory input?
2) Do you seem to be aware of subtleties in your environment?
3) Do other people's moods affect you?
4) Do you tend to be more sensitive to pain?
5) Do you find yourself needing to withdraw during busy days, into bed or into a darkened room or any place where you can have some privacy and relief from stimulation?
6) Are you particularly sensitive to the effects of caffeine?
7) Are you easily overwhelmed by things like bright lights, strong smells, coarse fabrics, or sirens close by?
8) Do you have a rich, complex inner life?
9) Are you made uncomfortable by loud noises?
10) Are you deeply moved by the arts or music?
11) Does your nervous system sometimes feel so frazzled that you have to get off by yourself?

12) Are you conscientious?
13) Do you startle easily?
14) Do you get rattled when you have a lot to do in a short amount of time?
15) When people are uncomfortable in a physical environment do you tend to know what needs to be done to make it more comfortable (like changing the lighting or the seating)?
16) Are you annoyed when people try to get you to do too many things at once?
17) Do you try hard to avoid making mistakes or forgetting things?
18) Do you make a point to avoid violent movies and TV shows?
19) Do you become unpleasantly aroused when a lot is going on around you?
20) Does being very hungry create a strong reaction in you, disrupting your concentration or mood?
21) Do changes in your life shake you up?
22) Do you notice and enjoy delicate or fine scents, tastes, sounds, works of art?
23) Do you find it unpleasant to have a lot going on at once?
24) Do you make it a high priority to arrange your life to avoid upsetting or overwhelming situations?
25) Are you bothered by intense stimuli, like loud noises or chaotic scenes?
26) When you must compete or be observed while performing a task, do you become so nervous or shaky that you do much worse than you would otherwise?
27) When you were a child, did your parents or teachers seem to see you as sensitive or shy?

Note: AES = Aesthetic Sensitivity; EOE = Ease of Excitation; LST = Low-sensory Threshold.

Researcher instructions:
To score each subfactor, use the following items:
EOE: 3, 4, 13, 14, 16, 17, 20, 21, 23, 24, 26, 27 (sum and divide by 12)
LST: 6, 7, 9, 18, 19, 25 (sum and divide by 6)
AES: 2, 5, 8, 10, 12, 15, 22 (sum and divide by 7)

Profile 23

HURIER Listening Profile

(Brownell, 1996)

Profiled by: Margarete Imhof, PhD

Gutenberg University Mainz, Germany

Construct

The letters in HURIER represent six interrelated listening subskills: Hearing, Understanding, Remembering, Interpreting, Evaluating, and Responding (Brownell, 1996, pp. 71ff.).

Instrument Type

Self-Report

Description

The HURIER Listening Profile (HLP) covers six components of the listening process.

- **Hearing**: Use attention to perceive, discriminate, and identify sounds; adjust focus for attention.
- **Understanding**: Comprehend information; ask clarifying questions; take notes.
- **Remembering**: Retain and recall information; use memory strategies.
- **Interpreting**: Take into account the communication context; see the speaker's perspective.
- **Evaluating**: Use principles of logic and reasoning; recognize bias and preconceived opinions.
- **Responding**: Analyze the communication situation and choose an appropriate response.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Six items measure both mental and overt behavior associated with each component. Test takers respond to the statements by specifying on a scale from 1 (*almost never*) to 5 (*almost always*) how they perceive their own listening behavior. Individual scores are calculated for each component and can range from 6 to 30. Results are used to identify resources and potentially problematic areas for prospective skill building.

Administration

The questionnaire is self-paced and is typically taken as a paper–pencil test. Test takers are instructed to think of a specific listening context and to answer all questions with this particular context in mind. Brownell suggested that, prior to responding to profile items, test takers receive some basic information about the listening components and begin by making a prediction of their strongest and weakest set of listening skills (see scale instructions below). This first estimate could be used as a backdrop for interpretation of the more detailed questionnaire results.

Scoring

Test takers are guided through a self-scoring process. Total scores for each component are calculated by summing values assigned to each of the six items pertaining to a subscale. Test takers are informed about their skill level in each of the listening components. In addition, they are asked to rank order the subscales from weakest to strongest based on total scores (see Scale section below).

Brownell proposed the following tentative ranges for interpretation of the scores:

25–30 points: You see yourself as an excellent listener.
20–25 points: You believe you are a good listener.
15–20 points: You consider your listening skills adequate.
10–15 points: You perceive some problems in your listening behavior.

Development

The HLP was developed by Brownell and based on her six-component model of listening. The HLP is based on a behavioral approach to listening and identifies mental and overt actions believed to be relevant for effective listening, primarily in the service industries (Brownell, 1994a, 1994b, 2009, 2010). The author based both the model and the questionnaire items on a comprehensive review of the listening literature and of listening tests (e.g., Purdy, 1996; Watson, Barker, & Weaver, 1995; Witkin & Trochim, 1997; Wolvin & Coakley, 1996). While the items have face validity, it is unclear how the items were written and piloted. Brownell offered a somewhat different selection of items in a 2010 publication (pp. 154f.). Brownell emphasized that the exact wording of the questions is of minor importance, because the goal of the instrument is to guide the learner through listening skill development with regard to the six components of listening (J. Brownell, personal communication, June 2015).

Reliability

The only publicly available report of item score reliability was published by Zohoori (2013), who used the HLP in a study comparing Iranian and US students. Zohoori (2013) reported a Cronbach's α for the total scale as .86. Computing internal consistency for the total score, however, assumes a measurement model that has yet to be supported empirically.

Validity

Face validity of the HLP is argued on the basis that the author derived the self-assessment items from the theoretical model of listening, which she uses in her listening course. To date, no validity studies have been published to investigate if the theoretically assumed factors hold up to empirical data.

Availability

The HLP is available in Brownell's textbook (2015) *Listening: Attitudes, Principles, and Skills*, which is in its fifth edition. It is also presented below, along with self-scoring instructions, with permission. Interested users should contact Brownell for permission to employ the scale.

Sample Studies

To date, the HLP has been utilized in one published empirical study. Zohoori (2013) investigated culture specific differences in the self-assessment of listening behavior between students from Iran and the United States. When scores were compared between the two groups, significant differences appeared in the components of hearing, remembering, and responding with US students scoring statistically higher than Iranian students. Effect sizes were small but consistent at about one third of a standard deviation. However, Zohoori does not appear to have tested the factor structure of the measure, nor are reliability estimates for subscales provided. Because individual scale reliabilities are not provided, it is hard to estimate how robust these results are. Moreover, measurement invariance is not supported, placing into question whether mean differences are accurate.

Critique

The HLP is a self-assessment questionnaire that stimulates learners to reflect on their listening behavior within six components of listening. It helps students understand and analyze their listening habits and prospective learning needs (Brownell, 1994b). As such, it is an intuitive, heuristic instrument to introduce learners into the field of listening and listening behavior. However, because the instrument relies on self-report, it is recommended to complement the assessment either by objective data from behavior observation or by a second-party report (Janusik, 2004).

In addition, caution is advised as the validity portfolio is lacking. Thus, more rigorous validation studies are encouraged to better understand the instrument and its implications. Although Brownell suggested categories from *excellent* to *problematic* for interpreting results, there is no evidence-based guideline for evaluating individual test scores, because empirically based norms have not been established. In addition, it is not clear why the categories overlap and why the values below 10 (6–9) have been left out. In a similar vein, the meaning of the information on the rank order of components is unclear. It remains an open question if all components are indeed of equal importance or if weakness in one can be compensated by strength in another.

The HLP could be useful, for instance, if instructors want to start a discussion on listening skills and behavior. Because psychometric characteristics have not been investigated, it is not safe to assume that the questionnaire would yield reliable test–retest values to measure skill development.

The instrument is consistent with a behavioral approach to listening and explicates the patterns of behavior that are typically involved in listening. The author drew upon the listening literature in formulating scale items, and, as a consequence, the composition of the instrument can be accredited with face validity; however, additional studies to explore and confirm the factor structure as well as a rigorous item analysis are needed. If validity evidence accumulates, the HLP could, ideally, be used for measuring the level of listening skills and the rate of skill development.

References

Brownell, J. (1994a). Managerial listening and career development in the hospitality industry. *International Journal of Listening*, *8*, 31–49. doi:10.1080/10904018.1994.10499130

Brownell, J. (1994b). Teaching listening: Some thoughts on the behavioral approach. *Business Communication Quarterly*, *57*(4), 19–26. doi:10.1177/108056999405700404

Brownell, J. (1996). *Listening: Attitudes, principles, and skills*. Boston: Allyn & Bacon.

Brownell, J. (2002). *Listening: Attitudes, principles, and skills* (2nd ed.). Boston: Allyn & Bacon.

Brownell, J. (2009). Fostering service excellence through listening: What hospitality managers need to know. *Cornell Hospitality Report*, *9*(6), 4–18. Retrieved February 3, 2016, from http://scholarship.sha.cornell.edu/chrpubs/8/

Brownell, J. (2010). The skills of listening-centered communication. In A. D. Wolvin (Ed.), *Dimensions of listening* (pp. 141–157). Oxford: Blackwell.

Brownell, J. (2015). *Listening: Attitudes, principles, and skills* (5th ed.). Boston: Allyn & Bacon.

Janusik, L. A. (2004). *Researching listening from the inside out: The relationship between conversational listening span and perceived communicative competence* (Unpublished doctoral dissertation). University of Maryland, College Park. http://hdl.handle.net/1903/1417

Purdy, M. (1996). What is listening? In M. Purdy & D. Borisoff (Eds.), *Listening in everyday life: A personal and professional approach* (2nd ed., pp. 1–20). Lanham, MD: University Press of America.

Watson, K. W., Barker, L. L., & Weaver, J. B., III. (1995). The listening style profile (LSP-16): Development and validation of an instrument to assess four listening styles. *International Journal of Listening*, *9*, 1–13. doi:10.1080/10904018.1995.10499138

Witkin, B. R., & Trochim, W. W. (1997). Toward a synthesis of listening constructs: A concept map analysis. *International Journal of Listening, 11*, 69–87. doi:10.1207/s1932586xijl1101_5
Wolvin, A. D., & Coakley, C. G. (1996). *Listening*. Madison, WI: Brown/Benchmark
Zohoori, A. (2013). A cross-cultural comparison of the HURIER listening profile among Iranian and US students. *International Journal of Listening, 27*, 50–60. doi:10.1080/10904018.2013.732415

Scale

HURIER Listening Profile (Brownell, 1996)

Source: Adapted from Brownell (2002). Reproduced with permission of Judie Brownell.

Instructions: Complete the listening questionnaire on the following pages. Each question corresponds with one of the six listening components you learned about in Chapter 1: Hearing, Understanding, Remembering, Interpreting, Evaluating, and Responding.[1]

> It might be fun, before you go any further, to guess how you will do.
> I think I will score highest on the component of ___________________.
> I will probably score lowest on the component of __________________.

Now, respond to each of the following questions concerning your perceptions of your listening behavior. Write the appropriate number in the blank to your left, using the key below. Unless your instructor gives you other directions, choose one specific listening context and answer all questions with that situation in mind. This will help you be more consistent in your responses.

Key:
5 = almost always
4 = usually
3 = sometimes
2 = infrequently
1 = almost never

_________ 1. I am constantly aware that people and circumstances change over time.
_________ 2. I take into account the speaker's personal and cultural perspective when listening to him.
_________ 3. I pay attention to the important things going on around me.
_________ 4. I accurately hear what is said to me.
_________ 5. I understand my partner's vocabulary and recognize that my understanding of a word is likely to be somewhat different from the speaker's.
_________ 6. I adapt my response according to the needs of the particular situation.
_________ 7. I easily follow conversations and can accurately recall which member contributed which ideas in small group discussions.

1 Chapter 1 refers to material covered earlier in the Brownell textbook from which the scale is drawn.

__________ 8. I consider my partner's personal expertise on the subjects when she tries to convince me to do something.

__________ 9. I do not let my emotions interfere with my listening or decision making.

__________ 10. I can remember what the instructor has said in class even when it's not in the book.

__________ 11. I recognize my "hot buttons" and don't let them influence my listening.

__________ 12. I take into account the person's motives, expectations, and needs when determining the meaning of a message.

__________ 13. I provide clear and direct feedback to others.

__________ 14. I let the speaker know immediately that he has been understood.

__________ 15. I overcome distractions such as the conversation of others, background noises, and telephones, when someone is speaking.

__________ 16. I enter communication situations with a positive attitude.

__________ 17. I am sensitive to the speaker's tone of voice in communication situations.

__________ 18. I listen to and accurately remember what my partner says, even when I strongly disagree with her viewpoint.

__________ 19. I encourage information sharing by creating a climate of trust and support.

__________ 20. I concentrate on what the speaker is saying, even when the information is complicated.

__________ 21. I consider how the speaker's facial expressions, body posture, and other nonverbal behaviors relate to the verbal message.

__________ 22. I weigh all evidence before making a decision.

__________ 23. I take time to analyze the validity of my partner's reasoning before arriving at my own conclusions.

__________ 24. I am relaxed and focused in important communication situations.

__________ 25. I listen to the entire message without interrupting.

__________ 26. I make sure that the physical environment encourages effective listening.

__________ 27. I recognize and take into account personal and cultural differences in the use of time and space that may influence listening effectiveness.

__________ 28. I ask relevant questions and restate my perception to make sure I have understood the speaker correctly.

__________ 29. I listen carefully to determine whether the speaker has solid facts and evidence or whether he is relying on emotional appeals.

__________ 30. I am sensitive to my partner's feelings in communication situations.

__________ 31. I have a wide variety of interests, which helps me approach tasks creatively.

__________ 32. I distinguish between main ideas and supporting evidence when I listen.

__________ 33. I am ready to focus my attention when a presenter begins her talk.

__________ 34. I readily consider new evidence and circumstances that might prompt me to reevaluate my previous position.

__________ 35. I can recall what I have heard, even when I am in stressful situations.

__________ 36. I take notes effectively when I believe it will enhance my listening.

Scoring

Transfer your self-ratings for each question to the corresponding question numbers below. Total the points you assigned for each of the six sets of questions. Place your total for each component in the Total space.		
Hearing 4 ________ 15 ________ 16 ________ 20 ________ 24 ________ 33 ________ Total ________	Understanding 5 ________ 11 ________ 25 ________ 28 ________ 32 ________ 36 ________ Total ________	Remembering 3 ________ 7 ________ 10 ________ 18 ________ 31 ________ 35 ________ Total ________
Interpreting 2 ________ 12 ________ 14 ________ 17 ________ 21 ________ 30 ________ Total ________	Evaluating 1 ________ 8 ________ 22 ________ 23 ________ 29 ________ 34 ________ Total ________	Responding 6 ________ 9 ________ 13 ________ 19 ________ 26 ________ 27 ________ Total ________

Transfer your totals for each component to the Total Points column below. Rank order each of the components according to your totals:

Component	*Total points*	*Rank*
Hearing	________	________
Understanding	________	________
Remembering	________	________
Interpreting	________	________
Evaluating	________	________
Responding	________	________

(Adapted from Brownell, 2002, pp. 31–33.)

Profile 24

Informational Reception Apprehension Test (IRAT)

(Wheeless, Preiss, & Gayle, 1997)

Profiled by: Shaughan A. Keaton

Young Harris College

Construct

Informational Reception Apprehension is a "trait-like anxiety that triggers deficiencies in an individual's ability to receive, process, and interpret, and/or adjust to information" (Schrodt & Wheeles, 2001, p. 57).

Instrument Type

Self-Report

Description

Although derivative of the original unidimensional Receiver Apprehension Test (RAT; Wheeless, 1975), the Informational Reception Apprehension Test (IRAT; Wheeless *et al.*, 1997) includes three subscales concerning apprehension about listening to information (IRAT-L; 13 items), reading (IRAT-R; 18 items), and the willingness to be receptive to abstract ideas (IRAT-IF; 7 items).

Administration

The IRAT can be administered via paper or online with items scaled along five points (from *strongly agree* to *strongly disagree*). The IRAT takes approximately 10 minutes to complete.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Scoring

After recoding positively worded scale items, responses within subscales are averaged to reveal aggregate scores for each subdimension.

Development

Wheeless (1975) originally developed the construct of receiver apprehension (RA) because he noticed that some listeners experienced anxiety when responding to messages. He defined RA as "the fear of misinterpreting, inadequately processing, and/or not being able to adjust psychologically to messages sent by others" (1975, p. 263). The construct later evolved into the three-dimensional form because research revealed that RA was inversely related to an individual's willingness to (a) listen and (b) consider abstract concepts (Wheeless *et al.*, 1997). Reading anxiety was added because people consider and process abstract concepts not only in conversation but textually as well.

In some versions, authors combined items from the RAT and the IRAT to produce a state version called the SRAT, which is supposed to "locate appropriate items that could be restated so as to reflect a specific communication interaction" (Goodboy *et al.*, 2014; Schumacher & Wheeless, 1997, p. 434). Others have used a revised version of the RAT called the Revised Receiver Apprehension Test (Preiss & Gayle, 1999; Winiecki & Ayres, 1999).

Reliability

The RAT was created as a 20-item, unidiminsional scale. Wheeless (1975) reported a principal component analysis that showed all items loaded on a primary component. Reported assessments of internal consistency, using Cronbach's alpha and split-half reliabilities, have ranged from .81 to .91 (Beatty, 1994; Beatty, Behnke, & Henderson, 1980; Bodie & Villaume, 2003; Wheeless, 1975). The derivative IRAT has produced scores with reliability estimates ranging from .86 to .91 for the listening subscale and from .71 to .84 for the intellectual flexibility subscale (Hayhurst, 2002; Keaton, 2013; Ledbetter & Schrodt, 2008; Schrodt, Wheeless, & Ptacek, 2000; Wheeless *et al.*, 1997; Wheeless & Schrodt, 2001). The reading subscale is not included in the subsequent discussion because it is a form of cognitive processing that, although related, is not relevant to listening.

Validity

One fundamental attribute important to the construction and validity assessment of any scale is that items be at least moderately correlated (DeVellis, 2003). This criterion seems to have support for the original and derivative versions claiming to measure informational reception apprehension. Reported interitem correlations have ranged from .81 to .91 for the RAT and .71 to .91 for the IRAT subscales. Dimensionality and data representation of measurements and factors are also important. Curiously, the original developers of the IRAT did not produce model fit estimates when presenting the newer version of the construct (Wheeless, Eddleman-Spears, Magness, & Preiss, 2005)—nor were any offered in subsequent uses of the scale. Instead, principal component analysis was used to produce items that comprise the scale, which is a technique more suitable for item reduction than for securing a stable factor structure (Tabachnick & Fidell, 2007). In fact, none of the studies cited above have assessed either scale for dimensionality or ability to represent data. Although the RAT received greater scrutiny by Beatty and colleagues (1994) in this regard (who also found evidence of concurrent validity), this report is concerned with the IRAT and subsequent adapted versions. In spite of scarce validity evidence available for the IRAT, it is used in many areas of research related to listening.

The construct validity of the IRAT has only been assessed in one report. Keaton (2013) conducted an independent assessment with data from 83 undergraduate students enrolled in communication studies courses at Louisiana State University A&M. This study focused exclusively on items that should represent listening anxiety (n = 13) and intellectual flexibility (n = 7). Although the measurement model containing these 20 items was not supported using confirmatory factor analysis, χ^2 (169) = 309.42, $p < .001$, CFI = .39, RMSEA = .10, CI90% = .09, .12, the sample size may not be adequate to fully test the measurement model. Thus, the dimensionality of the IRAT needs to be assessed in independent samples drawn from various populations.

Availability

The original RAT is available in *Communication Research Measures: A Sourcebook* (Rubin, Palmgreen, & Sypher, 1994), and the IRAT is widely accessible through most of the cited articles above. The IRAT is reprinted with permission at the end of this profile and is free to use, with appropriate citation, for research purposes.

Sample Studies

Examples of previous uses of the RAT (and the RRAT) include trait listening anxiety (Beatty *et al.*, 1980); listening preferences and communication apprehension (Bodie & Villaume, 2003); student attitudes and study habits (Preiss & Gayle, 1999); communication apprehension, personality, and listening behaviors (Hayhurst, 2002); and attitudes toward condom use (Goldman, Martin, Bryand, DiClemente, & Ditrinco, 2014).

Previous studies have examined the IRAT (and the SRAT) in a variety of contexts, including affective learning (Hsu, 2012), listening fidelity (Sawyer, Gayle, Topa, & Powers, 2014), family communication patterns (Ledbetter & Schrodt, 2008), cognition (Ledbetter & Schrodt, 2008; Wheeless *et al.*, 1997), uncertainty (Schumacher & Wheeless, 1997), educational motivation and achievement (Schrodt *et al.*, 2000), instructor self-disclosure (Goodboy *et al.*, 2014), workplace communication (Winiecki & Ayres, 1999), and technology aversion (Wheeless *et al.*, 2005).

Critique

The subscales of the IRAT have demonstrated a consistent ability to exhibit internal consistency, suggesting that researchers have little to fear in terms of the factors underestimating true relationships or contributing to Type II error. However, the validity portfolio for the IRAT is lacking, which indicates that its use as a research measurement is limited and should be questioned. The study by Keaton (2013) suggests problems with dimensionality and model fit that should be examined closely in future studies. All researchers should help contribute to the validity portfolio of scales by reporting reliability estimates and the results of confirmatory factor analyses. In this way, we can all provide evidence in the ongoing process that is validity assessment.

The IRAT as currently utilized may not fully assess receiver apprehension, especially in terms of intellectual flexibility. There may be more important dimensions to be discovered, or it may be that it is most useful as a tool for assessing the listening subdimension alone. Indeed, the listening anxiety factor was psychometrically sound in the data reported by Keaton. One may use the listening anxiety factor with a degree of confidence, but it is recommended to interpret the results derived from the IF factor conservatively.

References

Beatty, M. J. (1994). Receiver apprehension test. In R. B. Rubin, P. Palmgreen, & H. E. Sypher (Eds.), *Communication research measures: A sourcebook* (pp. 304–307). New York, NY: Guilford.

Beatty, M. J., Behnke, R. R., & Henderson, L. S. (1980). An empirical validation of the Receiver Apprehension Test as a measure of trait listening anxiety. *Western Journal of Speech Communication*, *44*, 126–129. doi:10.1080/10570318009373998

Bodie, G. D., & Villaume, D. W. A. (2003). Aspects of receiving information: The relationship between listening preferences, communication apprehension, receiver apprehension, and communicator style. *International Journal of Listening*, *17*, 47–67. doi:10.1080/10904018.2003.10499055

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Goldman, Z. W., Martin, M. M., Bryand, M. C., DiClemente, R. K., & Ditrinco, E. A. (2014). College students' communication apprehension, receiver apprehension, and views about discussing condom use. *North American Journal of Psychology*, *16*, 277–284.

Goodboy, A. K., Carton, S. T., Goldman, Z. W., Gozanski, T. A., Tyler, W. J. C., & Johnson, N. R. (2014). Discouraging instructional dissent and facilitating students' learning experiences through instructor self-disclosure. *Southern Communication Journal*, *79*(2), 114–129. doi:10.1080/1041794X.2013.865256

Hayhurst, J. L. (2002). *The Big Three as related to receiver apprehension and listening behaviors*. Master's thesis, West Virginia University.

Hsu, C.-F. (2012). The influence of vocal qualities and confirmation of nonnative English-speaking teachers on student receiver apprehension, affective learning, and cognitive learning. *Communication Education*, *61*, 4–16. doi:10.1080/03634523.2011.615410

Keaton, S. A. (2013). Informational reception apprehension. Paper presented at the National Communication Association, Washington, DC.

Ledbetter, A. M., & Schrodt, P. (2008). Family communication patterns and cognitive processing: Conversation and conformity orientations as predictors of informational reception apprehension. *Communication Studies*, *59*, 388–401. doi:10.1080/10510970802467429

Preiss, R. W., & Gayle, B. M. (1999). Receiver apprehension, educational attitudes, and study habits. *Journal of the Northwest Communication Association*, *27*, 91–100.

Rubin, R. B., Palmgreen, P., & Sypher, H. E. (1994). *Communication research measures: A sourcebook*. New York, NY: Guilford.

Sawyer, C. R., Gayle, K., Topa, A., & Powers, W. G. (2014). Listening fidelity among native and nonnative English-speaking undergraduates as a function of listening apprehension and gender. *Communication Research Reports*, *31*, 62–71. doi:10.1080/08824096.2013.844119

Schrodt, P. & Wheeless, L. R. (2001). Aggressive communication and informational reception apprehension: The influence of listening anxiety and intellectual inflexibility on trait argumentativeness and verbal agressiveness. *Communication Quarterly*, *49*, 53–69. doi:10.1080/01463370109385614

Schrodt, P., Wheeless, L. R., & Ptacek, K. M. (2000). Informational reception apprehension, educational motivation, and achievement. *Communication Quarterly*, *48*, 60–73. doi:10.1080/01463370009385580

Schumacher, B. K., & Wheeless, L. R. (1997). Relationships of continuing uncertainty and state-receiver apprehension to information-seeking and predictions in dyadic interactions. *Communication Quarterly*, *45*, 427–445. doi:10.1080/01463379709370075

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson Education.

Wheeless, L. R. (1975). An investigation of receiver apprehension and social context dimensions of communication apprehension. *The Speech Teacher*, *24*, 261–268. doi:10.1080/03634527509378169

Wheeless, L. R., Eddleman-Spears, L., Magness, L. D., & Preiss, R. W. (2005). Informational reception apprehension and information from technology aversion: Development and test of a new construct. *Communication Quarterly*, *53*, 143–158. doi:10.1080/01463370500090845

Wheeless, L. R., Preiss, R. W., & Gayle, B. M. (1997). Receiver apprehension, informational receptivity, and cognitive processing. In J. A. Daly, J. C. McCroskey, J. Ayres, T. Hopf, & D. M. Ayres (Eds.), *Avoiding communication: Shyness, reticence, and communication apprehension* (pp. 151–187). Cresskill, NJ: Hampton Press.

Wheeless, L. R., & Schrodt, P. (2001). An examination of cognitive foundations of informational reception apprehension: Political identification, religious affiliation, and family environment. *Communication Research Reports*, *18*, 1–10. doi:10.1080/08824090109384777

Winiecki, K. L., & Ayres, J. (1999). Communication apprehension and receiver apprehension in the workplace. *Communication Quarterly*, *47*, 430–440. doi:10.1080/01463379909385572

Scale

Informational Reception Apprehension Test

Source: Wheeless *et al.* (1997). Reproduced with permission of Hampton Press.

The following statements refer to your feelings about listening to others. Please read each statement and mark your level of agreement/disagreement using the following scale:

1 = Strongly Disagree
2 = Disagree
3 = Neither Agree nor Disagree
4 = Agree
5 = Strongly Agree

Items for the LA and IF subscales of the IRAT

<u>IRAT-L</u>

1) While listening, I get nervous when a lot of information is given at once.
2) I get impatient and anxious when listening to someone discuss theoretical, intellectual issues.
3) I have avoided listening to abstract ideas because I was afraid I could not make sense of what was said.
4) Many classes are annoying and uncomfortable because the teacher floods you with detailed information in the lectures.
5) I feel agitated or uneasy when someone tells me there is not necessarily a clear, concrete way to deal with an important problem.
6) While listening, I feel tense when I have to analyze feelings carefully.
7) It is frustrating to listen to people discuss practical problems in philosophical and abstract ways.
8) When I hear abstract material, I am afraid I will be unable to remember it very well.
9) I experience anxiety when listening to complex ideas others tell me.
10) When I listen to complicated information, I often fear that I will misinterpret it.
11) I feel relaxed and confident while listening, even when a lot of information is given at once. (R)
12) Listening to complex ideas is a pleasant, enjoyable experience for me. (R)
13) When listening, I feel relaxed and confident that I can remember abstract ideas that are being explained. (R)

<u>IRAT-IF</u>

14) I enjoy listening to people discuss intellectual problems.
15) I enjoy listening to abstract topics like politics, philosophy or religion where there are not clear, correct answers.
16) I believe there are at least two sides to every argument, and I enjoy listening to all sides.
17) When I listen to theoretical or hypothetical material, I like to consider the issues and think about the ideas.
18) It is fun and relaxing to seek out the opportunity to listen to new and different ideas.
19) I get curious and actually enjoy listening to someone with a foreign accent.
20) When others are talking, I enjoy "tuning in" to interpret the motives and emotions behind what is being said.

Note: Subscale titles should be removed and items randomly ordered prior to administration. Items marked as (R) should be reverse-coded prior to scoring.

Profile 25

Interaction Involvement Scale (IIS)

(Cegala, 1981; Cegala, Savage, Brunner, & Conrad, 1982)

Profiled by: Debra L. Worthington, PhD

Auburn University

Construct

Cegala (1981) conceptualized *interaction involvement* (II) as a dimension of communicative competence reflecting "the extent to which an individual partakes in a social environment" (p. 112).

Instrument Type

Self-Report

Description

The 18-item Interaction Involvement Scale (IIS) measures three facets of individual conversational involvement: attentiveness, perceptiveness, and responsiveness. (Cegala, 1981; Cegala *et al.*, 1982). Scale items describe an individual's perception of their own interaction behaviors. *Attentiveness* assesses an individual's awareness of factors affecting an interaction, *perceptiveness* addresses one's understanding of message meanings, and *responsiveness* measures a person's belief that she can respond appropriately to others during an interaction. Cegala (1981) originally described II as one aspect of communicative competence in the context of interpersonal communication. The state scale is administered immediately following a conversation to measure the level of individual involvement in the interaction (Cegala, 1981; Cegala *et al.*, 1982). The three dimensions are correlated and have been examined individually and as a composite score.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

The II scale was originally designed as a paper survey, although more recent studies have utilized computer-based and online assessments. The scale takes less than 10 minutes to administer. Respondents answer the 18 items using a 7-point scale, ranging from *not at all like me* (1) to *very much like me* (7). On the revised trait measure, four items measure perceptiveness, six items measure attentiveness, and eight are used to measure responsiveness. Higher scores reflect greater involvement.

Scoring

After reverse scoring 12 of the 18 items, answers to the items composing each subscale are summated. Participants receive subscores for each of the three areas (i.e., perceptiveness, attentiveness, and responsiveness) as well as an overall composite score.

Development

Based on Erving Goffman's early work on the nature of interpersonal interaction (Cegala, 1981, 2009), II reflects an individual's ability to "achieve interpersonal goals without result of loss of face to self or others" (Cegala, 1981, p. 111).

Cegala's initial studies (1981; Cegala *et al.*, 1982), using principal component analysis (PCA), identified three components explaining item variance: attentiveness (26% of item variance), perceptiveness (24%), and other-oriented perceptiveness (19%). A follow-up study found that individuals with higher perceptiveness scores had greater success in gaining information from others. In a later study, Cegala *et al.* (1982) reconsidered the original dimensions of II, adding the responsiveness dimension. Responsiveness accounted for the greatest percentage of total variance (33.8%) when items from 1802 respondents were submitted to PCA, followed by perceptiveness (10%) and attentiveness (9.3%). Correlations between the three factors ranged from .24 to .35. The authors argued that the revised scale provided an improved measure of a person's typical overt communication behaviors.

Attentiveness, as the most basic component, addresses the ability to attend to incoming visual and aural information in a social environment—that is, how well each member of an interpersonal interaction attends to verbal and nonverbal cues during a communicative event. *Perceptiveness* builds from attentiveness as the interlocutors interpret and assign meanings to these cues and attempt to make sense of the interaction. *Responsiveness* transpires as individuals respond to each other's verbal and nonverbal behavior and adapt to their social, communicative event.

Early studies using the IIS found that highly involved individuals are skilled at attending to, assessing, and responding to individuals and to the communicative elements of a conversation. They often engage in more eye gaze, use more immediate language, and tend toward gestures that are more object-focused rather than body-focused (Cegala, 1989; Cegala *et al.*, 1982). In contrast, individuals who are less involved in conversations are often psychologically and communicatively removed from the social context and may appear distracted or preoccupied. Cegala *et al.* (1982) described their speech as vague and inconsistent. They tend to speak with less certainty and use fewer personal pronouns.

A series of studies have examined the relation between II and a variety of psychological and communicative constructs (see the Validity and Sample Studies sections). Although both state and trait versions of the IIS were developed, the self-report trait version is most frequently used by researchers. Thus, less is known about the state version of the scale and findings associated with it (Frymier, 2005).

While initial studies focused on interpersonal contexts (Cegala, 1981; Huang & Huang, 2012), later studies expanded and adapted to other communication contexts including patterns associated with media use (e.g., email, mobile phones) as well as small group tasks, classroom and cultural differences (see for example, Cegala, Wall, & Rippey, 1987; Frymier, 2005; Myers & Bryant, 2002; Sun, Hullman, & Wang, 2011; Worthington, Fitch-Hauser, & Kim, 2008). These studies have used both student and non-student participants, as well as individuals in prescribed or established relationships.

It is not unusual for researchers to selectively choose subscales for specific research purposes or adapt items to specialized contexts (see for example, Nguyen & Fussell, 2016).

Reliability

Reliability of the IIS has been estimated by several means. Internal consistency has been tested by several researchers, with Cronbach alphas for the three subscales ranging widely, as low as .35 for the attentiveness subscale to a high of .88 also for the attentiveness subscale (Ragsdale, 1994). While the attentiveness subscale seems most problematic, most studies report internal consistency estimates in the .80s across the subscales and the overall scale.

Cegala *et al.* (1982) and Rubin and Graham (1988) estimated the test-retest reliability of the scale. After a six week delay, Cegala *et al.*, reported a correlation of .81; after a one year delay, Rubin and Graham reported a test-retest reliability of .61 of the trait measure. However, the reliability estimates were somewhat lower for the individual scales. Overall, the correlations between the three factors across the two tests ranged between .56 and .59.

Frymier (2005), utilizing the state version of the scale in a study of out-of-class student–teacher communication, reported alpha reliabilities ranging from .68 (perceptiveness) to .83 (attentiveness) and .86 (responsiveness), and .86 for the full scale. However, these reliabilities were based on items that, following PCA, did not load on the three components as expected (one item, "I feel sort of unplugged…" was removed from analysis because it did not meet minimum requirements).

Välikoski, Imhof, Worthington, Fitch-Hauser, and Kim (2008), in a cross-cultural study of mobile phone use, compared differences in subscale and total IIS scores across four countries: Finland, German, Korea, and the United States. The scale had been modified to reflect the mobile phone context under study. Reliability estimates across the four groups were moderate to good for all but the perceptiveness scale: attentiveness, $.62 < \alpha < .82$; perceptiveness, $.56 < \alpha < .74$; responsiveness, $.85 < \alpha < .86$; and, overall, $.78 < \alpha < .87$.

Validity

A number of studies support the scale's concurrent validity. These studies have utilized subscale and composite scores of the measure. Umphrey, Wikersham, and Sherblom (2008) found II to be positively correlated with immediacy (.45), composure (.35), receptivity (.40), connectedness (.55), communication satisfaction (.62), and communication quality (.59). Reflecting earlier findings by Cegala (1981), II is positively correlated with student question asking, behavioral flexibility, interaction management, and overt information seeking (Myers, Edwards, Wahl, & Martin, 2007). Of note, Umphrey *et al.* (2008) and Myers *et al.* (2007) did not provide evidence that they confirmed the three-factor model of the scale prior to their analysis and only report findings for the composite IIS score.

Looking at the three dimensions separately, responsiveness has been positively associated with sociability, intent to disclose, social confirmation and appropriate disclosure, and negatively associated with social anxiety and neuroticism (Cegala *et al.*, 1982; Chen, 1989; Duran & Kelly, 1988). Attentiveness is inversely related to neuroticism and impulsiveness, but positively related to intent to self-disclose and social experience (Cegala *et al.*, 1982; Chen, 1989; Duran & Kelly, 1988). Perceptiveness is positively correlated with communication competence, behavioral flexibility, and interaction management as well as self-consciousness and social composure, and two dimensions of Lennox and Wolfe's (1984) Revised Self-Monitoring Scales: ability to modify self-presentation, $r = .31$, and sensitivity to expressive behavior of others, $r = .27$ (Cegala *et al.*, 1982; Chen, 1989; Duran & Kelly, 1988). Using confirmatory factor analysis (CFA), Välikoski *et al.* (2008) confirmed the dimensionality of the IIS with students from Finland, Germany, and Korea.

Frymier (2005) assessed the state version of the IIS, using CFA procedures. Although the three extracted factors reflected that of Cegala's (1981) original structure, scale items did not load as expected.

Availability

The most commonly used trait version of the scale is provided below, with permission. Frymier (2005) presented a state version of the scale for use in the education context. All versions are free to use for research purposes.

Sample Studies

Research into II has typically focused on either the cognitive/affective characteristics or the verbal/nonverbal behaviors differentiating individuals of high and low involvement. These studies suggest that higher involvement is associated with higher self-esteem, greater emotional stability, positive moods, as well as greater belief in one's own communication competence. For example, Bodie (2010) found Active-Empathic Listening (see AELS profile, Profile 2) positively related to II – correlations for composite AELS and IIS ranged from .23 for attentiveness to .50 for perceptiveness.

The role of II in the education context has received increasing attention. In the classroom context, overall interaction scores were shown positively associated with question asking, out-of-class interactions, and overt information seeking, and inversely related to verbal aggressiveness (Myers *et al.* 2007). Frymier (2005) used an adapted form of the scale to examine the relation between effective student communication and positive student outcomes in out-of-class interactions. On the whole, these studies suggest that students with higher interaction involvement in a class are more motivated to study, are more satisfied with their classroom communication, and report higher grades. Frymier (2005) reported overall II is positively associated with student reports of satisfaction with instructor communication (.49).

In a study examining the relations between dimensions of religious commitment and communication characteristics, Ragsdale (1994) reported that II perceptiveness had a small positive relationship with several religious life inventory scales.

Researchers also are beginning to explore the role and effect of II in mediated interactions. Umphrey *et al.* (2008) compared face-to-face and video-conferenced classes, examining a number of factors including II. This study differed from most others in that it used path analysis to model the relations of the studied factors. A number of positive associations with II were found with correlations ranging from .35 to .62: composure, receptivity, immediacy, mutuality, communication quality, and communication satisfaction. Not surprisingly, higher levels of involvement were reported in the face-to-face class.

Välikoski *et al.* (2008) reported significant differences in a cross-cultural comparison of IIS scores. German students averaged higher attentiveness scores ($M = 15.07$), with US participants the lowest: ($M = 11.45$), F (3, 987) = 53.14, $p < .001$. German and Finnish participants averaged higher responsiveness scores ($M = 32.5$; 31.5, respectively) than US Americans ($M = 30.2$) or Koreans ($M = 29.6$), F (3, 979) = 11.49, $p < .001$, and US Americans averaged higher perceptiveness scores than the other three groups ($M = 18.34$, $SD = 10.23$), F (3, 984) = 51.79, $p < .001$. Välikoski *et al.* suggested that interaction involvement might affect mobile phone users' privacy management. Their regression analysis suggested that attentiveness predicted German, Korean, and US Americans' willingness to move to a more private location when interacting with someone on their mobile phone; perceptiveness was predictive of avoidance of perceived sensitive topics when in the presence of others for Finns, Koreans, and Americans.

Critique

Based on the published reports of reliability and validity, the trait measure of interaction involvement appears to be a sound scale for use in a range of contexts. The early development of the IIS was theory driven, incorporating principle component factor analyses. However, more recent researchers often do not test for model fit using confirmatory techniques, assuming that the scale items will fit Cegala's model. Unfortunately, this may not be case (see, e.g., Frymier, 2005).

Researchers using the scale should be aware of, and test for, the sometimes lower reliability estimates of the subscales. Also, the scoring procedure of the scale has not been assessed. Scores for each facet of the scale are summated, not averaged. As noted above, on the revised trait measure, four items measure perceptiveness, six items measure attentiveness, and eight are used to measure responsiveness. Subsequently, researchers should be aware that responsiveness carries greater weight in individual composite scores.

Although II and the many constructs associated with II have not been thoroughly studied in terms of its potential contributions to listening processes (for an exception, see Bodie, 2010), many of the constructs that have been associated with II have direct relations to those associated with listening processes and behaviors. However, listening as an active interaction has received less attention. Further study in additional contexts is needed as is research utilizing the state version of the scale. Cegala (2009) arguably offered the best critique of the II scale. He noted that research into the construct has utilized the self-report trait scale, thus limiting our understanding of the construct in actual interactions. Focusing on individuals at the extreme ends of the continuum means we have less of an understanding of those who lie between or how and why levels of involvement vary over the course of a communicative event (Cegala, 2009). Thus, despite its potential to measure individual cognitive and behavioral engagement in their conversations with others, few studies have employed this measure. Finally, researchers should exercise caution when shortening or modifying the scale.

References

Bodie, G. D. (2010) The Active-Empathic Listening Scale (AELS): Conceptualization and evidence of validity within the interpersonal domain. *Communication Quarterly*, *59*, 277–295. doi:10.1080/01463373.2011.583495

Cegala, D. J. (1981). Interaction involvement: A cognitive dimension of communicative competence. *Communication Education*, *30*, 109–121. doi:10.1080/03634528109378460

Cegala, D. J. (1989). A study of selected linguistic components in interaction. *Western Journal of Speech Communication*, *53*, 311–326. doi:10.1080/10570318909374309

Cegala, D. J. (2009) Interaction involvement. In S. W. Littlejohn & K. A. Foss (Eds.), *Encyclopedia of communication theory* (pp. 526–528). Sage: Thousand Oaks, CA.

Cegala, D. J., Savage, G. T., Brunner, C. C., & Conrad, A. B. (1982). An elaboration of the meaning of interaction involvement: Toward the development of a theoretical concept. *Communication Monographs*, *49*, 229–248. doi:10.1080/03637758209376087

Cegala, D. J., Wall, V. D., Jr., & Rippey, G. (1987). An investigation of interaction involvement and the dimensions of SYMLOG: Perceived communication behaviors of persons in task-oriented groups. *The Central States Speech Journal*, *38*, 81–91. doi:10.1080/10510978709368233

Chen, G. M. (1989). Relationships of dimensions of intercultural communication competence *Communication Quarterly*, *37*, 118–133. doi:10.1080/01463378909385533

Duran, R. L., & Kelly, L. (1988). An investigation into the cognitive domain of the competence II: The relationship between communicative competence and interaction involvement. *Communication Research Reports*, *5*, 91–96. doi:10.1080/08824098809359806

Frymier, A. B. (2005). Students' classroom communication effectiveness. *Communication Quarterly*, *53*, 197–212. doi:10.1080/01463370500089896

Huang, E. Y., & Huang, T. K. (2012). Investigating the antecedents of users' knowledge sharing intention. *The Journal of Computer Information Systems*, *53*, 93–102. doi:10.1080/08874417.2012.11645618

Myers, S. A., & Bryant, L. E. (2002). Perceived understanding, interaction involvement, and college student outcomes. *Communication Research Reports*, *19*, 146–155. doi:10.1080/08824090209384842

Myers, S. A., Edwards, C., Wahl, S. T., & Martin, M. M. (2007). The relationship between perceived instructor aggressive communication and college student involvement. *Communication Education*, *56*, 495–508. doi:10.1080/03634520701466398

Nguyen, D. T., & Fussell, S. R. (2016). Effects of conversational involvement cues on understanding and emotions in instant messaging conversation. *Journal of Language and Social Psychology*, *35*, 28–55. doi:10.1177/0261927X15571538

Ragsdale, J. D. (1994). Relational communication competence in high and low Christian religious commitment: A research note. *Review of Religious Research*, *35*, 268–274.

Rubin, R. B., & Graham, E. E. (1988). Communication correlates of college success: An exploratory investigation. *Communication Education*, *37*, 14–27. doi:10.1080/03634528809378700

Sun, S., Hullman, G., & Wang, Y. (2011). Communicating in the multichannel age: Interpersonal communication motivation, interaction involvement, and channel affinity. *Journal of Media and Communication Studies*, *3*, 7–15.
Umphrey, L. R., Wickersham, J. A., & Sherblom, J. C. (2008). Student perceptions of the instructor's relational characteristics, the classroom communication experience, and the interaction involvement in face-to-face versus video conference instruction. *Communication Research Reports*, *25*, 102–114. doi:10.1080/08824090802021954
Välikoski, T., Imhof, M., Worthington, D., Fitch-Hauser, M., & Kim, S-H. (2008, November). *Mobile telephony, listening and privacy management: A cross-cultural comparison of Finnish, German, Korean, and U.S. American students*. Paper presented at the European Communication Research and Education Association, Barcelona, Spain.
Worthington, D. L., Fitch-Hauser, M., & Kim, S. (2008, November). *Cell phones and listening Habits: How do differences in interaction involvement affect cell phone use of U.S. American and Korean students?* Paper presented at the National Communication Association convention, San Diego, CA.

Scale

Interaction Involvement Scale (Cegala, 1981; Cegala *et al.*, 1982)

Source: Cegala (1981) and Cegala *et al.* (1982).

Instructions: This questionnaire is designed to provide information about how people communicate. There are no right or wrong answers to any of the items. You only need to indicate the extent to which you feel each item describes your own behavior.

In responding to some of the items, you might say, "Sometimes I do that, and sometimes I don't." You should respond to each item in a way that best describes your *typical* manner of communication—how you behave in most situations. If you cannot decide how a particular item applies to you, circle the "not sure" alternative. However, please be sure to respond to all of the items.

Circle the one alternative for each item that best characterizes your communication in general.

Not at all like me	*Not like me*	*Somewhat like me*	*Not sure*	*Somewhat like me*	*Like me*	*Very much like me*
1	2	3	4	5	6	7

<u>Perceptiveness</u>

1) I am keenly aware of how others perceive me during my conversations.
2) Sometimes during conversations I'm not sure what the other really means or intends by certain comments. (R)
3) In my conversations I often do not accurately perceive others' intentions or motivations. (R)
4) In conversations I am very perceptive to the meaning of my partner's behavior in relation to myself and the situation.

Attentiveness

5) My mind wanders during conversations, and I often miss parts of what is going on. (R)
6) I am very observant of others' reactions while I'm speaking.
7) During conversations I listen carefully to others and obtain as much information as I can.
8) Often in conversations I will pretend to be listening when in fact I was thinking of something else. (R)
9) I carefully observe how the other is responding to me during a conversation.
10) Often I'm preoccupied in my conversations and do not pay complete attention to others. (R)

Responsiveness

11) Often in conversations I'm not sure what to say. I can't seem to find the appropriate lines. (R)
12) Often in conversations I'm not sure what my role is. I'm not sure how I'm expected to relate to others. (R)
13) Often during conversations I feel like I know what should be said (like accepting a compliment or asking a question), but I hesitate to do so. (R)
14) Often I feel withdrawn or distant during conversations. (R)
15) Often in conversations I'm not sure what others' needs are (e.g., a compliment, reassurance, etc.) until it is too late to respond appropriately. (R)
16) I feel confident during my conversations. I am sure of what to say and do.
17) Often I feel sort of "unplugged" during conversations. I am uncertain of my role, others' motives, and what is happening. (R)
18) Often during my conversation I can't think of what to say. I just don't react quickly enough. (R)

Note: Labels and reverse-coding designations (R) should be removed and items randomized prior to administration. Scores (1 = *not at all like me*; 7 = *very much like me*) are summated for each subscale and the composite score. Alternative versions can easily be developed (e.g., "I feel confident during my classroom conversations, I am sure of what to say and do"). Higher scores indicate greater involvement.

Profile 26

Feffer's Interpersonal Decentering

(Feffer, Leeper, Dobbs, Jenkins, & Perez, 2008)

Profiled by: Molly S. Tucker, MS and Sharon Rae Jenkins, PhD

University of North Texas

Construct

Feffer's Interpersonal Decentering is a developmental social cognitive construct designed to assess a person's tendency to interact maturely with others, such as spontaneously taking the perspective of others, anticipating or reflecting on social interactions, considering the other's mental state, and anticipating the other's response to one's own actions before acting.

Instrument Type

Narrative Content Analysis

Description

The Interpersonal Decentering (ID) scoring system gauges the functional maturity of a person's capacity for social cognitive information processing (often referred to as *social perspective-taking*, *role taking*, or *mentalizing*) by analyzing content of spontaneous narratives (see Role Category Questionnaire, Profile 56). The scoring system used to generate ID scores is best described as an implicit human-scored content analysis scoring system. The system can be applied to several types of data, including thematic apperceptive technique stories (TATs), expressive writing essays, and other personal or fictional narratives.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

Data may be gathered by a variety of methods that are nondirective as to content; the scoring system also may be applied post hoc to narratives gathered for other purposes. Standard TAT instructions, or reasonable approximations, may be utilized or adapted to elicit fictional stories:

> "This is a test of imagination. I am going to show you some pictures—one at a time—and your task will be to make up a story for each one. In your story, tell what has led up to the event shown in the picture, describe what is happening in the picture, what the characters are thinking or feeling, and then give the outcome. Speak your thoughts as they come to your mind. Do you understand?"
>
> *(Morgan & Murray, 1935)*

Feffer and Jahelka's (1968) "initial story" method was adapted into the current manual; this is now used as the standard scoring reference work and to teach Decentering scoring to new research assistants (Feffer *et al.*, 2008).

Scoring

Interaction units (denoted by same characters, time period, and location) are identified within a story. Each is given a Decentering score ranging from 1 to 9 depending on the intricacy with which characters are differentiated and/or internalized as opposed to undifferentiated and egocentric. *Basic concrete statements* without internalization are assigned scores ranging from 1 to 4. When characters are not differentiated from one another (e.g., "We wanted ice cream"), a score of 1 is assigned. If a directed action does not evoke a reaction (e.g., "He spoke to her"), a score of 2 is assigned; this increases to 3 if a response is evoked (e.g., "He spoke to her and she responded"), or a 4 if an additional reaction is evoked from the initiator (e.g., "He spoke to her and she responded, so he sat beside her").

Categories 5 through 9 necessitate one character internalizing another. The other may be undifferentiated (5, e.g., "He was wondering about her") or characterized by some distinguishing feature (6, e.g., "He was wondering about her uniform"). When a character internalizes another character and refers to the other's internal state, a score of 7 is assigned (e.g., "He wanted her opinion" or "He thought she might be angry"). A character internalizing another character, who is internalizing yet another character, is scored as an 8 (e.g., "He believed his mother was thinking of his father"). Finally, the highest score of mentalizing (9) involves reflection upon one's own beliefs, emotions, or behaviors with respect to another (e.g., "He wished he hadn't been honest with her").

This brief summary is not adequate for reliable scoring; study of the manual and coding practice stories with expert judges are necessary. Practice materials are available in Feffer *et al.* (2008); additional materials are available by emailing the second author at Sharon.Jenkins@unt.edu.

At this writing, three summary scores are used when there is more than one story per person:

1) the highest single score across a series of stories,
2) the average across stories of the highest score for each story, and
3) the average across interactions within each story, which is then averaged across stories.

First, the high score represents an individual's capacity for mature spontaneous decentering at least once across the sample of story situations. The average of highest scores suggests how consistently the person uses mature levels of decentering across the variety of situations sampled. This score is, however, often correlated with the total number of interactions and the total number of words, and if so, might be confounded with verbal fluency (story length) or loquacity. Finally, the score that averages the within-story averages represents the consistency of mature decentering within situations as well as across situations (Feffer *et al.*, 2008).

Development

"Theory of mind" constructs designed to explain children's understanding of the social and psychological world (role taking, perspective taking, and decentering) are often used interchangeably in the assessment of social cognition (Lewis & Carpendale, 2011). In Feffer's original Role-Taking Task (RTT), individuals are tasked with viewing and generating a story about a picture that depicts several characters; they are then prompted to retell the story from the perspective of each character. This procedure may be repeated with multiple pictures, each of which portrays different characters in ambiguous situations. RTT scores depend on the participant's propensity to individualize each character's experience while simultaneously producing an integrated story premise (Feffer, 1959). There was, however, some uncertainty about what exactly was determining manifest role taking: the integration of perspectives evident in the first story or the consistency between perspectives demonstrated in all parallel versions of the initial narrative. In response, Feffer and Jahelka (1968) elaborated upon the RTT model and studied the initial and subsequent narratives generated for a single TAT card. Their results suggested that the degree of integration of each character's perspectives within the initial TAT story was indeed associated with the maturity level of social interaction exhibited in the following stories. The current Interpersonal Decentering scoring manual is an elaboration of this initial-story methodology (Feffer *et al.*, 2008).

Feffer's Interpersonal Decentering narrative assessment method of perspective taking differs from other social cognitive assessments in the minimal structure provided by the instructions and the picture stimuli. Thus, this implicit or indirect process-sampling measure is not vulnerable to the social desirability bias or response set measurement error that is common in self-report scales; the response process (narrative) is the measure (also see Role Category Questionnaire, Profile 56). Because the purpose is to elicit spontaneous narrative material (either fictional stories or personal narrative), a variety of instructions may be used as long as the narrator is left free to structure the narrative at will. Variations on this approach (i.e., requesting a particular kind of story content, suggesting that there is a "right answer") would constitute an experimental condition. Any decentering processes that occur during the response phase should be attributable to spontaneous processes and the disposition of the storyteller in interaction with the sample of character roles and situations shown in the chosen pictures or requested by the instructions. TAT stories tend not to show systematic practice effects (so long as feedback is not offered) after multiple administrations, although to avoid ambiguity about the role of memory, second and successive administrations should include instructions to "write [tell] whatever comes to mind right now, don't worry about what you said before."

Reliability

Interpersonal Decentering scoring training can be completed by advanced undergraduates over the course of one month with 1–2 hours work per day after the initial studying. Training includes rigorous study of the scoring manual, daily scoring of practice story sets, and comparison of individual scores with expert consensus scores until the reliability criterion of rho > .80 is attained (Jenkins, 2008). Published reliability scores have ranged from .80 to .86 (Jenkins, Austin, & Boals, 2013; Jenkins, Nowlin, & Wilson, under review; Tucker, Baxley, Jenkins, & Johnson, in press). Pairs of advanced undergraduate scorers may score the same narratives independently and then convene twice a week to compare results and reconcile differences. For research purposes, these scores are then deliberated with other pairs in a weekly scoring meeting moderated by an expert; this helps to prevent coder drift (see Chapter 6, this volume). Pre-discussion scores are utilized to calculate interscorer reliability.

Validity

Feffer's Interpersonal Decentering scoring system demonstrates content validity through literal sampling of the respondent's social cognitive processes via constructed narratives. Integration of characters' viewpoints within the initial story of a TAT administration is related to subsequent level of social reciprocity (Feffer & Jahelka, 1968). There is concurrent validity shared between Feffer's Role Taking Task (RTT) and the Interpersonal Decentering scoring system, as well as discriminant validity with certain other TAT scoring systems. Convergent validity has been demonstrated between mature Interpersonal Decentering and counts of cognitive and insight words in Expressive Writing (EW) essays (Jenkins *et al.*, 2013). Additionally, a study of Interpersonal Decentering, treatment alliance, and therapeutic orientation in a clinical sample demonstrated good predictive validity (Jenkins *et al.*, under review), a strength of the scoring system. Other research has reported an association between the average level of Decentering and performance on the Picture Arrangement (PA) subtest of the Wechsler-Bellevue Intelligence Scale (W-BIS) for 18-year-old Berkeley Guidance Study participants (Wilson, Jenkins, & Tucker, 2014). This finding is demonstrative of convergent validity between two implicit measurement methods of role taking: One requires narrative construction, whereas the other necessitates visuospatial organization of inferred narratives.

Availability

The Feffer Interpersonal Decentering Manual (Feffer *et al.*, 2008) is available for purchase online at https://www.routledge.com. Additional practice scoring materials may be requested from Sharon.Jenkins@unt.edu.

Sample Studies

Feffer's ID scoring system has demonstrated utility in its ability to differentiate those individuals who have developed more mature social cognition from those who maintain a more egocentric view of the world. It has distinguished adolescents with schizophrenia diagnoses from those not diagnosed (Strober, 1979), predicted retention versus attrition from therapy as well as clients' perception of insight-oriented versus cognitive behavioral therapy processes (Jenkins *et al.*, under review), differentiated perpetrators of domestic violence from survivors (Jenkins, Dobbs, & Leeper, 2015), and predicted marital satisfaction (Tucker *et al.*, in press). When scoring Interpersonal Decentering from Expressive Writing (EW) essays about upsetting relational events, a clear association between more mature perspective taking and increased use of cognitive, insight, and positive emotion words emerged (Jenkins *et al.*, 2013). These patterns of word use during EW are associated with symptom improvement (Ramirez-Esparza and Pennebaker, 2006). For women, more mature perspective taking is associated with lower levels of depression symptoms and hyperarousal following a romantic breakup (Tucker, Jenkins, Sebastian, Dziurzynski, & Wilson, 2015). Perspective taking, as measured by ID, may be instrumental in recovery and resilience in the face of relational life stressors. What's more, ID appears to be related to performance-related subtests of the W-BIS during adolescence (Wilson *et al.*, 2014) and verbal subtests of the Wechsler Adult Intelligence Scale–Revised (WAIS-R) for subjects in their 30s (Tucker, Johnson, Wilson, & Jenkins, 2015). This suggests that the utility of ID skills may differ between adolescents (who may be verbally underdeveloped) and adults (who rely heavily on verbal organization of the social world).

Critique

The ID scoring system's greatest weakness is that it is only now accruing a supportive literature, in part because the methodological paradigm for narrative assessment measures against which it should be critiqued is itself not yet defined in a way that transcends theoretical traditions (Jenkins, 2017b). The largest story-based research literature is on the human motives of need for achievement, need for affiliation, and need for power (McClelland, 1985; Smith, 1992), and that paradigm's methodological specifications are theory specific. Clinical applications of storytelling assessment (e.g., Bellak & Abrams, 1997; Jenkins, 2008) typically rely on theory-based interpretation, with or without formal scoring. Although there exist critiques of this methodology based on inappropriate application of the psychometric criteria designed for self-report scales and ability tests (Jenkins, 2017a; e.g., empathy; see Interpersonal Reactivity Index, Profile 28), there is not yet a clear set of evaluative criteria by which to judge this system's soundness, although such a framework is currently in development (Jenkins, 2017a). As a human-scored content analysis measure, its most important vulnerability is poor scorer training and possible coder drift resulting in low interscorer reliability.

References

Bellak, L., & Abrams, D. M. (1997). *The Thematic Apperception Test, the Children's Apperception Test, and the Senior Apperception Technique in clinical use* (6th ed.) Boston, MA: Allyn & Bacon.

Feffer, M. H. (1959). The cognitive implications of role taking behavior. *Journal of Personality*, *27*, 152–168. doi:10.1111/j.1467-6494.1959.tb01826.x

Feffer, M., & Jahelka, M. (1968). Implications of the decentering concept for the structuring of projective content. *Journal of Consulting and Clinical Psychology*, *32*, 434–441. https://dx.doi.org/10.1037/h0026101

Feffer, M., Leeper, M., Dobbs, L., Jenkins, S. R., & Perez, L. (2008). Scoring manual for Feffer's Interpersonal Decentering. In S. R. Jenkins (ed.), *Handbook of clinical scoring systems for thematic apperceptive techniques* (pp. 157–180). Mahwah, NJ: Lawrence Erlbaum.

Jenkins, S. R. (2008). Teaching how to learn reliable scoring. In S. R. Jenkins (Ed.), *Handbook of clinical scoring systems for thematic apperceptive techniques* (pp. 39–66). Mahwah, NJ: Lawrence Erlbaum.

Jenkins, S. R. (2017a). Not your same old story: New rules for thematic apperceptive techniques (TATs). *Journal of Personality Assessment*, *99* (3), 238–253, doi:10.1080/00223891.2016.1248972.

Jenkins, S. R. (2017b). The narrative arc of TATs: Introduction to the JPA special section on thematic apperceptive techniques, *Journal of Personality Assessment*, *99* (3), 225–237, doi: 10.1080/00223891.2016.1244066.

Jenkins, S. R., Austin, H., & Boals, A. (2013). Content analysis of expressive writing narratives about stressful relational events using Interpersonal Decentering. *Journal of Language and Social Psychology*. doi:10.1177/0261927X13479188

Jenkins, S. R., Dobbs, L., & Leeper, M. (2015). Using the Thematic Apperception Test to assess Interpersonal Decentering in violent relationships. *Rorschachiana: Journal of the International Society for the Rorschach*, *36*, 180–200. doi:10.1027/1192-5604/a000069

Jenkins, S. R., Nowlin, R. B., & Wilson, J. A. (Under review). *Interpersonal Decentering predicts therapy retention and client perceptions of psychotherapy process.*

Lewis, C., & Carpendale, J.M. (2011). Social cognition. In P. K. Smith & C. H. Hart (Ed.), *The Wiley Blackwell handbook of childhood social development* (2nd ed., pp. 531–548). Oxford: Blackwell.

McClelland, D. C. (1985). *Human motivation*. Glenview: IL: Scott, Foresman.

Morgan, C., & Murray, H. A. (1935). A method for investigating fantasies: The Thematic Apperception Test. *Archives of Neurology and Psychiatry*, *34*, 289–306. doi:10.1001/archneurpsyc.1935.02250200049005

Ramirez-Esparza, N., & Pennebaker, J. W. (2006). Do good stories produce good health? Exploring words, language, and culture. *Narrative Inquiry*, *16*, 211–219. doi:10.1075/ni.16.1.26ram

Smith, C. P. (Ed.). (1992). *Motivation and personality: Handbook of thematic content analysis*. New York: Cambridge University Press.

Strober, M. (1979). The structuring of interpersonal relations in schizophrenic adolescents: A decentering analysis of Thematic Apperception Test stories. *Journal of Abnormal Child Psychology*, *7*, 309–316. doi:10.1007/BF00916540

Tucker, M. S., Baxley, C., Jenkins, S. R., & Johnson, B. (in press). Interpersonal Decentering predicts marital satisfaction. *Yansitma* [Projection].

Tucker, M. S., Jenkins, S. R., Sebastian, B. B, Dziurzynski, K. E., & Wilson, J. A. (2015). Interpersonal Decentering, depression, and trauma symptoms after a romantic breakup. Unpublished manuscript.

Tucker, M. S., Johnson, B., Wilson, J. A., & Jenkins, S. R. (2015, March). *Interpersonal Decentering and IQ: Changes in social cognitive maturity*. Poster presented at the annual meeting of the Society for Personality Assessment, Brooklyn, NY.

Wilson, J. A., Jenkins, S. R., & Tucker, M. S. (2014, August). *Interpersonal Decentering and picture arrangement: Social-cognitive maturity and performance IQ*. Poster presented at the annual convention of the American Psychological Association, Washington, DC.

Further Reading

Jenkins, S. R., Čermák, I., & Fikarová, T. (2013). Lucy's Interpersonal Decentering. In I. Čermák & T. Fikarová (Eds.), *The case of Lucy, seen through multimodal assessment*. Nove Zamky: Psychoprof.

Leeper, M., Dobbs, L., & Jenkins, S. R. (2008). Melvin Feffer's Interpersonal Decentering. In S. R. Jenkins (Ed.), *Handbook of clinical scoring systems for thematic apperceptive techniques* (pp. 149–156). Mahwah, NJ: Erlbaum.

Profile 27

Interpersonal Process Recall (IPR)

(Kagan, 1976, 1980a; Kagan & Kagan, 1997)

Profiled by: Christopher T. Belser, PhD

University of Central Florida

Construct

Interpersonal Process Recall (IPR) was designed as a method of helping clinical supervisees explore personal reactions within therapy sessions using video or audio recordings.

Instrument Type

Behavioral Assessment

Description

Interpersonal Process Recall (IPR) is a counseling strategy that utilizes audio or video recordings of prior-held therapy sessions to elicit the clinician's personal reactions to clients and the counseling process (Bernard & Goodyear, 2014; Kagan, 1976; Kagan & Kagan, 1990). The process involves the clinician, an inquirer (often the clinician's clinical supervisor), and sometimes the client (McQuellon, 1982). The inquirer uses non-judgmental, noninterpretive questions to help the clinician explore the interpersonal and intrapersonal dynamics of a therapy session in an environment that reduces the clinician's anxiety and fear of criticism (Clarke, 1997; McQuellon, 1982). The newfound knowledge gained through IPR can then be used in a therapy session to help the counselor develop self-awareness and critical thinking skills and to assist in client growth (Spivack, 1974). Although IPR was primarily developed for use in mental health fields, it has been utilized in medical and education professions, workplace development, and qualitative interview protocol (Kagan & Kagan, 1997; Larsen, Flesaker, & Stege, 2008; Macaskie, Lees, & Freshwater, 2015).

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

After a counseling session, the clinician and the inquirer meet to review the recorded session. Either person can stop the playback when he or she observes something important happening in the session or something that is not being verbally addressed in the session (Bernard & Goodyear, 2014; Clarke, 1997). IPR questions, or leads, can explore a variety of possibilities, such as affect, cognitions, nonverbal behaviors, expectations, fears, and fantasies. Both Bernard and Goodyear (2014) and Kagan (1980a) provided categorized lists of possible questions that the inquirer can ask. Borders and Brown (2005) recommended discussing the session in the present tense to allow the clinician to be present during the recall session. When using IPR, the inquirer should remain impartial, should refrain from taking on a teaching role, and should not point out alternate steps the clinician could have taken in session; instead, the inquirer should adhere to neutral, open-ended questions that encourage the clinician to explore thoughts and feelings (Cashwell, 1994).

Scoring

IPR does not involve scoring, as it involves open-ended questions in response to counseling sessions.

Development

Stemming from research into the dynamics of empathy, IPR began as a postsession interview process to help the clinician and the client discuss their here-and-now reflections while watching the playback of their recorded session (Kagan, Krathwohl, & Miller, 1963). Kagan and Krathwohl (1967) then turned their efforts to improving the supervision counselors-in-training received in graduate programs. They wanted to find a method for supervisors to process trainees' recorded sessions that did not rely on supervisor interpretations, as they believed that counselors-in-training could not assess their underlying thoughts and feelings about the counseling relationship during the counseling session. Building on the work of Bloom (1954), Kagan and Krathwohl hypothesized that replaying the recorded counseling session in a supervision session would allow the clinician to relive the experience and explore the underlying thoughts and feelings in deeper detail. They settled on the use of video recordings as the best avenue, as video allows the clinician to see both verbal and nonverbal cues.

Reliability

Because of the subjective and qualitative nature of IPR, its reliability is difficult to determine. IPR was originally intended for use in mental health professions, but has also been applied successfully in the medical field (Jason, Kagan, Werner, Elstein, & Thomas, 1971), law enforcement (Danish & Brodsky, 1970), and other areas (Kagan & Burke, 1976). Although these studies did not explicitly examine reliability, IPR's success across disciplines helps build a case for its consistency as a tool for those working with interpersonal situations. As an aid to the standardization of IPR training, the authors developed an instructor's manual and a video training series (Kagan & Kagan, 1990).

Validity

Kagan and Kagan (1997) noted that most studies assessing evidence for IPR's validity have evaluated its impact using a pre- and postevaluation model. Katz and Resnikoff (1977), however, asked individuals to provide an in-the-moment account of their feelings during an interpersonal situation and then again in a videotape recall session of that situation; the correlation between the in-the-moment account and the recall account was significant in all four of their experimental groups. Similarly, Archer and colleagues (1972) collected physiological feedback data on individuals as they watched vignettes and then again as they watched their video playback. The researchers noted that the individuals frequently exhibited the same physiological states in the recall session (sweating and increased heart rate) that they exhibited during the initial session; these physiological states were not altered until the inquirer helped the individual explore the situation and gain insight or awareness.

Availability

IPR is primarily a nonstandardized process that is discussed in most books about counseling or psychotherapy supervision. Kagan and Kagan (1995) also have released a video training series on IPR through Microtraining Associates, and these videos can be accessed within several academic video databases. The training series provides an overview of IPR's theoretical framework, strategies and techniques to be used in IPR, and case examples. Moreover, several authors have included extensive lists of IPR questions, or leads (Bernard & Goodyear, 2014; Kagan, 1980b).

Sample Studies

Several studies have been conducted by researchers outside of the original IPR team to evaluate outcomes of IPR or to compare IPR to other methods (Crews *et al.*, 2005; Spivack, 1972; West & Clark, 2004). Crews and colleagues (2005) found statistically significant score improvements on the Skilled Counseling Scale for both IPR-trained counselors ($t_8 = 4.58$, $p < .01$, $d = 1.53$) and counselors trained using Smaby, Maddux, Torres-Rivera, and Zimmick's (1999) Skilled Counselor Training Model (SCTM; $t_{46} = 55.24$, $p < .001$, $d = 8.05$). However, the authors did note that SCTM focuses more on skill development than does IPR, which explains why SCTM outgained IPR. Spivack (1972) found that pre-practicum counselor education students trained using IPR had higher ratings on the Counselor Verbal Response Scale (CVRS; Kagan, 1967) and the Empathic Understanding in Interpersonal Processes Scale (EUS; Truax &

Carkhuff, 1967) than students who were trained using only lectures, discussions, and demonstrations. In a qualitative study, West and Clark (2004) observed that using IPR in counseling supervision strengthened the supervisory relationship and produced moments of insight for both the supervisor and supervisee. Baker, Daniels, and Greeley (1990), in a meta-analytic review ($k = 10$, $N = 295$), found a positive, albeit small, average effect ($d = .20$) across studies of IPR. Kagan and Kagan (1990) criticized this meta-analysis for its exclusion of a study by Boltuch (1975) but did not indicate how the meta-analysis results would have changed.

Critique

Based on the number of studies indicating IPR's effectiveness, its use is warranted in clinical relationships to stimulate insight and awareness of internal processes and feelings. Although IPR techniques and strategies are discussed in counseling supervision literature and textbooks, the training modules by Kagan and Kagan (1995) offer the fullest preparation to use IPR. Moreover, those who wish to employ IPR in their practice are encouraged to seek supervision from someone with training and experience. It is also noteworthy that Kagan and Kagan (1990) have clarified that IPR training is not a skill acquisition model, as it has been portrayed in a few studies (Baker *et al.*, 1990; Crews *et al.*, 2005). Although IPR is highly adaptable to a variety of situations, helping professionals, educators, and researchers should take advantage of the training modules and available research to ensure that they are employing IPR as it was intended to be used.

References

Archer, J., Jr., Feister, T., Kagan, N. I., Rate, R., Spierling, T., & Van Noord, R. (1972). A new methodology for education, treatment and research in human interaction. *Journal of Counseling Psychology*, *19*, 275–281. doi:10.1037/h0033070

Baker, S. B., Daniels, T. G., & Greeley, A. T. (1990). Systematic training of graduate-level counselors: Narrative and meta-analytic reviews of three major programs. *The Counseling Psychologist*, *18*, 355–421. doi:10.1177/0011000090183001

Bernard, J. M., & Goodyear, R. K. (2014). *Fundamentals of clinical supervision* (5th ed.). Upper Saddle River, NJ: Pearson.

Bloom, B. S. (1954). The thought process of students in discussion. In S. J. French (Ed.), *Accent on teaching: Experiment in general education* (pp. 23–46). New York: Harper.

Boltuch, B. S. (1975). The effects of a pre-practicum skill training program, influencing human interaction, on developing counselor effectiveness in a master's level practicum (Unpublished doctoral dissertation). New York University, New York.

Borders, L. D., & Brown, L. L. (2005). *The new handbook of counseling supervision*. Mahwah, NJ: Lahaska Press.

Cashwell, C. S. (1994). *Interpersonal process recall*. Retrieved from ERIC Digest (EDO-CG-94-10).

Clarke, P. (1997). Interpersonal process recall in supervision. In G. Shipton (Ed.), *Supervision of psychotherapy and counselling: Making a place to think* (pp. 93–104). Buckingham, UK: Open University Press.

Crews, J., Smith, M. R., Smaby, M. H., Maddux, C. D., Torres-Rivera, E., Casey, J. A., & Urbani, S. (2005). Self-monitoring and counseling skills: Skills-based versus interpersonal process recall training. *Journal of Counseling & Development*, *83*, 78–85. doi:10.1002/j.1556-6678.2005.tb00582.x

Danish, S., & Brodsky, S. (1970). Psychology in action: Training of policemen in emotional control and awareness. *The American Psychologist*, *25*, 368–369. doi:10.1037/h0029496

Jason, H., Kagan, N. I., Werner, A., Elstein, A., & Thomas, J. B. (1971). New approaches to teaching basic interview skills to medical students. *American Journal of Psychiatry*, *127*, 1404–1407. doi:10.1176/ajp.127.10.1404

Kagan, H. K., & Kagan, N. I. (1997). Interpersonal process recall: Influencing human interaction. In C. E. Watkins, Jr. (Ed.), *Handbook of psychotherapy supervision* (pp. 296–309). New York, NY: Wiley.

Kagan, N. I. (1967). *Studies in human interaction* (Final Report, Grant No. OE7-32-0410-270). East Lansing, MI: Michigan State University, Educational Publication Services.

Kagan, N. I. (1976). *Influencing human interaction*. Mason, MI: Mason Media.

Kagan, N. I. (1980a). Influencing human interaction: Eighteen years with IPR. In A. K. Hess (Ed.), *Psychotherapy supervision: Theory, research, & practice* (pp. 262–286). New York, NY: Wiley.

Kagan, N. I. (1980b). *Interpersonal Process Recall: A method of influencing human interaction*. Houston, TX: Mason Media.

Kagan, N. I., & Burke, J. B. (1976). *Influencing human interaction using interpersonal process recall (IPR): A student manual*. East Lansing, MI: Michigan State University.

Kagan, N. I., & Kagan, H. K. (1990). IPR: A validated model for the 1990s and beyond. *The Counseling Psychologist*, *18*(3), 436–440. doi:10.1177/0011000090183004

Kagan, N. I., & Kagan, H. K. (1995). *Interpersonal process recall series [Video]*. Available from http://www.academicvideostore.com

Kagan, N. I., & Krathwohl, D. R. (1967). *Studies in human interaction: Interpersonal process recall stimulated by videotape*. East Lansing, MI: Michigan State University.

Kagan, N. I., Krathwohl, D. R., & Miller, R. (1963). Stimulated recall in therapy using video tape: A case study. *Journal of Counseling Psychology*, *10*, 237–243. doi:10.1037/h0045497

Katz, D., & Resnikoff, A. (1977). Televised self-confrontation and recalled affect: A new look at videotape recall. *Journal of Counseling Psychology*, *24*, 150–152. doi:10.1037/0022-0167.24.2.150

Larsen, D., Flesaker, K., & Stege, R. (2008). Qualitative interviewing using interpersonal process recall: Investigating internal experiences during professional-client conversations. *International Journal of Qualitative Methods*, *7*, 18–37.

Macaskie, J., Lees, J., & Freshwater, D. (2015). Talking about talking: Interpersonal process recall as an intersubjective approach to research. *Psychodynamic Practice*, *21*, 226–240. doi:10.1080/14753634.2015.1042517

McQuellon, R. P. (1982). Interpersonal Process Recall. In E. K. Marshall & P. D. Kurtz (Eds.), *Interpersonal helping skills* (pp. 161–202). San Francisco, CA: Jossey-Bass.

Smaby, M. H., Maddux, C. D., Torres-Rivera, E., & Zimmick, R. (1999). A study of the effects of a skills-based versus a conventional group counseling training program. *The Journal for Specialists in Group Work*, *24*, 152–163. doi:10.1080/01933929908411427

Spivack, J. D. (1972). Laboratory to classroom: The practical application of IPR in a master's level prepracticum counselor education program. *Counselor Education and Supervision*, *12*, 3–16. doi:10.1002/j.1556-6978.1972.tb01532.x

Spivack, J. D. (1974). Interpersonal process recall: Implications for psychotherapy. *Psychotherapy: Theory, Research, and Practice*, *11*, 235–238. doi:10.1037/h0086347

Truax, C., & Carkhuff, R. (1967). *Toward effective counseling and psychotherapy*. Chicago: Aldine.

West, W., & Clark, V. (2004). Learnings from a qualitative study into counselling supervision: Listening to supervisor and supervisee. *Counseling and Psychotherapy Research*, *4*, 20–26. doi:10.1080/14733140412331383903

Profile 28

Interpersonal Reactivity Index (IRI)

(Davis, 1980)

Profiled by: Shaughan A. Keaton

Young Harris College

Construct

Davis (1983) defined *empathy* as the "reactions of one individual to the observed experiences of another" (p. 113). It consists of a set of facets that reflect individual cognitive and emotional feelings of warmth, compassion, concern for others, and personal feelings of anxiety and discomfort from observing others' negative experiences (Davis, 1980).

Instrument Type

Self-Report

Description

The Interpersonal Reactivity Index (IRI) is a 28-item scale that features four, seven-item subscales: the Fantasy Scale (FS), Perspective-Taking Scale (PT), the Empathic Concern Scale (EC), and the personal distress scale (PD). EC refers to individuals' feelings of compassion and concern for others. PT assesses unplanned attempts to adopt others' points of view. FS describes the likelihood that a person identifies with a fictional character. PD indicates the extent that an individual feels uneasiness or worry when exposed to the negative experiences of others.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

Participants are asked to consider the extent to which each of 28 statements describes them, ranging from 0 (*does not describe me well*) to 4 (*describes me very well*). This survey can be administered via paper or online and takes approximately 10 minutes to complete. If researchers are interested in only a subset of the empathy facets, those subscales can be used in place of the entire scale.

Scoring

Responses to each seven-item subscale are averaged to reveal aggregate scores for each subdimension.

Development

To address limitations with a variety of other empathy measures, Davis (1980) developed the IRI. The two primary goals were to develop a scale (a) that was easy to administer and (b) that would capture individual differences in cognitive and emotional reactions. Davis (1980) began with 50 items, some borrowed from other measures such as Mehrabian and Epstein's (1972) Emotional Empathy Scale and the Fantasy-Empathy Scale (Stotland, Mathews, Sherman, Hansson, & Richardson, 1978), whereas others were created from scratch. He then conducted a Jöreskog Factor Analysis with oblique rotation on the items for males and females separately. The results revealed the same four dimensions for both male and female samples: fantasy items, perspective-taking items, empathic concern items, and personal distress items.

A second, 45-item measure was formed, which was tested again using a Jöreskog Factor Analysis with oblique rotation on the items (again, for men and women separately). The final result was a four-factor, 28-item scale composed of items that loaded heavily for both sexes. This version of the IRI was again factor analyzed on an independent sample with the same procedure (again confirming the four-factor structure). The final subscales were termed the Fantasy Scale (FS), Perspective-Taking Scale (PT), the Empathic Concern Scale (EC), and the Personal Distress Scale (PD).

Reliability

In Davis's original report (1980), the internal consistencies of the subscales were moderate for males (FS, $\alpha = .78$; PT, $\alpha = .75$; EC, $\alpha = .71$; PD, $\alpha = .78$) and for females (FS, $\alpha = .75$; PT, $\alpha = .78$; EC, $\alpha = .70$; PD, $\alpha = .78$). In all of the other studies included in this profile, the alphas have ranged as follows: (FS, $.63 < \alpha < .84$; PT, $.65 < \alpha < .81$; EC, $.65 < \alpha < .82$; PD, $.57 < \alpha < .82$).

Davis (1980) submitted the 45-item version of the IRI to a test–retest reliability procedure. Data from both males and females displayed adequate temporal stability for each subscale (.61 to .79 for males; .62 to .81 for females).

Validity

Davis (1983) provided validity evidence for the IRI's subscales by comparing them to previous cognitive and emotional measures of empathy, interpersonal functioning, self-esteem, emotionality, sensitivity to others, and intelligence. The PT scale was consistently associated with measures of interpersonal functioning. Those with higher PT scores reported less social dysfunction, more social competence, higher self-esteem, and a more selfless interest in others' feelings and reactions. FS scores were positively related with measures of verbal intelligence and emotional reactivity. EC scores were positively correlated with shyness and anxiety, and negatively related to an undesirable communication style (boastfulness and egotism), self-esteem, emotionality, and an unselfish concern for others. These associations were mostly small to moderate in magnitude ($-.54 < r < .56$). PD scores were strongly related to lower self-esteem, poor interpersonal skills, vulnerability, uncertainty, and fearfulness. The IRI, for the most part, was moderately associated with other measures of empathy ($.11 < r < .63$).

Cliffordson (2001) acquired evidence of convergent validity for the IRI (noting that it was subtly revised) by displaying that the concept of empathy is identical to empathic concern when comparing students' and parents' personality judgments.

In terms of construct validity, model fit estimates were not provided in Davis's original study, but the structures were repeatedly factor-analyzed using a Jöreskog Factor Analysis with oblique rotation, producing a consistent structure. The structural invariance of Davis's IRI model was examined between adolescents and their mothers (Hawk *et al.*, 2013). The scale was submitted to a confirmatory factor analysis (CFA) and found to demonstrate psychometric invariance and structural validity with comparative fit indices (CFI) ranging from .956 to .962, root mean square errors of approximation (RMSEA) ranging from .049 to .065, and the standardized root mean square residual (SRMR) ranging from .037 to .050. These estimates all fall within commonly accepted CFA model fit parameters.

Pulos, Elison, and Lennon (2004) examined the hierarchical structure of Davis's IRI, confirming a first-order, four-factor structure consistent with Davis's subscales and two second-order orthogonal factors of general empathy (EC, FS, and PT positively loaded on this dimension) and emotional control (PT positively loaded on this dimension, and FS and PD negatively).

The IRI has been translated into Spanish (Fernández, Dufey, & Kramp, 2011; Pérez-Albéniz, De Paúl, Etxeberría, Montes, & Torres, 2003), French (Gilet, Mella, Studer, Grühn, & Labouvie-Vief, 2013), German (Lauterbach & Hosser, 2007; Paulus, 2009), and Chinese (Chiang, Hua, Tam, Chao, & Shiah, 2014; Dong & Wang, 2010). Fernández and colleagues (2011) provided cross-cultural evidence of structural and predictive validity of both Davis's and Cliffordson's models of the IRI using structural equation model estimates. They found that both models exhibited acceptable fit, although the CFI estimates did not meet common standards (Davis: $\chi^2 = 781.74$ (344), $p < .001$, RMSEA = 0.054, CFI = 0.813, SRMR = 0.070; and Cliffordson: $\chi^2 = 810.34$ (346), $p < .001$, RMSEA = 0.056, CFI = 0.802, SRMR = 0.075). Gilet *et al.* (2013) also confirmed the factor structure of the IRI, although again with lower CFI than commonly accepted, $\chi^2(344) = 789$, $p < .01$, $\chi^2/df = 2.29$, CFI = .81, RMSEA = .065, 90% CI (.06, .07), SRMR = .07. No model fit estimates were provided for the German version.

Availability

The 28-item English version of the IRI is presented below (Davis, 1980) and is free to use for research purposes.

Sample Studies

Davis's IRI is arguably the most utilized measure of its kind. As such, this section can be no more than a sampling of available studies. Most of the heavily cited works provide examinations of the validity of the IRI as outlined above.

Particularly relevant for listening research, Riggio and Taylor (2000) tested the relationship between empathy and social skills; PT and EC were positively related with several self-reported social skills and negatively correlated with dogmatism. Kimmes, Edwards, Wetchler, and Bercik (2014) researched empathy in dyadic relationships, finding that perceived dyadic empathy and empathic congruity predicted relationship satisfaction.

Van Doesum, Van Lange, and Van Lange (2013) discussed how social mindfulness is positively related to empathy. PT ($r = .28$), EC ($r = .21$), and FS ($r = .17$) were related to social mindfulness, but as expected PD was not. Delič, Novak, Kovačič, and Avsec (2011) found PT and EC ($r = -.19$) negatively related to narcissism and positively associated with emotional intelligence (PT, $r = .23$; EC, $r = .17$) and the processing of social information (PT, $r = .29$; EC, $r = .22$). PT was also related to greater social skills ($r = .18$) and social awareness ($r = .35$). McLellan and McKinlay (2013) revealed that deficits in social skills are long lasting in patients who suffered traumatic brain injuries as children, whereas Saxton, Younan, and Lah (2013) found that those with traumatic brain injuries experienced impairments in aspects of emotional perception (empathy and PT) contributing to difficulties in social behaviors.

Lauterbach and Hosser (2007) investigated the relationship between empathy and whether prisoners were violent or nonviolent offenders. They found that (lack of) PT predicted a greater likelihood of future violent offenses within 2 years of a prisoner's release. They also reported that more aggression in prisoners corresponded with lower FS, PT, and EC scores. Loneliness was related to increased FS and PD; self-efficacy positively associated with PT and negatively with PD; and self-esteem negatively correlated with PD.

Critique

The IRI has ample evidence of psychometric validity, as outlined in this profile. In addition, the scale has been used in experimental medical research dealing with behaviors and cognition. For instance, it has been used to investigate traumatic brain injury patients' abilities to assess or detect emotional stimuli (McLellan & McKinlay, 2013; Saxton *et al.*, 2013). The IRI also has been extensively used in psychiatry and neuroscience, particularly in the study of personality disorders such as schizophrenia (for some recent examples of many studies on this topic, see Chiang *et al.*, 2014; Fujino *et al.*, 2014; Lehmann *et al.*, 2014; Michaels *et al.*, 2014).

Researchers examining the role of listening in several types of interactions—such as social support (e.g., Burleson, 1983), social or private relationships (e.g., Haas, Anderson, & Filkowski, 2015; Kimmes *et al.*, 2014), social skills and behaviors (e.g., Delič *et al.*, 2011; Van Doesum *et al.*, 2013; You, Kim, & No, 2015), or personality and personality disorders (e.g., Chiang *et al.*, 2014; Delič *et al.*, 2011; Fujino *et al.*, 2014; Lehmann *et al.*, 2014; Michaels *et al.*, 2014)—can confidently administer the scale. Like all self-report scales, the IRI potentially suffers from reporting biases.

References

Burleson, B. R. (1983). Social cognition, empathic motivation, and adults' comforting strategies. *Human Communication Research*, *10*, 295–304. doi:10.1111/j.1468-2958.1983.tb00019.x

Chiang, S.-K., Hua, M.-S., Tam, W.-C. C., Chao, J.-K., & Shiah, Y.-J. (2014). Developing an alternative Chinese version of the interpersonal reactivity index for normal population and patients with schizophrenia in Taiwan. *Brain Impairment*, *15*, 120–131. doi:10.1017/BrImp.2014.15

Cliffordson, C. (2001). Parents' judgments and students' self-judgments of empathy: The structure of empathy and agreement of judgments based on the interpersonal reactivity index (IRI). *European Journal of Psychological Assessment*, *17*, 36. doi:10.1027//1015-5759.17.1.36

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 1–19.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113. doi:10.1037/0022-3514.44.1.113

Delič, L., Novak, P., Kovačič, J., & Avsec, A. (2011). Self-reported emotional and social intelligence and empathy as distinctive predictors of narcissism. *Psychological Topics*, *20*, 477–488.

Dong, Y., & Wang, K. (2010). Chinese version of interpersonal reactivity index (IRI) reliabilty and validity/中文版人际反应指针量表 (IRI-C) 的信度及效度研究. *Chinese Journal of Clinical Psychology/中国临床心理学杂志*, *2*, 155–157.

Fernández, A. M., Dufey, M., & Kramp, U. (2011). Testing the psychometric properties of the Interpersonal Reactivity Index (IRI) in Chile: Empathy in a different cultural context. *European Journal of Psychological Assessment*, *27*, 179. doi:10.1027/1015-5759/a000065

Fujino, J., Takahashi, H., Miyata, J., Sugihara, G., Kubota, M., Sasamoto, A., … Murai, T. (2014). Impaired empathic abilities and reduced white matter integrity in schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *48*, 117–123. doi:10.1016/j.pnpbp.2013.09.018

Gilet, A.-L., Mella, N., Studer, J., Grühn, D., & Labouvie-Vief, G. (2013). Assessing dispositional empathy in adults: A French validation of the Interpersonal Reactivity Index (IRI). *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *45*, 42–48. doi:10.1037/a0030425

Haas, B. W., Anderson, I. W., & Filkowski, M. M. (2015). Interpersonal reactivity and the attribution of emotional reactions. *Emotion*, *15*, 390. doi:10.1037/emo0000053

Hawk, S. T., Keijsers, L., Branje, S. J., Graaff, J. V. d., Wied, M. d., & Meeus, W. (2013). Examining the interpersonal reactivity index (IRI) among early and late adolescents and their mothers. *Journal of Personality Assessment*, *95*, 96–106. doi:10.1080/00223891.2012.696080

Kimmes, J. G., Edwards, A. B., Wetchler, J. L., & Bercik, J. (2014). Self and other ratings of dyadic empathy as predictors of relationship satisfaction. *The American Journal of Family Therapy*, *42*, 426–437. doi:10.1080/01926187.2014.925374

Lauterbach, O., & Hosser, D. (2007). Assessing empathy in prisoners: A shortened version of the Interpersonal Reactivity Index. *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie*, *66*, 91. doi:10.1024/1421-0185.66.2.91

Lehmann, A., Bahçesular, K., Brockmann, E.-M., Biederbick, S.-E., Dziobek, I., Gallinat, J., & Montag, C. (2014). Subjective experience of emotions and emotional empathy in paranoid schizophrenia. *Psychiatry Research*, *220*, 825–833. doi:10.1016/j.psychres.2014.09.009

McLellan, T., & McKinlay, A. (2013). Sensitivity to emotion, empathy and theory of mind: adult performance following childhood TBI. *Brain Injury*, *27*, 1032–1037. doi:10.3109/02699052.2013.794965

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of personality*, *40*, 525–543. doi:10.1111/j.1467-6494.1972.tb00078.x

Michaels, T. M., Horan, W. P., Ginger, E. J., Martinovich, Z., Pinkham, A. E., & Smith, M. J. (2014). Cognitive empathy contributes to poor social functioning in schizophrenia: Evidence from a new self-report measure of cognitive and affective empathy. *Psychiatry Research*, *220*, 803–810. doi:10.1016/j.psychres.2014.08.054

Paulus, C. (2009). *Saarbrücken personality questionnaire SPF (IRI) for measuring empathy: Psychometric evaluation of the German version of the Interpersonal Reactivity Index.* Saarbrücken, Germany: Universität des Saarlandes.

Pérez-Albéniz, A., De Paúl, J., Etxeberría, J., Montes, M. P., & Torres, E. (2003). Adaptación de interpersonal reactivity index (IRI) al español. *Psicothema*, *15*, 267–272.

Pulos, S., Elison, J., & Lennon, R. (2004). The hierarchical structure of the Interpersonal Reactivity Index. *Social Behavior and Personality: An International Journal*, *32*, 355–359. doi:10.2224/sbp.2004.32.4.355

Riggio, R. E., & Taylor, S. J. (2000). Personality and communication skills as predictors of hospice nurse performance. *Journal of Business and Psychology*, *15*, 351–359. doi:10.1023/A:1007832320795

Saxton, M. E., Younan, S. S., & Lah, S. (2013). Social behaviour following severe traumatic brain injury: Contribution of emotion perception deficits. *NeuroRehabilitation*, *33*, 263–271.

Stotland, E., Mathews, K., Sherman, S., Hansson, R., & Richardson, B. (1978). *Empathy, fantasy, and helping*. Beverly Hills, CA: Sage.

Van Doesum, N. J., Van Lange, D. A., & Van Lange, P. A. (2013). Social mindfulness: Skill and will to navigate the social world. *Journal of Personality and Social Psychology*, *105*, 86. doi:10.1037/a0032540

You, S., Kim, E., & No, U. (2015). Impact of violent video games on the social behaviors of adolescents: The mediating role of emotional competence. *School Psychology International*, *36*, 94–111. doi:10.1177/0143034314562921

Scale

Interpersonal Reactivity Inded (Davis, 1980, 1983)

Source: Davis (1980, 1983).

Please indicate the extent that each statement describes you, using the following scale:

0	1	2	3	4
(does not describe me well)				(describes me very well)

<u>Fantasy Scale</u>

1) When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me.
2) I really get involved with the feelings of the characters in a novel.
3) I am usually objective when I watch a movie or play, and I don't often get completely caught up in it. (R)
4) After seeing a play or movie, I have felt as though I were one of the characters.
5) I daydream and fantasize, with some regularity, about things that might happen to me.
6) Becoming extremely involved in a good book or movie is somewhat rare for me. (R)
7) When I watch a good movie, I can very easily put myself in the place of a leading character.

<u>Perspective-Taking Scale</u>

8) Before criticizing somebody, I try to imagine how I would feel if I were in their place.
9) If I'm sure I'm right about something, I don't waste much time listening to other people's arguments. (R)
10) I sometimes try to understand my friends better by imagining how things look from their perspective.
11) I believe that there are two sides to every question and try to look at them both.
12) I sometimes find it difficult to see things from the other person's point of view. (R)
13) I try to look at everybody's side of a disagreement before I make a decision.
14) When I'm upset at someone, I usually try to "put myself in their shoes" for a while.

<u>Empathic Concern Scale</u>

15) When I see someone being taken advantage of, I feel kind of protective toward them.
16) When I see someone being treated unfairly, I sometimes don't feel very much pity for them. (R)
17) I often have tender, concerned feelings for people less fortunate than me.
18) I would describe myself as a pretty soft-hearted person.
19) Sometimes I don't feel sorry for other people when they are having problems. (R)
20) Other people's misfortunes do not usually disturb me a great deal. (R)
21) I am often quite touched by things that I see happen.

Personal Distress Scale

22) When I see someone who badly needs help in an emergency, I go to pieces.
23) I sometimes feel helpless when I am in the middle of a very emotional situation.
24) In emergency situations, I feel apprehensive and ill-at-ease.
25) I am usually pretty effective in dealing with emergencies. (R)
26) Being in a tense emotional situation scares me.
27) When I see someone get hurt, I tend to remain calm. (R)
28) I tend to lose control during emergencies.

Note: Labels should be removed and items randomly ordered prior to administration. After reverse scoring, scores from each subscale are averaged.

Profile 29

Language Style Matching (LSM)

(Ireland & Pennebaker, 2010)

Profiled by: Kaitlin Cannava, PhD

Visiting Assistant Professor, Xavier University

Construct

Language style matching (LSM) (Ireland & Pennebaker, 2010) was designed to capture the amount of conversational involvement or "psychological similarity" between two people. There are several existing measures of dyadic language coordination (DLC), the degree to which two individuals within an interaction (written or spoken) match on certain language variables, and LSM is one of several variables used to capture and conceptualize coordination.

Instrument Type

Behavioral Assessment

Description

LSM captures the coordination of language during a conversation as the degree of function word matching between two or more interlocutors. Function words fall into nine word categories, namely auxiliary verbs, articles, common adverbs, personal pronouns, indefinite pronouns, prepositions, negations, conjunctions, and quantifiers. These words have no lexical meaning but express grammatical and structural relationships between other words. For example, the word *her* is only understood as an anaphoric reference, or a reference to something that has previously been stated; *her* refers to a specific reference at a prior point in the conversation. If two people are talking about a person named Gina, instead of saying *Gina* over and over, the two people can start referring to Gina as *her*.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

LSM only measures the matching of function words in a conversation and ignores content words. *Gina* would not be recognized in the LSM calculation, but *her* would be recognized. When two individuals use function words in a similar manner in their talk (or another language medium like writing), that similarity is said to be a signal of similar styles of thinking (Ireland & Pennebaker, 2010). LSM also is thought to reflect engagement attempts between partners in a conversation and the degree to which those attempts are reciprocated. In general, LSM might suggest that conversational partners are listening to one another on a fundamental level; it would make sense that people who are engaged with each other on a topic would speak about the topic in the same way.

LSM can be analyzed through any set of texts. Written texts such as emails or text messages, spoken texts such as transcripts of conversation, group talk, and even numerous texts from the same person such as journal entries are all potential data for LSM.

Administration

To compute an LSM score, a researcher needs a set of texts. These texts can be part of a face-to-face or mediated conversation that is synchronous or asynchronous. The texts also can be transcribed from the spoken language of interlocutors or can be written (e.g., emails). After gathering text samples, an LSM score is computed after running the transcripts through a computerized textual analysis program like the Linguistic Inquiry and Word Count (LIWC) (Pennebaker, Booth, Boyd, & Francis, 2015; Pennebaker, Booth, & Francis, 2007). LIWC provides the researcher with data about how often certain types of words (including function words) were used (e.g., 10% of Person A's email were I-Pronouns). Both LIWC and LSM are considered unobtrusive ways to measure language.

Scoring

LSM is calculated as the percentage of words that occur in a text. Counting numerous categories of words by hand has become obsolete with the advent of computerized textual analysis software programs such as LIWC. LIWC was designed to provide information about 80 variables, including the nine function word categories used to calculate LSM.

The LSM score for a dyad or group is calculated by taking the absolute value of the difference between speakers and then dividing by the total for each category. For example, if we were calculating the degree of matching between two interacting speakers, the equation would look like:

$$\frac{1-(|\text{Person 1's function words}-\text{Person 2's function words}|)}{(\text{Person 1's function words}+\text{Person 2's function words}+.001)}$$

This score ranges between 0 and 1. Scores of .60 reflect relatively low synchrony, and .85 or above represents high synchrony (Ireland & Pennebaker, 2010).

Development

LSM was first introduced as Linguistic Style Matching (Niederhoffer & Pennebaker, 2002), a measure of "clicking" and rapport with another person. The Niederhoffer and Pennebaker study used informal chat rooms and the Watergate transcripts as data. It was originally thought that one person primed another during conversation to use particular types of words. This original measure did not initially use the LSM calculation shown above; instead, they correlated the degree to which one person used a "comparable number of words and types of words as the other person" (Niederhoffer & Pennebaker, 2002, p. 344). The original calculation of a matching score was based on an overall conversational level and on a turn-by-turn level. Since this study, LSM has been refined to the nine function word categories and has been assessed almost exclusively at the overall conversational level; the turn-by-turn level of analysis has been used much less frequently.

LSM, now known as Language Style Matching, was introduced with the new name and new calculation in a study of speed dating and relationship maintenance (Ireland *et al.*, 2011). Since that study, LSM has not been changed with respect to its calculation, types of words used, or any measurement criteria, at least as of the writing of this profile.

Reliability

Because LSM is an observed variable and is computed from data obtained by the automated word count function of LIWC, its reliability can be said to rest on two criteria. First, there is the question of whether function words are reliably counted across various types of transcripts. Research published by Pennebaker and colleagues on the LIWC program shows this to be the case (Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007). The second question has to do with the psychometrics of LSM. Previous studies have consistently shown that the LSM of each function word category is positively correlated with the LSM of every other function word category (Chung & Pennebaker, 2015). This internal consistency means that if a dyad matches in one function word category, it likely matches in the other eight function word categories (Ireland & Pennebaker, 2010). Using Cronbach's alpha, reliability estimates have ranged from .49 to .80 (Ireland & Pennebaker, 2010; Ireland *et al.*, 2011).

Validity

In terms of convergent validity, there has been little research on how LSM correlates with other measures of coordination. Babcock, Ta, and Ickes (2014) found that LSM was empirically distinct from Latent Semantic Similarity (LSS) (Landauer, Foltz, & Laham, 1998). The LSS index is a measure of how similar two speech samples are in the words that were used (the semantic space) and how similar those words were used with other words. LSS analyzes how certain words group together, similar to Latent Semantic Analysis (LSA) (see http://lsa.colorado.edu/), but also calculates the degree to which groups of words match the contributions of another speaker. The correlation between LSS and LSM "was significant but not large ($r = .35$, $p < .01$)" (Babcock *et al.*, 2014, p. 5), suggesting that these two indices are related but not isomorphic. This conclusion was further validated by Babcock *et al.* (2014), who showed that LSS and LSM correlated differently with dyad-level behaviors and post-interaction measures. In particular, LSS was positively and strongly correlated with behaviors thought to be indicative of interest and involvement in a conversation, such as total number of speaking turns ($r = .59$), total number of self-disclosures ($r = .48$), and total number of verbal acknowledgments ($r = .48$). LSS also was correlated with the perception of involvement ($r = .32$). On the other hand, LSM was only correlated with three variables, specifically, the average of self-esteem scores ($r = .34$), feeling a need to communicate ($r = .33$), and the percentage of third-person pronouns ($r = .37$). Babcock *et al.* (2014) interpreted their results as showing LSM was high when people did not want to talk to each other, when they talked about other people more, and when they had low self-esteem, whereas LSS scores were high when both parties perceived the other to be involved in the conversation and were actively behaving in a way to further the conversation. Published studies have not yet compared LSM to other DLC measures.

The predictive validity of LSM has been demonstrated in studies showing this measure is related to empathic behavior (Ireland & Pennebaker, 2010), relationship initiation and stability (Ireland *et al.*, 2011), and relationship closeness (Ireland & Pennebaker, 2010; Ireland *et al.*, 2011). LSM is thought to be a construct of how two people have similar psychological states or similar thinking styles. The research done by Pennebaker and colleagues restricted LSM research to identify relational closeness, but his team is now attempting to use LSM in a new way. More recently, Pennebaker and colleagues have attempted to use LSM as a measure of group cohesion and group dynamics (Chung & Pennebaker, 2015). Because previous research on LSM is a marker of two people coordinating, LSM can also be applied to groups of people and how they talk together to accomplish goals, to communities of people, and even to cultural dynamics.

Availability

LIWC can be purchased on www.liwc.net. This program gives researchers the raw data needed to calculate LSM scores. Depending on the typical size of data files that a researcher utilizes, she can opt for the full version (LIWC2007 or LIWC2015) or a version with limited features (LIWClite7).

Sample Studies

LSM has been applied to track social and group processes. LSM has been used as a measure to predict outcomes in heterosexual speed-dating scenarios (Ireland *et al.*, 2011), showing that high LSM scores predict post-interaction liking and romantic stability three months later. LSM also shows that matching of function words is related to group cohesion (Gonzales, Hancock, & Pennebaker, 2010) and cooperation (Scissors, Gill, Geraghty, & Gergle, 2009) although a new study suggests that LSM was indicative of couples who are unable to cooperate or reach an agreement during a negotiation (Ireland & Henderson, 2014). More recently, LSM was shown to predict reported emotional improvement and cognitive reappraisal of a distressed other talking about an upsetting event (Cannava & Bodie, 2016). When a supportive listener and a distressed discloser matched function words, this coordination made the distressed person feel better and think differently about the situation.

Critique

LSM has been studied mostly using writing samples; only a few studies have explored LSM within actual conversations (Cannava & Bodie, 2016; Gonzales *et al.*, 2010; Taylor & Thomas, 2008). The focus on writing is surprising given that LSM is thought to reflect engagement attempts between partners in a conversation and the degree to which those attempts are reciprocated. Most of the initial studies on LSM were analyzed using journal entries, poems, or speeches; basically, LSM was not originally a conversation-based measure.

More recently, scholars using LSM are turning to conversation-based data (e.g., Cannava & Bodie, 2016). Not only has LSM not been studied consistently in conversation, but also one study suggests that LSM is not a conceptual measure of conversational involvement. Babcock *et al.* (2014) found that LSM was highest in conversations when members talked about other people more (using third-person pronouns) and were more disinclined to talk to each other. Babcock *et al.* (2014) interpreted their results by suggesting that when people are reluctant to talk to each other, they will "mindlessly" match others to accomplish a conversation.

Another critique of LSM is that it is just one out of many options to quantify and measure dyadic language coordination. Other measures such as language style synchrony (LSS; Lord, Sheng, Imel, Baer, & Atkins, 2014), latent semantic analysis (LSA; Landauer *et al.*, 1998), and alignment (Du Bois, 2007) are all ways to analyze language coordination. Each operationalization places value on different parts of speech, and researchers should take into consideration how dyads are likely to coordinate when creating a study. For example, LSM places value on only function words (pronouns, articles, common adverbs, etc.) rather than LSS and LSA (all semantic categories) or alignment (phrases).

Not only does LSM place value on specific types of words, but also the output and interpretation of an LSM score might be a bit misleading. The LSM score reflects how often each person uses the nine categories of function words and compares that frequency to another person. For example, if Person A says, "I drank coffee yesterday afternoon," and Person B says, "My sister ate pizza today," the LSM score would be 100%. Each person used 20% pronouns (Person A: "I" and Person B: "My") and 0% in all other function word categories. Therefore, each category of words is 100% matching because if both people use the same percentage of words in a category, then that category gets a 100% matching score. By this logic, a high LSM score can represent people *not* matching their language style—that people can match by not matching. This could be a problem for how we conceptualize matching, coordination, and interpret results about what exactly LSM is and how it functions to predict interpersonal outcomes. Other measures might lend themselves to better calculate what it means to coordinate linguistically.

References

Babcock, M. J., Ta, V. P., & Ickes, W. (2014). Latent semantic similarity and language style matching in initial dyadic interactions. *Journal of Language and Social Psychology*, 33, 78–88. doi:10.1177/0261927X13499331

Cannava, K., & Bodie, G. D. (2016). Language use and style matching in supportive conversations between strangers and friends. *Journal of Social and Personal Relationships*. Advance online publication. doi:10.1177/0265407516641222 DOI:10.1177%2 F0265407516641222

Chung, C. K., & Pennebaker, J. W. (2015). Using computerized text analysis to track social processes. In T. Holtgraves (Ed.), *The Oxford handbook of language and social psychology*. New York: Oxford University Press.

Du Bois, J. W. (2007). The stance triangle. In R. Englebretson (Ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction* (pp. 139–182). Amsterdam: Benjamins.

Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (2010). Language style matching as a predictor of social dynamics in small groups. *Communication Research*, 31, 3–19. doi:10.1177/0093650209351468

Ireland, M. E., & Henderson, M. D. (2014). Language style matching, engagement, and impasse in negotiations. *Negotiation & Conflict Management Research*, 7, 1–16. doi:10.1111/ncmr.12025

Ireland, M. E., & Pennebaker, J. W. (2010). Language style matching in writing: Synchrony in essays, correspondence, and poetry. *Journal of Personality and Social Psychology*, 99, 549–571. doi:10.1037/a0020386

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, 22, 39–44. doi:10.1177/0956797610392928.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25, 259–284. doi:10.1080/01638539809545028

Lord, S. P., Sheng, E., Imel, Z. E., Baer, J., & Atkins, D. C. (2014). More than reflections: Empathy in motivational interviewing includes language style synchrony between therapist and client. *Behavior Therapy*, 46, 296–303. doi:10.1016/j.beth.2014.11.002.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Sharing one's story: On the benefits of writing or talking about emotional experience. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 573–583). Oxford: University Press.

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic Inquiry and Word Count (LIWC2007): A computer-based text analysis program [Computer software]*. Austin, TX: LIWC.net.

Pennebaker, J. W., Chung, C. K., Ireland, M. E., Gonzales, A. L., & Booth, R. J. (2007). *The development and psychometric properties of LIWC2007*. Austin, TX: LIWC.net.

Scissors, L., Gill, A. J., Geraghty, K., & Gergle, D. (2009). In CMC we trust: The role of similarity. In *Proceedings of CHI 2009* (pp. 527–536). New York: ACM Press.

Taylor, P. J., & Thomas, S. (2008). Linguistic style matching and negotiation outcome. *Negotiation and Conflict Management Research*, 1, 263–281. doi:10.1111/j.1750-4716.2008.00016.x

Profile 30

Leader-Member Exchange 7 Questionnaire (LMX-7)

(Graen & Uhl-Blen, 1995)

Profiled by: Lisa K. Hanasono, PhD

Bowling Green State University

Construct

The Leader-Member Exchange 7 questionnaire (LMX-7) was developed to measure the quality of professional relationships between leaders and followers (Graen, Novak, & Sommerkamp, 1982; Graen & Uhl-Bien, 1995).

Instrument Type

Self-Report

Description

The LMX-7 (Graen & Uhl-Bien, 1995) measures the quality of professional relationships between leaders and followers. Specifically, this instrument quantifies the amount of respect, trust, and obligation exchanged within leader–follower relationships, traits that are often associated with listening (Churchill, Ford, Hartley, & Walker, 1985; Drollinger, Comer, & Warrington, 2006).

The LMX-7 operates on the assumption of differentiated dyadic relationships, meaning that interpersonal interactions may vary from one leader–follower dyad to another, and these relationships may change over time (Dulebohn, Bommer, Liden, Brouer, & Ferris, 2012). As noted, the instrument examines three dimensions: mutual respect, reciprocal trust, and expectations about relational obligations. Focusing on interval-level data, participants provide their responses to the questionnaire's seven items, each of which uses a different 5-point scale (see further in this profile). The questionnaire

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

can be adapted easily to be completed from the perspective of a leader or follower (e.g., "How would you characterize your working relationship with your leader?" vs. "How would you characterize your working relationship with your follower?").

Administration

The LMX-7 is self-administered and takes between 5 and 15 minutes to complete, depending on the number of individuals a respondent is assessing. Participants should respond to the questionnaire from the perspective of a leader or a follower. Focusing on their role as a leader, participants should complete the questionnaire multiple times to assess the working relationships they have with each of their subordinates. When examining one's perspective as a follower, participants should complete the questionnaire to assess their professional relationships with their direct superiors (e.g., bosses, coaches, and/or advisors).

Scoring

For each leader–follower relationship, all seven items from the questionnaire should be summed. The total scores can be interpreted as very high (30–35), high (25–29), moderate (20–24), low (15–19), and very low (7–14). The LMX-7 provides an index of relational quality between leaders or followers, so higher scores indicate stronger relationships (i.e., ingroups), whereas lower scores indicate weaker relationships (i.e., outgroups).

To assess leader–member agreements on LMX ratings, researchers can examine the correlation between the average scores from leaders' perspectives and followers' perspectives. The correlation is usually positive in direction and moderate in size, $r \approx .22$ (Gerstner & Day, 1997).

Development

Efforts to develop a measurement of LMX, including the LMX-7, reflect the ongoing evolution of LMX theory and its construct (Schriesheim, Castro, & Cogliser, 1999). Graen and Uhl-Bien (1995) identified four stages that defined and refined the conceptualization and operationalization of LMX. First, researchers shifted their attention from studying the characteristics of individual leaders and followers to examining differentiated leader–follower relationships within professional contexts. Scholars asserted that leaders do not treat all of their followers in the same manner; instead, leaders form unique relationships with their subordinates (Dansereau, Graen, & Haga, 1975; Graen, Novak, & Sommerkamp, 1982). Second, LMX researchers explored characteristics of leader–follower relationships and their impact on organizational variables, such as job satisfaction (e.g., Stepina, Perrewe, Hassell, Harris, & Mayfield, 1991) and turnover (e.g., Graen, Liden, & Hoel, 1982). Third, LMX scholarship examined the relational development between leaders and followers over time, spanning from strangers (with low LMX levels) to mature partnerships (with high LMX levels). Fourth, scholars expanded their treatment of LMX beyond independent dyads to explore how interdependent dyads coexist within a larger system of professional networks.

With each new stage, the theoretical construction of LMX and its measurement have shifted (Schriesheim *et al.*, 1999). LMX instruments have ranged from 2-item measures (e.g., Dansereau *et al.*, 1975), 4-item measures (e.g., Graen & Schiemann, 1978), 5-item measures (e.g., Graen, Liden, *et al.*, 1982), and 7-item measures (e.g., Graen, Novak, *et al.*, 1982) to multidimensional scales like the LMX-MDM (Liden & Maslyn, 1998) and LMX-17 (Graen & Scandura, 1987). Despite the existence of numerous LMX scales, the LMX-7 remains one of the most widely used instruments. In their meta-analysis of LMX theory, Gerstner and Day (1997) concluded that the LMX-7 "appears to provide the soundest psychometric properties of all LMX measures" (p. 837).

Reliability

The internal consistency of scores generated from the LMX-7 has been acceptable, with Cronbach's alpha generally ranging from .80 to .90 (Gerstner & Day, 1997). Due to high correlations among items measuring respect, trust, and obligation (Graen & Uhl-Bien, 1995), the LMX-7 Questionnaire is usually interpreted as quantitatively unidimensional.

Validity

Despite the availability of longer and more elaborate LMX scales, many scholars endorse the LMX-7 due to its conciseness and evidence of convergent validity. In their review of LMX measures, Graen and Uhl-Bien (1995) explained that "even though items [from alternative LMX measures] were added to tap into possible multiple dimensions, the expanded measure was highly correlated with the more concise 7-item LMX and produced the same effects" (p. 236). Agreeing with this conclusion in their meta-analysis, Dulebohn and colleagues (2012) completed paired comparisons between the LMX-7 and other LMX scales like the LMX-MDM and LMX-17 and found no statistically significant differences among the measures.

Availability

Originally published by Graen and Uhl-Blen (1995), the LMX-7 is provided at the end of this profile, with permission. It is free to use for research purposes. Instructions should be changed to reflect the relationship(s) upon which participants should focus.

Sample Studies

For more than 40 years, scholars have studied LMX to understand how effective leadership operates in professional relationships (Dansereau *et al.*, 1975; Graen, 1976; Graen & Cashman, 1975). More recently, researchers have used the LMX-7 to examine superior–subordinate relationships in a broad scope of professional contexts and cultures. Studies continue to examine the main effects of LMX on individual, interpersonal, and organizational factors. For example, Myers (2006) found that college students at a Midwestern university with higher levels of LMX were more motivated to communicate with their instructors for relational, functional, and participatory reasons than were students with lower levels of LMX. Lloyd, Boer, and Voelpel's (2015) study on German supervisor–employee interactions reported that LMX was positively related to—yet distinct from—active empathic listening and perceived listening quality. While studying workers at civil engineering companies in South Korea, Lee and colleagues (Lee, Park, Lee, & Lee, 2007) found a positive relationship between LMX and employees' preference for using direct communicative strategies to seek feedback from their superiors, and Volmer and coauthors' (Volmer, Spurk, & Niessen, 2011) longitudinal study found that LMX was predictive of job satisfaction after a 3-month period.

Other studies examined how factors mediate the relationship between LMX and its outcome variables. For example, Katrinli and colleagues' (Katrinli, Atabay, Gunay, & Guneri, 2008) study on nurses at a private hospital in Turkey indicated that the positive relationship between LMX and organizational identity was mediated by job involvement. While studying a sample of nurses (59% Hispanic, 35% Caucasian) at a southwestern US medical facility, Walumbwa, Cropanzano, and Goldman (2011) reported that the positive relationship between LMX and job performance was mediated by participants' commitment to their supervisor.

Finally, some scholars studied how LMX serves as a mediating variable. For example, while studying workers at a Fortune 500 telecommunications company in China, Xu and colleagues (Xu, Huang, Lam, & Miao, 2012) reported that LMX fully mediated the negative effects of abusive supervising behaviors (e.g., verbal insults) on employees' task performance.

Critique

Guided by the principles of LMX theory (Graen, 1976; Graen & Scandura, 1987), the LMX-7 provides an important contribution to the study of leadership because it redefines the level of analysis from an individual person (i.e., a leader-based or follower-based) to a dyadic unit (i.e., based on the superior–subordinate relationship). However, scholars have identified three major criticisms of the LMX-7. First, several versions of the LMX questionnaire have been utilized by scholars (Dulebohn *et al.*, 2012). To encourage the consistent use of the same 7-item scale, Graen and Uhl-Bien (1995) made their recommended LMX-7 instrument more easily accessible by publishing it in *Leadership Quarterly* in 1995.

Second, the questionnaire includes several double-barreled items, which might negatively affect the instrument's clarity and exactitude. For example, one item asks, "How well does your leader understand your job problems and needs?" When analyzing data, researchers cannot ascertain the degree to which participants' numerical responses uniquely measure each issue (i.e., job problems vs. needs).

Third, scholars continue to debate whether LMX should be a unidimensional or multidimensional construct, which has direct implications on the scale's construct validity and measurement. As LMX theory evolved, researchers offered varying conceptualizations of LMX and its dimensions (for a review, see Schriesheim *et al.*, 1999). For example, Dienesch and Liden (1986) argued that LMX could be understood in terms of three dimensions: affect (i.e., liking), loyalty, and task-related contributions. Liden and Maslyn (1998) added a fourth dimension: professional respect. While reviewing more than 100 studies on LMX research, Schriesheim *et al.* (1999) identified at least 18 subdimensions of LMX that have been examined in the literature, including innovativeness, authority, understanding, and mutual control. Offering a more parsimonious conceptualization of LMX, Graen and Uhl-Bien (1995) argued for three underlying and interrelated dimensions: mutual respect, trust, and obligation. Although these three dimensions offer theoretical value, Graen and Uhl-Bien (1995) acknowledged that "these dimensions are so highly correlated that they can be tapped into with the single measure of LMX" (p. 237). As LMX theory and scholarly conceptualizations of LMX continue to evolve, future research should remain attentive to concurrent and construct validities of LMX measures.

References

Churchill, G. A., Jr., Ford, N. M., Hartley, S. W., & Walker, O. C., Jr. (1985). The determinants of salesperson performance: A meta-analysis. *Journal of Marketing Research*, *22*, 103–118. http://www.jstor.org/stable/3151357

Dansereau, F., Jr., Graen, G. B., & Haga, W. J. (1975). A vertical dyad linkage approach to leadership within formal organizations: A longitudinal investigation of the role making process. *Organizational Behavior and Human Performance*, *13*, 46–78. doi:10.1016/0030-5073(75)90005-7

Dienesch, R. M., & Liden, R. C. (1986). Leader-member exchange model of leadership: A critique and further development. *Academy of Management Review*, *11*, 618–634. doi:10.5465/AMR.1986.4306242

Drollinger, T., Comer, L. B., & Warrington, P. T. (2006). Development and validation of the active empathetic listening scale. *Psychology and Marketing*, *23*, 161–180. doi:10.1002/mar

Dulebohn, J. H., Bommer, W. H., Liden, R. C., Brouer, R. L., & Ferris, G. R. (2012). A meta-analysis of antecedents and consequences of leader-member exchange: Integrating the past with an eye toward the future. *Journal of Management*, *39*, 1715–1759. doi:10.1177/0149206311415280

Gerstner, C. R., & Day, D. V. (1997). Meta-analytic review of leader-member exchange theory: Correlates and construct issues. *Journal of Applied Psychology*, *82*, 827–844. doi:10.1037//0021-9010.82.6.827

Graen, G. B. (1976). Role-making processes within complex organization. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 1201–1245). Chicago, IL: Rand McNally.

Graen, G. B., & Cashman, J. (1975). A role-making model of leadership in formal organizations: A development approach. In J. G. Hunt & L. L. Larson (Eds.), *Leadership frontiers* (pp. 143–166). Kent, OH: Kent State University Press.

Graen, G. B., Liden, R. C., & Hoel, W. (1982). Role of leadership in the employee withdrawal process. *Journal of Applied Psychology*, *67*, 868–872. doi:10.1037/0021-9010.67.6.868

Graen, G. B., Novak, M., & Sommerkamp, P. (1982). The effects of leader-member exchange and job design on productivity and satisfaction: Testing a dual attachment model. *Organizational Behavior and Human Performance*, *30*, 109–131. doi:10.1016/0030-5073(82)90236-7

Graen, G. B., & Scandura, T. A. (1987). Toward a psychology of dyadic organization. *Research in Organizational Behavior*, *9*, 175–208.

Graen, G. B., & Schiemann, W. (1978). Leader-member agreement: A vertical dyad linkage approach. *Journal of Applied Psychology*, *63*, 206–212. doi:10.1037/0021-9010.63.2.206

Graen, G. B., & Uhl-Bien, M. (1995). Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective. *Leadership Quarterly*, *6*, 219–247. doi:10.1037/0021-9010.63.2.206

Katrinli, A., Atabay, G., Gunay, G., & Guneri, B. (2008). Leader-member exchange, organizational identification and the mediating role of job involvement for nurses. *Journal of Advanced Nursing*, *64*, 354–362. doi:10.1111/j.1365-2648.2008.04809.x

Lee, H. E., Park, H. S., Lee, T. S., & Lee, D. W. (2007). Relationships between LMX and subordinates' feedback-seeking behaviors. *Social Behavior and Personality*, *35*, 659–674. doi:10.2224/sbp.2007.35.5.659

Liden, R. C., & Maslyn, J. (1998). Multidimensionality of leader-member exchange: An empirical assessment through scale development. *Journal of Management*, *24*, 43–72. doi:10.1016/S0149-2063(99)80053-1

Lloyd, K. J., Boer, D., & Voelpel, S. C. (2015). From listening to leading: Toward an understanding of supervisor listening within the framework of leader-member exchange theory. *International Journal of Business Communication,* 1–21. doi:10.1177/2329488415572778

Myers, S. A. (2006). Using leader-member exchange theory to explain students' motives to communicate. *Communication Quarterly*, *54*, 293–304. doi:10/1080/01463370600878008

Schriesheim, C. A., Castro, S. L., & Cogliser, C. C. (1999). Leader-member exchange (LMX) research: A comprehensive review of theory, measurement, and data-analytic practices. *Leadership Quarterly*, *10*, 63–113. doi:10.1016/S1048-9843(99)80009-5

Stepina, L. P., Perrewe, P. L., Hassell, B. L., Harris, J. R., & Mayfield, C. R. (1991). A comparative test of the independent effects of interpersonal, task, and reward domains on personal and organizational outcomes. *Journal of Social Behavior and Personality*, *6*, 93–104.

Volmer, J., Spurk, D., & Niessen, C. (2011). Leader-member exchange (LMX), job autonomy, and creative work-involvement. *The Leadership Quarterly*, *12*, 456, 465. doi:10.1016.j.leaqua.2011.10.005

Walumbwa, F. O., Cropanzano, R., & Goldman, B. M. (2011). How leader-member exchange influences effective work behaviors: Social exchange and internal-external efficacy perspectives. *Personnel Psychology*, *64*, 739–770. doi:10.1111/j.1744–6570.2011.01224.x

Xu, E., Huang, X., Lam, C. K., & Miao, Q. (2012). Abusive supervision and work behaviors: The mediating role of LMX. *Journal of Organizational Behavior*, *33*, 531–543. doi:10.1002/job.768

Scale

The LMX-7 (Graen & Uhl-Blen, 1995)

Source: Graen and Uhl-Blen (1995). Reproduced with permission of Elsevier.

Instructions: This questionnaire contains items that ask you to describe your relationship with either your leader or one of your subordinates. For each of the items, indicate the degree to which you think the item is true for you by circling one of the responses that appear below the item.

1) Do you know where you stand with your leader (follower) … [and] do you usually know how satisfied your leader (follower) is with what you do?

Rarely	*Occasionally*	*Sometimes*	*Fairly often*	*Very often*
1	2	3	4	5

2) How well does your leader (follower) understand your job problems and needs?

Not a bit	*A little*	*A fair amount*	*Quite a bit*	*A great deal*
1	2	3	4	5

3) How well does your leader (follower) recognize your potential?

Not at all	*A little*	*Moderately*	*Mostly*	*Fully*
1	2	3	4	5

4) Regardless of how much formal authority your leader (follower) has built into his or her position, what are the chances that your leader (follower) would use his or her power to help you solve problems in your work?

None	*Small*	*Moderate*	*High*	*Very high*
1	2	3	4	5

5) Again, regardless of the amount of formal authority your leader (follower) has, what are the chances that he or she would "bail you out" at his or her expense?

None	*Small*	*Moderate*	*High*	*Very high*
1	2	3	4	5

6) I have enough confidence in my leader (follower) that I would defend and justify his or her decision if he or she were not present to do so.

Strongly disagree	*Disagree*	*Neutral*	*Agree*	*Strongly agree*
1	2	3	4	5

7) How would you characterize your working relationship with your leader (follower)?

Extremely ineffective	*Worse than average*	*Average*	*Better than average*	*Extremely effective*
1	2	3	4	5

Profile 31

Listenability Style Guide (LSG)

(Rubin, 2012)

Profiled by: Graham D. Bodie, PhD

Meek School of Journalism and New Media, The University of Mississippi

Construct

Listenability, or listenable discourse, is the degree to which a message "is characterized by linguistic and rhetorical structures that ease the particular cognitive burdens listeners face" (Rubin, 2012, p. 178). The listenability of a spoken passage is a function of several features that make messages easier to comprehend. Thus, listenability resides in discourse, even though it is thought to primarily influence how much information listeners understand and retain.

Instrument Type

Behavioral Assessment

Description

The Listenability Style Guide (LSG) is a "menu of stylistic resources" that can be used to assess the degree to which a given message "[qualifies] as listenable" (Rubin, 2012, pp. 178–179). Listenability is a function of several factors that make spoken discourse easier to follow, understand, and comprehend. The four primary factors that make up listenability are summarized in Table P31.1.

Listenability is relevant for any text that is intended to be heard, regardless of whether that text was initially written and later read aloud or initially spoken and later transcribed. The LSG attempts to provide an analogous assessment of oral-based discourse to the readability metrics that rate literate-based discourse. Contrary to readability indices,

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Table P31.1 Factors that make discourse more listenable.

Factor	Description
Oral-based lexis and syntax	Verbal rather than nominalized constructions; relatively low lexical diversity
Readily accessible rhetorical structures	Discourse structure in which high-level rhetorical predicates are highlighted with organizational cues or multiple causal links rather than multiply chained support for warrants
Evocation of oral-based interaction contexts	Use of first-person reference, second-person direct address, exophoric pronouns
Considerateness	Coherence, consistency, predictable information flow, conformity to conversational maxims, and useful redundancies

Source: Rubin, Hafer, & Arata (2000, p. 123). Reproduced with permission of Taylor & Francis.

however, the LSG "deliberately eschews the reductionistic formulae that characterize readability research and practice" (Rubin, 2012, p. 176). The LSG provides guidelines for assessing listenability and is thus described as a more qualitative assessment rather than a numeric scoring rubric.

Administration

The LSG "is intended to assist communicators in crafting messages for audiences to hear rather than to read" (Rubin, 2012, p. 178). It should not be used "as a checklist of obligatory features … [as messages do not] require all of the elements listed in the guide in order to qualify as listenable" (pp. 178–179). The LSG is not a "rubric for quantifying listenability," suggesting that researchers desiring to use the LSG should be well versed in the differences between oral- and literate-based discourse (discussed further in this profile) as well as "textual norms for messages intended for the two different modalities (Rubin, Hafer, & Arata, 2000, p. 123).

Scoring

Because the LSG is more a set of guidelines than a rubric, it does not generate a number like measures of readability (e.g., reading grade-level). Past work has classified existing messages as moderately listenable, then generated highly listenable versions of those messages in subsequent experiments that explored comprehension, retention, and processing of these messages (Rubin, 2012; Rubin *et al.*, 2000). Researchers looking for a numeric valuation of a text's listenability will be disappointed with the LSG.

Development

For decades, educators and researchers alike have struggled with how best to measure comprehension of information presented orally. Most tests of listening comprehension were developed to mirror best practices for measuring reading comprehension—present a text (written or oral), and ask a series of multiple-choice questions that ask about details or general themes presented in the text. Indeed, some tests of listening comprehension merely present recorded versions of reading comprehension passages; the implicit assumption of this strategy is that comprehension works the same regardless of modality (Harwood, 1955).

The development of listenability and appropriate metrics for its assessment was motivated by the recognition that written and spoken language "differ as a matter of degree" rather than any "sharp distinction" between modalities (Rubin & Rafoth, 1986, pp. 140–141). Regardless of whether it was designed to be read or heard, "a passage may exhibit more or less orality depending upon the density of those features which are characteristic of typical spoken language" (Rubin & Rafoth, 1986, p. 140). Orality is a feature of any discourse, and "to say then that some sample of discourse is 'oral-based' reveals nothing of the modality in which it was actually composed" (Rubin *et al.*, 2000, p. 124). Instead of distinguishing between spoken and written discourse, texts can be placed on a continuum from very literate-based to very oral-based (Biber, 1988). Foreshadowing the LSG, Rubin and Rafoth presented 10 characteristics "useful … for educators wishing to select materials particularly suited for listening" (p. 148):

1) Sentence structure is relatively simple rather than complex. Highly embedded or subordinated sentences, final free modifiers (such as nominative absolutes), long prepositional phrases, and sequences of participial phrases are rare or nonexistent.
2) Sentences are joined by coordinating conjunctions.
3) Sentences string propositions out with a verbal style instead of compacting propositions with nominalizations, appositives, and participial phrases.
4) Passages contain a relatively high degree of redundancy.
5) Passages resolve thematic units quickly, although thematic inconsistency between distal portions may be present.
6) Passages have clear structural characteristics that minimize the amount of inferencing required to understand underlying structural elements and their relations to one another.
7) Passages presuppose face-to-face interaction by occasional use of features like dexis (i.e., situationally dependent terms, such as *this-that* and *here-there*), definite articles where no unique reference has been established, first-person point of view, and second-person address.
8) Passages may include ellipses and "minor sentences."
9) Passages use the vocabulary conventionally associated with speech events.
10) Passages include formulaic expressions that evoke oral situations.

Reliability

The LSG is used to judge texts for their listenability, and so any given text should be judged in a similar manner regardless of the individual doing the judging. Although standard measures of interrater reliability are applicable to the LSG, no such measures have been reported in studies using the guidelines. Future work should establish how consistent trained judges are when assessing texts of various types.

Validity

The LSG seems to have a high degree of face validity as the domains used to assess texts for listenability have been found to distinguish "between oral-based and writing-based discourse" (Rubin, 2012, p. 178). There is an implicit assumption in the LSG such that oral-based discourse should be better understood than literate-based discourse when delivered verbally; literate-based discourse should be better understood than oral-based discourse when delivered in written form. Rubin *et al.* (2000) tested this assumption in an experimental study that asked participants to read or listen to a text designed to be spoken or a text designed to be read. The primary hypothesis was an interaction effect between the modality factor (spoken/written) and the text factor (oral/literate). Results showed only main effects. Perhaps part of the reason for the lack of an interaction was that the texts used did not appreciably differ when analyzed for stylistic dimensions of orality-literateness. Better manipulations are needed in future work.

Rubin (2012) attempted a similar study using postsurgical discharge instructions that he modified to create moderately and highly listenable versions; participants were asked to read or listen to one of these versions. These versions were not tested for orality-literateness nor were any manipulation checks offered to provide evidence the versions differed in listenability. More importantly, no interaction terms were statistically significant, mirroring the Rubin *et al.* (2000) findings. Perhaps even more important, the highly listenable version was also scored as more readable than the moderately listenable version, suggesting some conceptual overlap between listenability and readability.

A third study has provided some evidence, however, that readability metrics may not be the best way to manipulate listenability, at least with respect to the degree of comprehension of spoken text. Eastwood, Snook, and Chaulk (2010) found that increasing the readability of police cautions (the Canadian equivalent of Miranda rights) did not improve comprehension when these texts were presented orally, causing the authors to conclude "that the reading complexity measures examined here may not be useful predictors of listening comprehension of police cautions" (pp. 465–466). A similar conclusion can be gleaned from early work by Harwood (1955). The extent to which readability and listenability are distinguishable is thus still an open question.

Availability

The LSG was first published in a 2012 *Journal of Health Communication* article; this article is open-source and free to download from http://www.tandfonline.com/doi/pdf/10.1080/10810730.2012.712622. It is reproduced at the end of this profile and is free to use for research purposes.

Sample Studies

Measuring listenability as a distinct construct is much less common than using indices of readability to approximate likely comprehension. To date, the LSG has been applied in one study to postsurgical instructions (Rubin, 2012) and in two studies to police cautions (Eastwood & Snook, 2012; Snook, Luther, Eastwood, Collins, & Evans, 2016).

Eastwood and Snook created eight police caution statements: 2 (instructions vs. no instructions) × 2 (listing vs. no listing) × 2 (explanation vs. no explanation). The only factor to exhibit a main effect on comprehension was explanation such that cautions with built-in redundancy increased comprehension by as much as 30%. The results suggested that providing instruction prior to presenting the caution (the instruction's manipulations) and signposting transitions (the listing manipulation) did not improve comprehension. No interaction terms were statistically significant either.

Snook *et al.* (2016) conducted a conceptual replication using a mock interrogation design. In the first study, participants were randomly assigned to commit a fake crime or to remain innocent; both groups were then interrogated, during which their legal rights were either scripted to be highly listenable or not. In the second study, legal rights were manipulated to include variations of listenability along the Instruction, Listing, and Explanation factors in a manner similar to that of Eastwood and Snook. Results from both studies show that increased comprehension is a function of listenable discourse with a 25% improvement over the base statement. Thus, listenable material can improve comprehension even in more cognitively taxing situations. The extent to which these results can be replicated in actual police interrogation situations remains to be demonstrated.

Critique

To date, the LSG has not been used extensively, but research employing its principles provides promising evidence that spoken discourse can be made more comprehensible by adding a few simple features known to improve attention and involvement. As shown by work on police cautions, the elements of the LSG can be manipulated individually and tested for their relative importance toward the prediction of retaining information. Also evident from these data, however, is the fact that some types of information are more easily retained than other types of information, perhaps regardless of how listenable it is. In the Snook *et al.* (2016) study, for instance, 70% of participants recalled they had a right to hire a lawyer without any listenability manipulation; the addition of the listenability factors increased comprehension to 100%. Thus, some information may be readily known by a high percentage of participants. An interesting study would include a police caution without the right to hire a lawyer to see if listeners retained that information (false comprehension). For the right to apply for legal aid, however, average comprehension ranged from 0 to 10% with no substantive improvement with the manipulation of listenability criteria. These results suggest that familiarity with rights influences comprehension.

Perhaps because of the contexts within which listenability has been applied, the sole focus has been on how to improve comprehension. But comprehension is not the only important outcome for orally based discourse. Listeners often listen for enjoyment or to simply retain the gist of a conversation. The listenability criteria emphasized in the LSG may or may not be applicable to making oral discourse more pleasant or aurally aesthetic, for instance. Research on the degree to which the factors that make up the LSG guidelines do more than improve comprehension is needed.

Because it is a set of guidelines rather than a prescriptive rubric, it is currently impossible to ascertain the degree to which different users are applying the various factors in similar ways. Presently, the only way to test whether a text has been made more listenable is to present that text along with a text that is thought to be less listenable and measure comprehension or retention of information. This begs the question as to whether listenability is simply a characteristic of messages that makes them more comprehensible.

Finally, the degree to which the criteria outlined in the LSG produce a set of guidelines that are distinct from those already established for readability is questionable. Perhaps the most damaging evidence comes from my own analysis of the phrases found in the LSG (see Scale below). The listenable examples had an average reading ease (Flesch–Kincaid) of 77.35 ($SD = 12.38$) and reading level of 6.17 ($SD = 2.16$), whereas the "not" examples had an average reading ease of 36.87 ($SD = 23.36$) and reading level of 11.58 ($SD = 3.46$); both sets of means were statistically different with a large effect size ($r^2 = .54$, reading ease; $r^2 = .47$, reading level). As a result, the LSG is advising users to generate listenable material by also making that material more readable. It might be difficult, although not impossible, to create highly listenable material that is not also highly readable. The police cautions used in Eastwood and Snook (2012), however, decreased reading ease and increased reading level compared to the baseline condition (see Scale below). The difference between the readability and listenability of texts still deserves additional research attention.

References

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press

Eastwood, J., & Snook, B. (2012). The effect of listenability factors on the comprehension of police cautions. *Law and Human Behavior*, *36*, 177–183. doi: 10.1037/h0093955

Eastwood, J., Snook, B., & Chaulk, S. J. (2010). Measuring reading complexity and listening comprehension of Canadian police cautions. *Criminal Justice and Behavior*, *37*, 453–471. doi:10.1177/0093854810362053

Harwood, K. A. (1955). Listenability and readability. *Speech Monographs*, *22*, 49–53. doi:10.1080/03637755509375133

Rubin, D. L. (2012). Listenability as a tool for advancing health literacy. *Journal of Health Communication*, *17*, 176–190. doi:10.1080/10810730.2012.7126622

Rubin, D. L., Hafer, T., & Arata, K. (2000). Reading and listening to oral-based versus literate-based discourse. *Communication Education*, *49*, 121–133. doi:10.1080/03634520009379200

Rubin, D. L., & Rafoth, B. A. (1986). Oral language criteria for selecting listenable materials: An update for reading teachers and specialists. *Reading Psychology: An International Quarterly*, *7*, 137–151. doi:10.1080/0270271860070302

Snook, B., Luther, K., Eastwood, J., Collins, R., & Evans, S. (2016). Advancing legal literacy: The effect of listenability on the comprehension of interrogation rights. *Legal and Criminological Psychology*, *21*, 174–188. doi:10.1111/lcrp.12053

Scale

The Listenability Style Guide (Rubin, 2012)

Source: Rubin (2012). Reproduced with permission of Taylor & Francis.

Listenability domain	Recommended language or discourse practice	Example
Oral-based sentence structure	When you link clauses, use coordinating conjunctions when possible. Don't overuse subordination	To help this medication to work better, you should eat a low-fat diet, and you should exercise, and you should not smoke. NOT: In addition to eating a proper diet (such as a low-cholesterol/ low-fat diet), it is known that this medication will work better if lifestyle changes include exercising, losing weight if overweight, and stopping smoking.
	When you do use subordinate clauses, put them at the ends of sentences.	People with hypertension should avoid fast foods, because fast foods contain a lot of salt. NOT: Fast foods, because they contain excessive salt, should be avoided by hypertension patients.
	Keep clauses to moderate length. Avoid many long phrases, which are really "reduced clauses."	Don't eat grapefruit and don't drink grapefruit juice while you are taking this medicine. But listen to your doctor's instructions about this grapefruit rule. NOT: Absent your doctor's instructions to the contrary, avoid consumption of grapefruit products while under this medication.
	Make your sentences express actions by using verb forms. Avoid dense noun forms that derive from verbs.	Consume lots of liquids so you don't dehydrate. NOT: Consumption of plentiful liquids prevents dehydration.
Oral-based vocabulary	Use personal pronouns to evoke a face-to-face context. Imperative sentence (understood *you*) also work well.	The day after surgery your knee and lower leg might swell. So keep your leg raised. NOT: Swelling in the knee and lower leg is common 24 hours after surgery. Raising the leg will help reduce that symptom.

(*Continued*)

Listenability domain	Recommended language or discourse practice	Example
	Repeat content nouns and verbs across sentences. Don't aim for diverse vocabulary for its own sake (low lexical diversity).	To see if your tumor contains malignant cells, your doctor will do a needle biopsy. The needle biopsy will draw out a small section of the tumor that your doctor can examine under a microscope. NOT: Your doctor will perform a needle biopsy on your tumor. This procedure allows microscopic examination of sampled cells to search for malignancies.
	Use everyday words (but don't be afraid of using specialized language if it is common to the patient/ consumer community).	Your incision from the surgery may turn red for a day or two. It's okay so long as the redness does not spread. NOT: The surgical incision may become inflamed for a day or two. You need to become concerned only if the inflammation spreads to surrounding tissue.
Features of face-to-face conversation	Use contractions freely.	You've got to take it before breakfast. This medicine won't work if you don't take it on an empty stomach. NOT: This medication will not function properly if administered with milk or milk products.
	Use questions to focus attention.	Now would you like to know some of the easiest ways to keep your blood sugar steady all day long? NOT: Several strategies are available to reduce fluctuations in blood sugar levels.
	Use conjunctions at the beginning of sentences to create conversational flow.	And so you've probably heard a lot about the importance of getting lots of calcium for women's bones to stay strong.
	Use tag questions.	There is so much conflicting advice about the value of back surgery, isn't there?
	Use simple and common idioms.	When you spend too much time being a couch potato, it'll get you in the end. NOT: Sedentary lifestyles carry health risks.

Listenability domain	Recommended language or discourse practice	Example
	Periodically call your interlocutor by name.	I'll tell you, Ms. Lawson, you really know a lot about how insulin works.
Considerateness toward listeners	Use internal summaries.	So now we've talked about two different ways to keep your child's teeth from going bad. One way is for you to give your child more water to drink and less soda. The second way is to get your child's teeth sealed at the dentist's office. Now here's a third thing you can do to keep your child's teeth healthy.
	Give listener's "advance organizers" to help them predict what information will follow.	In this little talk I'm going to cover three simple things you can do to reduce the chance of getting a stroke. You can cut back the amount of red meat that you eat. You can walk 1000 steps every day. And you can take one baby aspirin at night. First, let's talk about the meat in your diet.
	Explicitly signal transitions between topics.	So that's what I wanted to tell you about causes of reflux. If you don't have any questions, let's talk now about what you can do about it.
	Announce important topics.	Vitamin A. Vitamin A in vegetables like carrots can keep your eyes healthy.
	Convey information in little stories.	I know this one lady, my neighbor Rosie, who didn't even want to walk up her driveway to get her mail her knees hurt her so much. At first she was just using ice on her knees, and it helped at first, but then icing stopped working for her. So she started using heat. Finally, she figured out that the best thing for her was to switch between ice packs and heating pads. NOT: Alternative heat and cold can sometimes relieve chronic knee pain.

(*Continued*)

Listenability domain	Recommended language or discourse practice	Example
	Be redundant. State and then paraphrase important points.	Suppose you lose consciousness or you can't stay awake during times when you are usually awake. If that happens, you should go back to the emergency room. Abnormal sleepiness or blacking out are absolutely reasons to return to the hospital right away. NOT: Loss of consciousness and inability to remain awake are signs that it would be wise to return to the emergency room.
	Convey information with vivid analogies to everyday objects and events.	So you see, using an expired prescription is a little like eating spoiled food. It doesn't give you the benefit you need, and it can actually hurt you. NOT: Expired prescriptions often fail to deliver the necessary therapeutic effect and may cause additional harm.

Police Caution Messages used in Eastwood and Snook (2012)

<u>Base Legal Counsel Caution (Reading ease = 88.7, Reading level = 4.7)</u>

You have the right to hire and talk to your own lawyer right away. You have the right to free legal advice from a government lawyer right away. If you want this free advice I will give you the number to call. If you are charged with a crime you can apply for a free lawyer to help with your case.

<u>Instructions (read prior to caution, or not) (Reading ease = 85.4, Reading level = 5)</u>

I am going to read you the police caution. The police caution describes the rights that you have when being interviewed by the police. I want you to listen carefully to the caution as I am reading it and think about the information that you hear. This is important, as I will ask you to tell me what the caution means when I finish reading it. I will start reading the caution now.

<u>Listing (Reading ease = 82.9, Reading level = 6.5)</u>

You have four rights that you need to know about: First, you have the right to hire and talk to your own lawyer right away. Second, you have the right to free legal advice from a government lawyer right away. Third, if you want this free legal advice, I will give you a telephone number to call. Fourth, if you are charged with a crime, you can apply for a free lawyer to help with your case.

Explanations (Reading ease = 83.5, Reading level = 6.4)

You have the right to hire and talk to your own lawyer right away. This means that you can hire and talk to any lawyer you want before I ask you any more questions. You have the right to free legal advice from a government lawyer right away. This means that you can talk to a free lawyer and get free legal advice before I ask you any more questions. If you want this free legal advice, I will give you a telephone number to call. This means that you can get a phone number from me that will let you call for the free legal advice I just mentioned. If you are charged with a crime, you can apply for a free lawyer to help with your case. This means that if you do end up being charged with a crime, you can apply to get a lawyer to help you for free.

Profile 32

Listening Concepts Inventory (LCI and LCI-R)

(Imhof & Janusik, 2006; Bodie, 2010)

Profiled by: Debra L. Worthington, PhD

Auburn University

Construct

The Listening Concepts Inventory (LCI) is a multidimensional measure designed to assess an individual's subjective conceptualization of listening.

Instrument Type

Self-Report

Description

The Listening Concepts Inventory (LCI; Imhof & Janusik, 2006) and its revised version (LCI-R; Bodie, 2010) were conceived as means of assessing individual and situational differences in how people conceptualize listening. Participants identify the degree to which they believe a list of activities reflect listening. The LCI consists of 33 items. The LCI-R reduced the number of items to 15. The listed activities fall into four broad categories: listening as a means of *organizing information*, listening as a means of *relationship building*, listening as a means of *learning*, and listening as a means of *critical evaluation*. Responses reflect an individual's view of the role and functions of listening (i.e., their listening belief system).

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

Both the original and revised versions of the LCI may be presented via paper or computer. Individuals are asked to consider how similar each listed activity is to listening (e.g., retaining information, helping, analyzing, and arguing) using 5-point scaling options ranging from 1 (*not at all similar*) to 5 (*identical*).

Scoring

Item responses are averaged within each subscale, and participants receive a score for each of the previously identified areas (i.e., organizing information, relationship building, learning, and critical evaluation). Higher scores are an indication that an individual's conceptualization of listening aligns more closely with a particular area. Normative data have not yet been established.

Development

The LCI was first introduced by Imhof and Janusik (2006) with the goal of creating a diagnostic means of measuring perceptions of listening cognitions. Drawing on a cognitive model of listening, Imhof and Janusik based their measure on the assumption that listening concepts "determine listening behavior, the process, and the outcome" (p. 79). They offered an integrated systems model of listening, grounded in the work of Biggs (1999), that established three interdependent listening facets: listening presage (i.e., personal and contextual factors), the listening process (cognitive and behavioral), and the listening product.

The conceptualizations that people hold about listening are important to this process model. These concepts form a listening framework that affects how individuals engage with others.

Using an inductive approach, Imhof and Janusik (2006) developed and presented 204 US and 154 German participants with 65 potential conceptualizations of listening. Data were collapsed across the two groups, and then the list of conceptualizations was tested for sampling adequacy (Hair, Tatham, Anderson, & Black, 1998) (KMO-MSA = .93; Bartlett chi-square = 1275.72, $df = 2016$, $p < .001$). Having met the requirements for data reduction, an initial principal component analysis (PCA) extracted 12 components, which were then reduced to 6 based on parallel analysis (O'Connor, 2000). Based on item loadings, the authors decided to interpret four components (two of the six components had fewer than four items loadings above .60) (Guadagnoli & Velicer, 1988). These four components accounted for 47% of the item variance. Next, they used the Fürntratt criterion to remove items from the factors. This process ultimately led to the identification of the previously described four areas—listening as organizing information, relationship building, critical assessment, and as learning/integrating information—and a total of 33 items.

Bodie (2010) designed a series of studies to provide validity evidence for the LCI. Study 1 was designed to provide further validation of the factor structure of the scale. Results of a confirmatory factor analysis with data gathered from US undergraduates did not replicate the hypothesized factor structure (33 items). Instead, data were more consistent with a 15-item scale reflecting the four means of conceptualizing listening suggested in the original LCI. This revised version was replicated and subjected to further tests of validity (see below).

Reliability

Reliability of the four factors identified by Imhof and Janusik (2006) appears to be stable. In their initial study, they reported Cronbach's alpha coefficients for the subscales ranging between .87 and .90. Davis, Thompson, Foley, Bond, and DeWitt (2008) reported a similar range (.88 to .98). Bodie (2010) reported reliability estimates ranging from .68 to .84 for the LCI-R, including stable estimates across time.

Validity

The majority of studies utilizing the LCI and LCI-R have focused on building its validity portfolio (see Development section). Bodie (2010) undertook the most rigorous assessment of the LCI, which, as previously described, resulted in the LCI-R.

One of his four studies was designed to explore the scale's measurement invariance, particularly its generalizability. Replicating results from his initial study, Bodie (2010) reported a good model fit of the LCI-R, χ^2 (84) = 207.40, $p < .001$, CFI = .93, RMR = .088, RMSEA = .08, CI90% = .07, .10. In addition, when comparing US undergraduate and healthcare workers, the model was found to be invariant with respect to measurement weights and structural covariances.

Bodie (2010) also addressed the temporal stability of the LCI and its association with other individual differences in listening and cognitive style. The nomological validity of the scale was also assessed. Bodie examined the relationship between LCI-R scores and individual listening styles, listening competency, and active listening as well as other cognitive measures, such as need for cognition and need to evaluate. Among the findings were positive associations between relationship building of the LCI-R relationally oriented listening (LSP-R) and the need to express opinions; positive associations between the LCI-R factors information acquisition and learning and LSP-R critical and extroversion; and a positive correlation between the LCI-R critical and neuroticism. Notably, the LCI-R variables accounted for 70% of the variance associated with the measured listening constructs and 4% of the variance in listening styles. Follow-up bivariate correlations revealed that the more listening constructs endorsed by an individual, the greater their self-reported listening competence (SPLCS, see profile 57): discriminative listening ($r = .17$), comprehensive listening ($r = .20$), critical listening ($r = .19$), and therapeutic listening ($r = .18$).

Availability

The original form of the LCI is available in the Imhof and Janusik (2006) article published in the *Journal of Intercultural Communication Research*. Bodie's (2010) revised version of the instrument (LCI-R) initially appeared in *Imagination, Cognition and Personality*. All 65 items utilized by Imhof and Janusik are presented at the end of this profile with notes regarding the 33 items that constitute the LCI and the 15 that comprise the LCI-R.

Sample Studies

Imhof and Janusik (2006) compared listening conceptualizations of German and US American undergraduate students. Although the effect size estimate was small ($\eta^2 = .04$), Germans linked listening more closely to relationship building than did their American counterparts. In contrast, US students appear to align listening more closely with learning and integrating information, and relatedly to critical evaluation ($\eta^2 = .22$ and .16, respectively).

Ala-Kortesmaa and Välikoski (2011) explored listening concepts in the legal context. Using an exploratory factor analysis, they compared 114 US American and 96 Finnish attorneys' conceptualizations of listening. Their analysis did not support the factor structure previously reported by Imhof and Janusik (2006). Their findings echo those of Imhof and Janusik, suggesting that individuals from different cultures may hold differing conceptualizations of listening. More specifically, Finnish attorneys tended to align listening more closely with critical listening and organizing information, whereas American attorneys appear to view listening primarily in terms of relationship building. The authors noted that this difference could be related to differences in the legal systems between the two countries. Methodologically, however, because factorial invariance was not established, any group differences could be the result of different factor structures rather than reflecting true differences.

Davis *et al.* (2008) explored conceptions of listening held by healthcare professionals (administrators, physicians, and nurses). This study is significant because it addressed differing conceptualizations of individuals holding differing roles in the same healthcare context. Statistically significant differences across the groups were reported for all conceptualizations, with the exception of critical listening. Again, however, because factorial invariance was not tested, we are left to question whether group differences are true.

Bodie (2010) also examined the impact of context on individual conceptualizations of listening. Study participants completed the LCI-R after being asked to imagine themselves in four specified listening situations designed to evoke one of the four listening conceptualization factors (e.g., a friend giving directions, a friend's relationship breakup, attending a classroom lecture, or participating in a small group in a class activity). After dropping three items that did not reflect their latent construct ($\lambda < .40$), the model fit was satisfactory (CFIs > .90, SRMR < .08). Results of the study provide evidence that listening conceptualizations vary with the listening situation.

Critique

Knowledge of how people conceptualize listening has important implications for understanding how people think about listening as well as how those perceptions affect their listening behaviors. Thus, more work should focus on building a scale that can invariably tap conceptualizations of listening. Although the LCI-R appears to have some advantages over the LCI (e.g., shorter length and replicated factor structure), it has yet to be tested across cultures or in populations other than US undergraduate students and healthcare workers.

The previously described sample populations are just a beginning to the type of study necessary to more fully study the measurement invariance across cultures. Collectivistic cultures may view listening processes differently than Western, more individualistic cultures, for instance, but an equivalent measure needs to be used to test possible differences.

References

Ala-Kortesmaa, S. & Välikoski, T. (2011). *Professional communication and listening concepts of Finnish and American legal professionals*. Paper presented at the annual meeting of the International Communication Association, Boston, MA. Retrieved from http://citation.allacademic.com/meta/p488673_index.html

Biggs, J. (1999). *Teaching for quality learning at university*. Buckingham, UK: Society for Research on Higher Education.

Bodie, G. D. (2010). The revised Listening Concepts Inventory (LCI-R): Assessing individual and situational differences in the conceptualization of listening. *Imagination, Cognition and Personality*, *30*, 301–339. doi:10.2190/IC.30.3.f

Davis, J., Thompson, C. R., Foley, A., Bond, C. D., & DeWitt, J. (2008). An examination of listening concepts in the healthcare context: Differences among nurses, physicians, and administrators. *International Journal of Listening*, *22*, 152–167. doi:10.1080/109040/10802181524

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, *103*, 265–275. doi:10.1037/0033-2909.103.2.265

Hair, J. F., Tatham, R. L., Andersonn, R. E., & Black, W. (1998). *Multivariate data analysis*. Boston: Prentice Hall.

Imhof, M., & Janusik, L. (2006). Development and validation of the Imhof-Janusik Listening Concepts Inventory to measure listening conceptualization differences between cultures. *Journal of Intercultural Communication Research*, *35*, 79–98. doi:10.1080/17475750600909246

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments and Computers*, *32*, 396–402.

Scale

Listening Concepts Inventory (LCI; LCI-R) (Imhof & Janusik, 2006; Bodie, 2010)

Source: Imhof and Janusik (2006) and Bodie (2010).

Instructions

The questionnaire tries to understand how people like you think about the word "listening." Below are listed several activities that may or may not be similar to what you think of when you think of listening. We are interested in your opinions on this matter.

For each activity listed, please consider the degree to which it is similar to listening. The more you think that activity is similar to listening, the higher the number you should circle. The less you think that activity is similar to listening, the lower the number you should circle.

Take for example the activity "Smelling." If you think "smelling" is an activity identical to listening, you would circle 5. If you think "smelling" is only somewhat related to listening, you might circle 3.

There are no right or wrong answers, just your opinion. Circle the number closest to your view for each activity.

1 = Not at all similar
2 = Somewhat related
3 = Rather similar
4 = Almost identical
5 = Identical

<u>Organizing Information (Information Acquisition)</u>

1) Organizing information
2) Gathering information
3) **Retaining information**
4) **Storing information**
5) **Drawing conclusions**
6) Apprehending
7) Comparing
8) **Becoming aware**

<u>Relationship Building Listening (Relationship-building)</u>

9) **Bonding**
10) Accepting
11) **Comforting**
12) Socializing
13) **Helping**
14) Welcoming
15) Minding

<u>Learning & Integrating Information (Learning)</u>

16) **Learning**
17) **Interpreting**
18) **Analyzing**
19) **Understanding**
20) Making gestures
21) Selecting
22) Appreciating
23) Decoding
24) Responding
25) Making inferences
26) Enjoying

<u>**C**ritical Listening (Critical)</u>

27) **Arguing**
28) Inquiring
29) Testing
30) **Conceding**

31) **Being critical**
32) **Answering**
33) Obeying

Additional conceptualizations tested in the original Imhof & Janusik study (2006), but excluded from the subscales of the LCI and LCI-R

34) Hearing
35) Straining your ears
36) Smelling
37) Taking in
38) Ignoring
39) Memorizing
40) Showing empathy
41) Sharing
42) Evaluating
43) Overhearing
44) Imaging
45) Discriminating
46) Eavesdropping
47) Constructing
48) Smiling
49) Perceiving
50) Thinking
51) Make an effort
52) Receiving
53) Feeling
54) Sensing
55) Attending
56) Lending an ear
57) Making sense
58) Being mentally active
59) Problem solving
60) Agreeing
61) Relaxing
62) Communicating
63) Confronting
64) Observing
65) Interrogating

Note: Instructions for the original LCI: "Listening is a rich and multifaceted concept and may be represented in various ways, e.g., depending on each individual's experience and culture. Since everybody has their own expertise in listening, the objective of this questionnaire is to gain a survey of listening concepts. Please consider the degree to which the following activities are similar to listening:".

Subscale titles in parentheses reflect the subscale descriptor of the LCI-R (Bodie, 2010).

Items in bold are elements of the 15-item LCI-R (Bodie, 2010).

Profile 33

Listening Fidelity (LF)

(Mulanax & Powers, 2001)

Profiled by: Margarete Imhof, PhD

Gutenberg University, Mainz, Germany

Construct

Listening Fidelity is conceptually defined as “the degree of congruence between the cognitions of a listener and the cognitions of a source following a communication event” (Mulanax & Powers, 2001, p. 70).

Instrument Type

Cognitive Assessment

Description

The Listening Fidelity (LF) test requires listeners to draw a set of geometric forms on an 8.5" × 11" sheet of paper according to video-recorded instructions. The LF test represents an objective performance measure of listening competence that was designed to complement the Basic Communication Fidelity (BCF) test (Powers & Lowry, 1984), which focuses on the speaker’s oral communication performance.

Administration

To administer the measure of LF, researchers will need (a) a standard stimulus, (b) blank recording sheets, (c) pens or pencils, and (d) the capacity to code participant drawings. In the standard version, participants are asked to listen to a recorded description of a

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

series of geometrical shapes and to draw these shapes on a blank piece of paper (8.5" × 11"). The instructed goal is to reconstruct the original set of shapes as accurately as possible. The idea is to test if, and to what degree, the receiver is able to comprehend and reproduce the cognitions that are communicated by the speaker.

A video recording of the standard video-assisted instructions is available here: https://vimeo.com/154069191. General scoring is described in the Scoring section, and a copy of the diagram is provided at the end of this profile, along with specific scoring information. If a different type of instruction is used (i.e., an instructor reading the instructions), variations in speech rate, accent, accuracy, comprehensibility, and consistency of the verbal input may affect test results.

Scoring

Drawings of the complex geometric figure produced by listeners are scored according to the accuracy with which the drawings reflect the original image. The listener earns points for: (a) drawing the figure in the appropriate quadrant of the paper, (b) drawing the correct form (triangle, circle) and size of individual shapes that make up the figure, and (c) correctly reproducing relations among shapes (connections and alignments). The full scoring rubric is included along with the geometrical figure at the end of this entry. A high score indicates a high degree of congruence between the mental image of the speaker and that of the listener. Although the LF test claims to measure listening performance objectively, the scoring guidelines advise test administrators that "good judgment must be used in determining points; some of these criteria are subjective evaluations" (M. Fitch-Hauser, personal communication, January 2014). Because of the qualitative nature of interpreting drawings, it is important to establish clear and consistent scoring procedures and to use multiple raters so measures of intercoder reliability can be calculated.

Development

The LF test was developed by Mulanax and Powers (2001) as a complement to the Basic Communication Fidelity assessment, which focuses on the oral communication skills of a speaker. The actual form of the LF test was patterned after the Geometric Figures Test proposed by Brilhart (1965), who also aimed at tapping into communication accuracy from both the speaker's and the listener's perspective. Mulanax and Powers (2001) pursued several goals: (a) to measure the decoding skills of the receiver in oral communication, (b) to focus on current cognitive processes associated with creating a mental image from an oral message and to remove the influence of reading/writing skills and memory and information retention capacity, and (c) to control for possible confounds associated with the sender (e.g., nonverbal, accents, speaking rate, etc.). Controlling for source characteristics allows individual differences in receivers to be more clearly reflected in the variance of the LF test scores. Additional publications review and discuss the theoretical underpinnings and possible extensions of the LF test and the LF construct (Powers & Bodie, 2003; Powers & Sawyer, 2011; Powers & Witt, 2008).

Reliability

Reliability of scores is assessed as interrater consistency as multiple raters are used for scoring a subset of collected LF data. Consistency for a single rater can also be estimated if she is asked to assess a single figure multiple times after an appropriate time lag. Mulanax and Powers (2001) do not report assessments of interrater agreement in their original publication; neither do they mention if multiple raters were used. To ensure LF can be measured in a reliable manner, more work needs to be done to standardize the scoring procedure and to develop appropriate methods to train raters in scoring the figures.

Validity

Concurrent validity of the LF test has been investigated. Mulanax and Powers (2001) found LF and Receiver Apprehension (RA) inversely related. According to RA studies, message processing suffers when communicators express fear of communication failures, such as misinterpreting, inadequately processing information, not understanding, and not being able to adjust to the emotional content of a message (Ayres, Wilcox, & Ayres, 1995; Buhr & Pryor, 1988; see Informational Reception Apprehension profile, Profile 24). The results of the study by Mulanax and Powers (2001) suggest that, indeed, individuals with high RA produce lower LF test scores. The authors proposed that apprehensive individuals have reduced cognitive capacity, leading them to be more distracted and subsequently perform more poorly on the test.

To illustrate the empirical potential of the LF test, two additional studies are reviewed (Fitch-Hauser, Powers, O'Brien, & Hanson, 2007). The first study investigated the effect of three varieties of verbal input on LF. The oral messages that represent the LF instructions were produced with varying transmission speed and level of detail in the description. Although all messages contained the full information, the messages differed in duration from 132 sec (227 words, representing low potential) to 190 sec (338 words, representing moderate potential) and 280 (614 words, representing high potential). LF scores differed significantly between groups (F [2, 65] = 5.346, $p = .008$, $\eta^2 = .20$). The authors concluded that the LF test is sensitive to critical differences between sources. As a side note, no significant differences were found between LF scores of men and women.

Fitch-Hauser *et al.* (2007) also looked at the relationship between LF and other listening tests, that is, the Listening Styles Profile (LSP; Watson, Barker, & Weaver, 1995) and Watson-Barker Listening Test (WBLT; Watson & Barker, 1988). Although no significant correlations could be detected between the four dimensions of listening styles, LF scores were correlated with the subscales of the WBLT. However, later research has called into question the viability of these early listening measures (see Profile 36 for the LSP and Profile 64 for the WBLT).

Availability

The materials used in past work are available online. The scoring rubric and geometric figure used in past studies are reproduced below, with permission. All materials are free to use for research purposes with appropriate citation. All other uses require permission.

Sample Studies

Listening Fidelity represents a potentially important aspect of the listening process, helping researchers to understand the dynamics of the sender–receiver interaction. As seen in this profile, research suggests that the LF test has construct validity, as studies have confirmed the predicted relationships with other listening and communication measures (e.g., Fitch-Hauser *et al.*, 2007). However, the potential of the LF test has not been fully realized for both theoretical contexts and practical applications (Powers & Bodie, 2003). At this time, additional research is needed to better understand the place of Listening Fidelity in the broader context of the listening process.

Critique

Notably, the LF test rests on a sound theoretical foundation. It was developed as an extension of the basic communication fidelity test, and the results can be viewed in the context of both sending and receiving messages.

As Worthington and Fitch-Hauser (2012) noted, the LF test is one-dimensional in nature because it singles out one aspect of the listening process, namely, the reconstruction of a mental image by the receiver; thus, it fails to encompass the complexity of the entire listening process. However, this unidimensionality may be considered an asset because the test addresses the translation of a mental image from sender to receiver. If the listening process breaks down when LF can be assured, we have a good reason to look for the cause of the communication failure elsewhere. Given that "accuracy in listening should be the base on which a theory competence in listening is founded" (Powers & Bodie, 2003, p. 24), an instrument to measure degrees of accuracy is certainly indispensable.

The LF test also attempts to avoid confounding constructs and keeps the focus on the cognitive process of listening. The ongoing issue with many tests of listening is that they include memory, reading, and/or writing skills, and are based on prior subject knowledge. The LF test aims at limiting the influence of these variables.

Finally, the LF test is potentially universal in its application because it uses common knowledge (shapes and sizes) and skills (rough drawing) and does not require specific subject-related competences. Of course, test administrators must make sure that test takers have the knowledge of geometric shapes, both linguistically and conceptually. Moreover, vocabulary can still influence results. For instance, some listeners may be more familiar with different types of triangles (e.g., isosceles and equilateral) that could make drawing more accurate and instructions more streamlined.

Although the clarity of instructions ensures that test administration is standard, the issue of ecological validity must still be raised. For example, LF test scores largely depend on understanding instructions, whereas comprehension may be affected by the complexity and subtext (e.g., irony, emotional overtones).

The LF test can be useful when listener–speaker congruence is relevant, as, for example, in theoretical as well as practical investigations of misunderstandings (Powers & Sawyer, 2011), detecting problems in the physical environment that impair listening skills, and identifying LF in second-language learning (Cook, Powers, Fitch-Hauser, & Worthington, 2009).

The potential of LF for further research in listening needs additional exploration. Drawing on a variety of theoretical perspectives, Powers and Sawyer (2011) suggested that LF should be a function of age and sensory and neurological capabilities of the receivers across the lifespan, higher order thinking skills (in particular regarding the awareness of different perspectives), prior subject knowledge and expertise, and motivational resources (e.g., willingness to listen and give attention).

References

Ayres, J., Wilcox, A. K., & Ayres, D. M. (1995). Receiver Apprehension: An explanatory model and accompanying research. *Communication Education*, *44*, 223–235. doi:10.1080/03634529509379013

Brilhart, B. L. (1965). The relationship between some aspects of communicative speaking and communicative listening. *Journal of Communication*, *15*, 35–46.

Buhr, T. A., & Pryor, B. (1988). Receivers' apprehension and need for cognition. *Psychological Reports*, *62*, 995–996. doi:10.2466/pr0.1988.62.3.995

Cook, J., Worthington, D. L., Fitch-Hauser, M., & Powers, W. (2008, November). Listening fidelity: Testing for cross-cultural differences in U.S. Caucasian & U.S. Hispanic listeners. Paper presented at the National Communication Association convention, San Diego, CA.

Fitch-Hauser, M., Powers, W. G., O'Brien, L., & Hanson, S. (2007). Extending the conceptualization of Listening Fidelity. *International Journal of Listening*, *21*, 81–91. doi:10.1080/10904010701301982

Mulanax, A., & Powers, W. G. (2001). Listening Fidelity development and relationship to receiver apprehension and locus of control. *International Journal of Listening*, *15*, 69–78. doi:10.1080/10904018.2001.10499045

Powers, W. G., & Bodie, G. D. (2003). Listening Fidelity: Seeking congruence between cognitions of the listener and the sender. *International Journal of Listening*, *17*, 19–31. doi:10.1080/10904018.2001.10499045

Powers, W. G., & Lowry, D. N. (1984). Basic communication fidelity: A fundamental approach. In R. N. Bostrom (Ed.), *Competence in communication* (pp. 57–71). Beverly Hills, CA: Sage.

Powers, W. G., & Sawyer, C. R. (2011). Towards a theoretical framework for Listening Fidelity research. *Human Communication*, *14*, 313–326.

Powers, W. G., & Witt, P. L. (2008). Expanding the theoretical framework of communication fidelity. *Communication Quarterly*, *56*, 247–267. doi:10.1080/01463370802240957

Watson, K. K., & Barker, L. L. (1988). Listening assessment: The Watson-Barker listening test. *International Journal of Listening*, *2*, 20–32. doi:10.1080/10904018.1988.10499095

Watson, K. K., Barker, L. L., & Weaver, J. (1995). The Listening Styles Profile (LSP-16): Development and validation of an instrument to assess four listening styles. *International Journal of Listening*, *9*, 1–13. doi:10.1080/10904018.1995.10499138

Worthington, D. L., & Fitch-Hauser, M. E. (2012). *Listening: Processes, functions, and competency*. Boston: Allyn & Bacon.

Scale

Listening Fidelity Materials

Source: Powers. Reproduced with permission of Dr. Will Powers.

Instructions

In the original video, available online at https://vimeo.com/154069191, *this description is read by a white, college-aged male speaking in a Standard American English accent. The rate of speaking is approximately 140 words per minute. He sits at a table, holding a sheet of paper from which he reads the following instructions. Along with a transcript, speaking time and pauses are presented below. Timing (in seconds) of the reading and pauses are indicated.*

(0:01) Before we begin let me say that we are going to be drawing a series of connected geometric shapes. Please do not begin drawing a shape until you have heard all the instructions regarding the size and its position relative to the other shapes and the paper itself.

(0:21) Let me say also that all four squares are the same size, all two rectangles are the same size, and all three triangles are the same size. Let's begin.

(0:34) Find the center at the top of the paper and move down two inches. Draw an equilateral triangle, one inch tall with the bottom right corner at this center point. (Pause 0:03).

(0:48) Next, touching the bottom right hand corner of that triangle will be the point of a new triangle, which will also be a one-inch equilateral triangle. The point of the triangle will be pointing towards the top of the page. (Pause 0:02).

(1:09) Next, draw a one inch by one inch square. The upper left hand corner of the square will begin on the bottom center of the second triangle. (Pause 0:03).

(1:24) Now, draw another one inch by one inch square with the upper right hand corner starting at the bottom of the left hand corner of the previous square. The sides of the square will run parallel to the sides of the paper. (Pause 0:03).

(1:44) Next, draw a rectangle two inches wide by one inch tall. The square above it will be on the middle of the top line of this rectangle. The rectangle will equally protrude on each side of the square. (Pause 0:03).

(2:03) Now, on the base of the rectangle draw another one inch equilateral triangle with the point facing downward. The base of the triangle will be centered on the rectangle. (Pause 0:03).

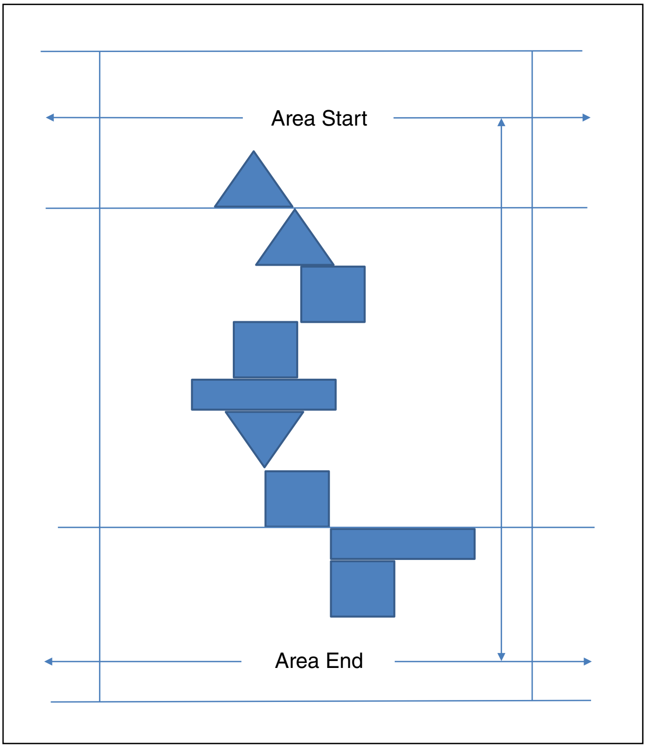
(2:18) Using the point of the triangle facing the bottom of the page as the upper left hand corner draw a new one inch by one inch square. (Pause 0:03).

(2:30) Now, the bottom right hand corner of your square will be the top left hand corner of a new two inch wide by one inch tall rectangle. There will be no shapes directly above touching the rectangle. (Pause 0:04).

(2:48) Lastly, draw a one inch by one inch square by extending the left side of the rectangle you just drew and using it as the left side of the new square. (Pause 0:03).

(3:02) This concludes our communication exercise. Thank you for your participation. (3:10)

Diagram



Note: The lines included in this figure are not to be drawn; they are presented for scoring purposes.
Source: Powers (2015). Reproduced with permission of Sage Publications.

LF Scoring Procedures (completed on 8.5″ × 11″ paper) [1]

Positioning and size analysis = 4 points

Structure:

1 point if the initial shape appears in the upper 1/4 of the paper (the instructions say to start two inches from the top)
1 point if the final shape appears in the lower quadrant of the paper
1 point if all the shapes appear in the centered 5.5” of the paper (instructions state to begin in the middle)
1 point if all the shapes are connected to another shape

1 Significant training is needed, particularly when using multiple coders, to increase consistency across coders and to enhance reliability of the evaluations (M. Fitch-Hauser, personal communication, January 2014).

Shapes:

1 point for each correct shape = 9 points; there should be four squares, two rectangles, and three triangles.

Relationships:

2 points for the correct connection and alignment of each succeeding pair of shapes = 16 points

RANGE = 0 - 29

Data entry:

Position: up to 4 points
Shape: up to 9 points
Relationship: up to 16 points

Code for Relationships:	3 = 10 points
	5 = 12 points
	7 = 14 points
	9 = 16 points

Profile 34

Listening Practices Feedback Report (LPFR)[1]

(Brandt, Brandt, Emmert, & Emmert, 1992)

Profiled by: Debra L. Worthington, PhD

Auburn University

Construct

The Listening Practices Feedback Report (LPFR) provides a 360° assessment of a business leader's perception of their own listening ability as well as perceptions from others.

Instrument Type

Self- and Other-Report

Description

The LPFR is a 28-item instrument that allows organizational leaders to self-assess and for their colleagues and subordinates to provide feedback on six facets of listening (Brandt Management Group, 1999, pp. 4–12):

- **Attention**: Giving full attention to the speaker without being preoccupied; avoiding interrupting; maintaining eye contact; permitting proper closure or agreement before changing topics; minimizing calls and distractions.

1 This profile reflects the 1998 revised version of the LPFR. Descriptions and assessments are based on secondary sources, as original materials and conference papers were unavailable for review. For readability, the profile cites original sources where secondary sources provide a level of confidence in the material being described. The reliance on secondary sources means that this profile may be incomplete, particularly in the area of scoring. All sources—original and secondary—are provided in the bibliography.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

- **Empathy**: Repeating, paraphrasing, or summarizing comments to ensure understanding; placing oneself in another person's position, understanding their concerns and feelings; encouraging others to share their views; considering the subject under discussion before responding; and correctly anticipating conversational flow.
- **Memory**: Following agreed-upon instructions or guidelines; accurately recalling comments or positions at a later date; accurately relating messages to others; taking notes when appropriate.
- **Open mind**: Appearing to listen free from personal bias; considering the content and logic without criticizing delivery, appearance, grammar, vocabulary, and so on; avoiding becoming emotional or defensive when encountering a difficult situation; balancing listening and talking; avoiding emotion-packed (trigger) words, phrases, or clichés.
- **Respect**: Keeping confidences; sincerely listening beyond just going through the motions; taking time and having patience during conversations/meetings; acknowledging others' ideas/words regardless of business, social, or economic status.
- **Response**: Asking questions to clarify technical or misunderstood points; following up with prompt actions; showing appropriate nonverbal responses; preparing or becoming properly informed as necessary; smiling and acknowledging humorous remarks.

Administration

The LPFR is a self-administered questionnaire that takes approximately 10 minutes to complete. In addition to administrating the questionnaire to (typically) a supervisor, multiple colleagues, associates, and subordinates complete an other-report version of the survey (typically 5–10; Brandt, 2003; McCord, 2011; Orick, 2002). Increasing the number of other-reports increases the robustness of the feedback and subordinate anonymity (McCord, 2011). Delivery has included in-person, postal mail, and online (Brandt *et al.*, 1992; McCord, 2011; Orick, 2002). Several studies reported that the supervisor being assessed chose the individuals who completed the other-report version (Emmert, Emmert, & Brandt, 1993; Orick, 2002). Brandt (2003) argued that "a person cannot be given a feedback report without comment and preparation" (p. 16).

Scoring

Responses to LPFR items range from 1 (*almost never*) to 10 (*almost always*), although some modified forms have use 5-point scaling (Shoho *et al.*, 2006). Responses for each dimension are averaged, and averages are computed for self-report and other-report versions (Emmert *et al.*, 1993). Global scores above 225 (80th percentile) are indicative of highly effective listeners, whereas those below 196 (the 70th percentile) are considered average listeners (Brandt Management Group, 1999, as cited in McCord, 2011). It is unclear how cutoff scores were determined.

Development

The LPFR was developed by utilizing multiple samples (860 participants) drawn from 22 companies representing an assortment of industries. Employees were asked to identify six colleagues—three good listeners and three poor ones—and to provide five characteristics associated with each type. Responses were used to identify the categories outlined above, although decision criteria are unclear. In addition, the number and descriptions of indices have changed over time (see, e.g., Brandt *et al.*, 1992; Emmert *et al.*, 1993; Emmert, Emmert, Brandt, Watson, & Barker, 1994). Notably, using the same items, different factor solutions were generated for respondent scores (acceptance of the listening role, avoiding emotional responses, memory, empathic nonjudgmental attitude, listener as speaker, and business-like/professional listening), averaged associates' scores (acceptance of listening role, appropriate professional listening, supportive listening), and respondent/associate LPFR difference scores (non-status information seeking, supportiveness, attentiveness, gatekeeping, listening acceptance, lack of interruptions, listener speaking and deliberation) (Emmert, Emmert, & Brandt, 1992). These different factor solutions make it difficult to assess reported results, especially because items load differently on what are seemingly the same factors.

As part of a study on differences in listening practices of male and female business leaders, Emmert *et al.* (1994) conducted an unspecified factor analysis with varimax rotation on the original items, which resulted in dropping three original factors from analysis. Lack of specifics of the analysis and changes in factor labels make it difficult to determine which factors were actually dropped. Emmert *et al.* (1994) reported results of an earlier FA, which found the six self-report factors accounted for 50% of item variance, whereas the three factors of the associates' perceptions accounted for 59%. Unfortunately, early research on the development of the LPFR is not readily available for review.

Reliability

Reliability estimates have varied widely. Emmert *et al.* (1992) and McCord (2011) reported Cronbach alphas in the 80s for the global self-report version and in the 90s for the observer version. Individual indices are much lower, with most estimates listed in the 50s and 60s (see Emmert *et al.*, 1992; Miller, 2008; Orick, 2002; Williams, 2006; for an exception, see Lieb, 2014). More importantly, if the first-order factors do not load on a single second-order factor (a model never tested), then an overall estimate of reliability makes little sense.

Validity

Validity evidence for the LPFR remains largely unavailable. Although the measure may have face validity, only one study has administered it with other listening measures—the individual Listening Styles Profile (Emmert *et al.*, 1994), whose own validity has been questioned (see the Listening Styles Profile, Profile 36).

Availability

Janice Brandt has provided permission for the copyrighted scale, reproduced here, to be used for research purposes. Original and modified versions of the scale have been used in several master's and doctoral theses (see References).

Sample Studies

Studies have used the LPFR to examine listening practices of business leaders in hotel management, the US Coast Guard, and nonprofit contexts (Ellis, 2003; McCord, 2011; Miller, 2008; Orick, 2002; Williams, 2006).

McCord (2011) tested the relations between leadership style (transformational, transactional, or passive/avoidant) and leader listening practices. With the exception of the Open Mind dimension, he reported weak to moderate positive relations between transformational leadership scores and the LPFR indices. In contrast, the passive/avoidant leadership style was inversely related to the same indices.

Miller (2008), examining the listening practices of African-American leaders/managers, reported a number of differences in self and others' perceptions of listening behavior. For example, subordinates rated their supervisors' attention ($M = 7.06$) higher than supervisors rated their own attention behaviors ($M = 6.93$), $t(185) = -2.04$, $p < .05$. Similar small, statistically significant differences were found for all but the Empathy dimension.

Orick (2002) explored differences in associate perceptions of leader listening practices based on the leader's profession (staff, management, and administration), time in the position, and a variety of demographic items. Her primary finding: Hispanic women averaged higher on the Memory index, $F(2, 89) = 5.41$, $p < .05$.

Examining sex-related differences in supervisor listening practices, Emmert *et al.* (1993) compared the averaged self-report and averaged other-report scores of men and women.

Williams (2006) examined the effect of listening training, educational level, and workplace tenure on managers' listening practices. They found that managers typically had high mean LPFR scores (see also Orick, 2002) and that listening training did not discriminate scores, nor did education level or time on the job.

Lieb (2014) examined the relation between LPFR dimensions and employees' affective organizational commitment. Notably, he did not confirm the LPFR factors as part of his analysis. Only two LPFR indices were associated with organizational commitment (Memory, $r = .23$; Open Mind, $r = .27$), but neither emerged as predictors in a regression equation, $F(2, 60) = 2.32$, $p > .05$. High multicollinearity across the indices likely affected the results.

Critique

The LPFR is an improvement over simple self-report measures because it provides feedback from others and the opportunity for self-reflection. In order for it to be accepted as a research instrument, however, its validity portfolio must be fully established. This measure faces several issues:

- Although the use of 10 response categories allows for greater nuances in assessment, the wording traditionally utilized may be confusing (e.g., 4 = *once in a while* vs. 5 = *occasionally*) and introduce error in responses.
- Supervisors who pick the subordinates to complete the observer version may choose individuals who are more likely to provide favorable ratings.
- Virtually no researchers using the LPFR confirmed its proposed dimensions.

- Internal reliability estimates of each dimension are seldom reported.
- There is a lack of normed data for the LPFR.

Researchers are urged to confirm and report the factor structure of the LPFR prior to conducting statistical analyses to aid in building its validity and reliability portfolio.

References

Brandt, J. (2003). Leading through listening: Utilizing the LPFR 360° to evaluate your skills. *The Listening Professional*, *2*(1), 16–17.

Brandt, J., Brandt, R., Emmert, P., & Emmert, V. (1992, March). *The development of the listening practices feedback report*. Paper presented at the International Listening Association convention, Seattle, WA.

Brandt Management Group. (1999). *The complete guide: Listening Practices Feedback Report- 360*. Richmond, VA: Brandt Management Group.

Emmert, P. (1996). President's perspective. *ILA Listening Post*, *56*, 2–3.

Emmert, V. (1986, March). *Listening in interpersonal relationships*. Paper presented at the International Listening Association convention, San Diego, CA.

Emmert, P., Emmert, V., & Brandt, J. (1992, March). An examination of the dimensional structure of the Listening Practices Feedback Report. Paper presented at the International Listening Association convention, Seattle, WA.

Emmert, V., Emmert, P., & Brandt, J. (1992, March). *An analysis of respondent/average associate differences on the Listening Practice Feedback Report*. Paper presented at the International Listening Association convention, Seattle, WA.

Emmert, P., Emmert, V., & Brandt, J. (1993). An analysis of male-female differences on the Listening Practices Feedback Report. *International Journal of Listening*, *7*, 43–55. doi:10.1080/10904018.1993.10499124.

Emmert, V., Emmert, P., & Brandt, J. (1994). *Psychometric properties of the Listening Practice Feedback Report: A validity and reliability study*. Unpublished report. Richmond, VA: Brandt Management Group.

Lieb, K. (2014). *Exploring perceived leader listening dimensions as predictors of employees' affective organizational commitment*. Unpublished doctoral dissertation. Barry University, Miami, FL.

McCord, S. K. (2011). *Leadership style and listening practices of IMB team leaders: A correlational study*. Unpublished doctoral dissertation. Southern Baptist Theological Seminary, Louisville, KY.

Miller, K. L. (2008). *A quantitative study of the listening practices of African-American leader/managers of faith-based nonprofit organizations in the US*. Unpublished doctoral dissertation. Capella University, Minneapolis, MN.

Orick, L. M. (2002). *Listening practices of leaders*. Unpublished doctoral dissertation. The University of New Mexico, Albuquerque.

Shoho, A. R., Woods, J., & Smith, P. (2006). *Is anyone listening? The dichotomous perspective of administrator listening skills*. Paper presented at the University Council for Education Administration conference, San Antonio, TX.

Williams, B. L. (2006). *The listening behaviors of managers in the workplace*. Unpublished thesis. George Washington University, Washington, DC.

Scale

Listening Practices Feedback Report, Observer Version

Source: Brandt (1987). Reproduced with permission of Janice Brandt.

Directions

There is no “right” or “wrong” answer to the questions. They describe behavioral criteria used in determining how people are perceived as listeners. The questions do not determine listening ability.

You have been asked to complete this questionnaire regarding your leader/manager with whom you frequently interact. You will be asked to answer questions about how you perceive, how you feel about his/her listening habits. Your responses will be completely confidential.

Some questions are reversed for research validity. Please read each item carefully. Then determine from the rating scale how frequently they typically use the behaviors described in most situations with most people. Do not confer or consult with co-workers: your own perception is needed.

Choose one of the ten possible responses for each of the 28 listening practices.

1 = Almost Never
2 = Rarely
3 = Seldom
4 = Once in a While
5 = Occasionally
6 = Sometimes
7 = Fairly Often
8 = Usually
9 = Very Frequently
10 = Almost Always

The Listener …

<u>Attention Index</u>

1) Gives full attention and is not preoccupied with other concerns.
2) Changes the topic before proper closure or agreement.(R)
3) Maintains comfortable eye contact with speaker.
4) Allows others to finish without interrupting.
5) Holds outside calls and distractions to a minimum during meetings and conversations.

<u>Empathy Index</u>

6) Encourages others to give their views on subjects under discussion.
7) Correctly anticipates where the conversation is going.
8) Repeats, paraphrases, or summarizes comments to ensure understanding.
9) Thinks about the subject under discussion before responding.
10) Places him/herself in others’ position and understands their concerns and feelings.

Memory Index

11) Takes notes when notes are appropriate.
12) Accurately recalls comments or positions at a later date.
13) Accurately relates messages to a third party.
14) Produces results consistent with agreed upon instructions or guidelines.

Open Mind Index

15) Talks more than listens.(R)
16) Appears to listen with an open mind free from personal biases.
17) Becomes defensive or emotional when encountering a difficult situation.(R)
18) Avoids emotion-packed (trigger) words, phrases, or clichés.
19) Considers content and logic and is not critical of others' delivery, appearance, grammar, vocabulary, etc.

Respect Index

20) Seems hurried or impatient during conversations and meetings.(R)
21) Respects others' ideas and words regardless of business, social, or economic status.
22) Keeps a confidence.
23) Sincerely listens without going-through-the-motions.

Response Index

24) Shows appropriate non-verbal responses, such as nodding and facial expressions.
25) Asks relevant questions for clarification of points that are technical or misunderstood.
26) Prepares or becomes informed when such preparation or knowledge is necessary.
27) Follows up with prompt actions.
28) Smiles or otherwise acknowledges humorous remarks.

Note: Factor labels should be removed and items randomized prior to administration. Items marked with (R) should be reverse coded prior to scoring. Copyright 1987 and revised 1998, Brandt Management Group, Inc., Richmond, VA. Used with permission. For the self-report version of the measure, the instructions are revised to begin with "You have been asked to complete this questionnaire regarding your leader/manager with whom you frequently interact" and "I, the listener" replaces "The listener."

Profile 35

Listening Span Tests

Profiled by: Margarete Imhof, PhD

Gutenberg University, Mainz, Germany

Construct

As an indicator of auditory short-term storage capacity, *listening span* is defined as the maximum amount of information an individual can store in the so-called phonological loop (see the Description section). Listening Span tests qualitatively differ from Conversational Listening Span tests (see Profile 12).

Instrument Type

Cognitive Assessment

Description

As noted above, Listening Span is defined as the maximum amount of information an individual can store in the so-called phonological loop. The phonological loop is an aspect of the larger construct of working memory (WM), which is critical for selecting, processing, and manipulating information (Alloway & Copello, 2013). WM consists of three components (Baddeley, 1986, 2003; Baddeley & Hitch, 1974): the central executive control, the visuospatial sketchpad, and the articulatory or phonological loop. The phonological loop serves as short-term storage for acoustic stimuli (speech and sound). Its purpose is to rehearse and recycle auditory information for immediate recall and to hold relevant information active so that it may be used for higher level processing (e.g., comprehension, problem solving, and learning).

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

The constructs of working memory and short-term memory (STM) are frequently confused or even used synonymously (Berti, 2010). The critical distinction between the two is that working memory includes attentional control processes for mental operations in addition to short-term storage (Alloway & Copello, 2013; Cowan, 1999; Engle, Kane, & Tuholski, 1999). Strictly speaking, listening span is an indicator of auditory short-term storage capacity.

Listening span tests present sets of items that can vary in content. For instance, some include words, others include numbers, and still others include sentences or larger chunks of information. In a typical listening span task focusing on auditory STM, subjects listen to a sequence of items (i.e., a set) and then repeat back the items from the list in serial order. The set sizes increase across trials. If test takers need to manipulate the items (e.g., the task is to recall the items in reverse order), executive functions of WM are required in addition to storage. A systematic variation of listening span task characteristics generates differential degrees of task complexity (LaPointe & Engle, 1990), which can be used for diagnostic purposes.

Individual listening span is determined by the maximal set size that a person is able to recall in perfect order. Both in research and for practical purposes, different types of items are utilized, including sets of numbers (also known as *digit span*) and verbal items. In the case of verbal items (words and sentences), the same sets of items have been used to assess both reading and listening span (Daneman & Carpenter, 1980).

Administration

Listening span tasks are ideally presented as recorded stimuli to control for presentation speed and clarity. It also is critical that the intervals allotted for recall are consistent across individuals. In addition, listening span task performance has been shown to be susceptible to situational interferences (e.g., changes in volume, background noise, and acoustic distractions), making it critical for the test administrator to create a testing environment free of disturbances (Petersen, Wostmann, Obleser, Stenfelt, & Lunner, 2015; Pichora-Fuller, Schneider, & Daneman, 1995). As a rule, the final set is defined if a person fails to reproduce a set of items in three subsequent presentations. For example, when a person has mastered the task to recall the last four items of a list, but fails to reproduce the last five items of the next list, the listening span is determined as four, which, according to experimental research (Cowan, 2001), should be a sound estimate of what can be expected as the average listening span.

Scoring

The scoring of listening span tests is rather straightforward, as it is basically a count of the number of items that a person is able to recall from a list. The critical issue is more about the definition of what is considered an *item*, because the number of items that a person is able to actively recall depends on item characteristics, such as word length and number of syllables in a word.

In a test that uses consistent items such as digits or monosyllabic words, the listening span score is the number of items a person can reliably reproduce after listening to the item presentation.

Development

Theoretical and empirical work on working memory (Baddeley, 2003; Baddeley & Hitch, 1974) has produced a large body of literature that has explored the nature of cognitive components critical for information processing and learning. Daneman and Carpenter (1980) suggested that specific learning difficulties could be reasonably well explained in terms of individual differences in STM and WM capacity, both the phonological and the visual components.

Listening span tasks have frequently been included as subscales in comprehensive cognitive assessments. The Wechsler Intelligence Scale for Children (WISC; Wechsler, 2014) includes a rough estimate of a Working Memory Index, but it fails to distinguish between verbal and visual material and between STM and WM functions (Alloway, Gathercole, & Pickering, 2006). A more fine-tuned diagnostic instrument has been developed by Alloway (2007). Her instrument, the Automated Working Memory Assessment (AWMA), is available as a PC-based assessment (Alloway, 2012) and has been translated into a Spanish version (Injoque-Ricle, Calero, Alloway, & Burin, 2011). Similar principles have been applied in other test batteries that are more specifically tailored to different age groups (Hasselhorn *et al.*, 2012) and to individuals with specific medical conditions (Ivanova & Hallowell, 2014). One popular listening test, the Kentucky Comprehensive Listening Test (KCLT), also used a listening span task to measure short-term listening ability (Bostrom & Waldhart, 1983).

Janusik (2007) developed the Conversational Listening Span (CLS) measure, which conceptually differs from traditional listening span tests (see Profile 12). The CLS measures conversational listening capacity by estimating "the number of items that one can hold active, can paraphrase, and can respond to in the course of a conversation" (Janusik, 2007, p. 144).

Reliability

As a rule, working memory capacity, including listening span, is a trait characteristic, and test–retest reliability can be expected to be rather high. In the case of the AWMA, test–retest reliability was shown to range between .69 and .90 across an interval of 4 weeks. Hasselhorn *et al.* (2012) found test–retest reliabilities from .67 to .89 for 9- to 12-year-olds and from .66 to .85 for 5- to 8-year-olds for an interval of 1–2 weeks. The somewhat low minimum reliabilities may be due to the fact that aspects of WM capacity, and in particular listening span, are susceptible to physiological changes during the day (Morris & Sarll, 2001). As a consequence, it is advised that, in case of low listening span performance, a second testing be administered.

Validity

For the AWMA, Alloway and colleagues investigated construct validity of their test battery and reported that AWMA scores and the WISC-IV Memory Index converge quite substantially (Alloway, Gathercole, Kirkwood, & Elliott, 2008). In addition, behavior ratings of children with WM difficulties and measures of WM were also highly correlated (Gathercole *et al.*, 2008). Indicators for predictive validity of pertaining measures also can be found in studies that used WM scores as predictors for learning difficulties (for a review, see Cowan, 2014).

Availability

Standardized listening span tests are available as assessments through major publishers. For example, they can be administered by professionals, such as school psychologists and teachers. The tests come with norms for a variety of age groups. Tests can be located on the following websites:

- Automated Working Memory Assessment (2012) (1st ed.; ages 4–79; multiple languages available; available at www.pearsonclinical.com).
- Wechsler Intelligence Scale for Children® (2014) (5th ed.; ages 6–16; available at www.pearsonclinical.com).
- Arbeitsgedächtnisbatterie für Kinder von 5 bis 12 Jahren (AGTB 5–12) (2012) (Working memory testing battery for children ages 5–12; in German; available at http://www.testzentrale.de/).
- The Working Memory Rating Scale (2008) (ages 5–8; available at www.pearsonclinical.com).

Sample Studies

Listening span measures have turned out to be valid predictors of scholastic attainment (Alloway & Alloway, 2010; Daneman & Merikle, 1996; Gathercole, Alloway, Willis, & Adams, 2006). A variety of learning disabilities have been associated with low working memory capacity in both the visual and the articulatory systems. In a longitudinal study, WM and STM measures were taken from children with literacy disorders (Daneman & Carpenter, 1980; Fischbach, Könen, Rietz, & Hasselhorn, 2014). Overall, children with literacy disorders were outperformed by their typical developing peers in all phonological and in dynamic visual-spatial storage and central-executive tasks. The most consistent deficit in children with literacy disorders was found in the storage capacity of the phonological loop. In addition, there is empirical evidence that listening span is critical not only for language-related performance but also for the development of mathematical skills (de Weert, Desoete, & Roeyers, 2013). Although executive functions of WM have been shown in a meta-analysis (David, 2012) to be the strongest predictor for mathematical skills, effect sizes for the impact of phonological loop measures on math performance are still moderate.

Individual differences in L2 learning have frequently been shown to be related to differences in listening span measures (Juffs & Harrington, 2011; Mackey & Sachs, 2012; Martin & Ellis, 2012). Listening span seems to be critical for such diverse aspects as vocabulary learning in L2 and applying grammatical knowledge to infer meaning and structure.

Critique

Listening span has been shown to be a prolific construct to capture individual differences in the capacity to store auditory information. Empirical evidence supports the notion that listening span is distinct from other storage and WM functions (Alloway *et al.*, 2006; de Carvalho, Kida, Capellini, & de Avila, 2014; Lecerf & Roulin, 2006; Siegel, 1994). A broad base of empirical literature illustrates that WM in general, and listening span in particular, are critical for learning and for the development of learning disabilities (Gathercole *et al.*, 2006), for language comprehension (Daneman & Merikle, 1996), and for communication competences as diverse as sentence comprehension, word learning, use of context information (Janse & Jesse, 2014), and perspective taking in a face-to-face conversation (Wardlow, 2013).

A diagnostic measure of an individual's listening span may be important to identify and allocate listening comprehension difficulties of low-performing listeners. Building on these measures, training interventions to foster the size of the individual listening span can be planned. For practical purposes, it could be helpful to know how to tailor communication for listeners with WM limitations (Gathercole & Alloway, 2011).

From the point of view of research, a measure of listening span could be an important covariate to explain variance in measures of listening comprehension. Given the variation of listening span across the lifespan (Alloway & Alloway, 2013; Pichora-Fuller *et al.*, 1995), it might make sense to use pertaining measures for studies that investigate development of communication skills. In addition, current research would need to be expanded to more complex listening situations. Because real-life conversations typically contain redundant information (e.g., in the language, through gestures, and through context), we need to learn more about how listening span limitations affect communication efficiency. For this purpose, listening span measures need to be investigated in more detail; in particular, norms need to be established for a larger variety of age groups. Most listening span tests have been normed for children and youths only. A closer look at the norms provided for the AWMA reveals that, although norms are provided across the lifespan, sample sizes were small, so that additional studies are necessary.

It is interesting to note that listening span measures depend on the type of material used. There is empirical evidence that differences in listening span are found not only when digits and words are used as stimuli, but also when speech and music are presented (Schulze & Koelsch, 2012). Listening span may be an interesting construct to complement the array of factors that determine comprehension and communication competences.

References

Alloway, T. P. (2007). *Automated working memory assessment*. London: Pearson Assessment.

Alloway, T. P. (2012). *Automated working memory assessment–2*. London: Pearson Assessment.

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*, 20–29. doi:10.1016/j.jecp.2009.11.003

Alloway, T. P., & Alloway, R. G. (2013). Working memory across the lifespan: A cross-sectional approach. *Journal of Cognitive Psychology, 25*, 84–93. doi:10.1080/20445911.2012.748027

Alloway, T. P., & Copello, E. (2013). Working memory: The what, the why, and the how. *The Australian Educational and Developmental Psychologist, 30*, 105–118. doi:10.1017/edp.2013.13

Alloway, T. P., Gathercole, S. E., Kirkwood, H. J., & Elliott, J. E. (2008). Evaluating the validity of the Automated Working Memory Assessment. *Educational Psychology, 7*, 725–734. doi:10.1080/01443410802243828

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuo-spatial short-term and working memory in children: Are they separable? *Child Development, 77*, 1698–1716. doi:0009-2910/2006/7706-0014

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829–839. doi:10.1038/nrn1201

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory, VIII* (pp. 47–89). New York: Academic Press.

Berti, S. (2010). Arbeitsgedächtnis: Vergangenheit, Gegenwart und Zukunft eines theoretischen Konstruktes [Working memory: The past, the present, and the future of a theoretical construct]. *Psychologische Rundschau, 61*, 3–9.

Bostrom, R. N., & Waldhart, E. S. (1983). *The Kentucky Comprehensive Listening Test (KCLT)*. Lexington, KY: The Kentucky Listening Research Center.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge: Cambridge University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review, 26*, 197–223. doi:10.1007/s10648-013-9246-y

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466. doi:10.1016/S0022-5371(80)90312-6

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*, 422–433. doi:10.3758/BF03214546

David, C. V. (2012). Working memory deficits in math learning difficulties: A meta-analysis. *International Journal of Developmental Disabilities, 58*, 67–84. doi:10.1179/2047387711Y.0000000007

de Carvalho, C. A. F., Kida, A. de S. B., Capellini, S. A., & de Avila, C. R. B. (2014). Phonological working memory and reading in students with dyslexia. *Frontiers in Psychology, 5*. Retrieved from http://journal.frontiersin.org/article/10.3389/fpsyg.2014.00746/full.

de Weerdt, F., Desoete, A., & Roeyers, H. (2013). Working memory in children with reading disabilities and/or mathematical disabilities. *Journal of Learning Disabilities, 46*, 461–472. doi:10.1177/0022219412455238

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and function of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge: Cambridge University Press.

Fischbach, A., Könen, T., Rietz, C. S., & Hasselhorn, M. (2014). What is not working in working memory of children with literacy disorders? Evidence from a three-year-longitudinal study. *Reading and Writing*, *27*, 267–286. doi:10.1007/s11145-013-9444-5

Gathercole, S. E., & Alloway, T. P. (2011). *Working memory and learning: A practical guide for teachers*. Thousand Oaks, CA: Sage.

Gathercole, S. E., Alloway, T. P., Kirkwood, H. J., Elliott, J. G., Holmes, J., & Hilton, K. A. (2008). Attentional and executive function behaviours in children with poor working memory. *Learning and Individual Differences*, *18*, 214–223. doi:10.1016/j.lindif.2007.10.003

Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A.-M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology*, *93*, 265–281. doi:10.1016/j.jecp.2005.08.003

Hasselhorn, M., Schumann-Hengsteler, R., Gronauer, J., Grube, D., Mähler, C. Schmid, I., Seitz-Stein, K., & Zoelch, C. (2012). *Arbeitsgedächtnisbatterie für Kinder von 5–12 Jahren* [Working Memory Testing Battery for Children 5–12]. Göttingen: Hogrefe.

Injoque-Ricle, I., Calero, A. D., Alloway, T. P., & Burin, D. I. (2011). Assessing working memory in Spanish-speaking children: Automated Working Memory Assessment battery adaptation. *Learning and Individual Differences*, *21*, 78–84. doi:10.1016/j.lindif.2010.09.012

Ivanova, M. V., & Hallowell, B. (2014). A new modified listening span task to enhance validity of working memory assessment for people with and without aphasia. *Journal of Communication Disorders*, *52*, 78–98. doi:10.1016/j.jcomdis.2014.06.001

Janse, E., & Jesse, A. (2014). Working memory affects older adults' use of context in spoken-word recognition. *The Quarterly Journal of Experimental Psychology*, *67*, 1842–1862. doi:10.1080/17470218.2013.879391

Janusik, L. A. (2007). Building listening theory: The validation of the Conversational Listening Span. *Communication Studies*, *58*, 139–156. doi:10.1080/10510970701341089.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, *44*, 137–166. doi:10.1017/S0261444810000509

LaPointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 1118–1133. http://dx.doi.org10.1037/0278-7393.16.6.1118

Lecerf, T., & Roulin, J.-L. (2006). Distinction between visuo-spatial short-term-memory and working memory span tasks. *Swiss Journal of Psychology*, *65*, 37–55. doi:10.1024/1421-0185.65.1.37

Mackey, A., & Sachs, R. (2012). Older learners in SLA research: A first look at working memory, feedback, and L2 development. *Language Learning*, *62*, 704–740. doi:10.1111/j.1467-9922.2011.00649.x

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, *34*, 379–413. doi:10.1017/S0272263112000125

Morris, N., & Sarll, P. (2001). Drinking glucose improves listening span in students who miss breakfast. *Educational Research*, *43*, 201–207. doi:10.1080/00131880010021311

Petersen, E. B., Wostmann, M., Obleser, J., Stenfelt, S., & Lunner, T. (2015). Hearing loss impacts neural alpha oscillations under adverse listening conditions. *Frontiers in Psychology*, *6*(Art. 177), 1–11. doi:10.3389/fpsyg.2015.00177
Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, *97*, 593–608. doi:10.1121/1.412282
Schulze, K., & Koelsch, S. (2012). Working memory for speech and music. *Annals of the New York Academy of Sciences*, *1252*, 229–236. doi:10.1111/j.1749-6632.2012.06447.x
Siegel, L. S. (1994). Working-memory and reading: A life-span perspective. *International Journal of Behavioral Development*, *17*, 109–123.
Wardlow, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. *Psychonomic Bulletin & Review*, *20*, 766–772. doi:10.3758/s13423-013-0396-1
Wechsler, D. (2014). *Wechsler Intelligence Scale for Children–V*. San Antonio, TX: Pearson.

Profile 36

Listening Styles Profile-Revised (LSP-R)

(Bodie, Worthington, & Gearhart, 2013; Watson, Barker, & Weaver, 1995)

Profiled by: Graham D. Bodie, PhD[1] and Debra L. Worthington, PhD[2]

[1] *Meek School of Journalism and New Media, The University of Mississippi*
[2] *Auburn University*

Construct

Listening styles were originally defined as "attitudes, beliefs, and predispositions about the how, where, when, who, and what of the information reception and encoding process" (Watson *et al*., 1995, p. 2). More recently, listening styles have been conceptualized as representing various goals that listeners can attempt to achieve in a conversational context (Bodie *et al*., 2013).

Instrument Type

Self-Report

Description

The original Listening Styles Profile (LSP-16; Watson *et al*., 1995) was developed to capture individual variability in the tendency to listen in habitual ways. The four orientations measured by the original 16-item scale (people, content, action, and time) reflected preferences of listening with respect to how, where, when, with whom, and what types of information people reported enjoying most. In this view, listening styles are habitual reactions that individual listeners oriented toward, especially in novel situations (Imhof, 2004).

Following the discovery of psychometric issues with the LSP-16 (Bodie & Worthington, 2010), Bodie *et al*. (2013) developed a revised, 24-item measure (LSP-R) that captures four "goals that listeners have when engaged in situations that call them to be a particular

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

kind of listener" (p. 86). Relational listening (RL) is a concern with and awareness of others' feelings and emotions. Analytical listening (AL) reflects an intentional focus on the full message of a speaker prior to forming a judgment. Task-oriented listening (TOL) refers to a concern with the amount of time spent listening as well as a desire to interact with focused speakers. Critical listening (CL) is a tendency to evaluate and critically assess messages for accuracy and consistency. All items are assessed along 7-point Likert scaling.

More recently, a shortened version of the LSP-R—the LSP-R8—has been introduced (Rinke & Moy, 2015).

Administration

The LSP-R is a self-administered questionnaire that takes between 5 and 10 minutes to complete. The items can be printed on paper or administered through an online survey system.

Scoring

There are six items for each of the four listening goals. Responses to each of the items within a single factor are averaged, producing four scores per participant (1–7). No known normative data exist to suggest standard cutoff values.

Development

To classify general manners in which people prefer to listen, Watson *et al.* (1995) proposed the construct of listening style and suggested four listening responses that individuals naturally orient toward. The original Listening Styles Profile (LSP-16) has consistently produced inadequate reliability estimates, in the range of .50 to .60, and has failed to factor appropriately, motivating the creation of a more stable instrument (Bodie & Worthington, 2010; Bodie *et al.*, 2013).

The LSP-R includes 24 items that tap four factors: relational listening (RL), which is concern and awareness of others' feelings and emotions; analytical listening (AL), which is attention to the full message of a speaker before coming to judgment; task-oriented listening (TOL), or disdain for listening to speakers who are lengthy in getting their point across; and critical listening (CL), or focus on the accuracy and consistency of a speaker's message. In addition to relabeling the factors, the LSP-R also shifts the general conceptualization of listening styles from habitual reactions that remain relatively constant across various listening situations to goals that are triggered by both individual predispositions and elements of the listening situation.

Rinke and Moy (2015) introduced a short form of the LSP-R. The shortened scale was formed from the two highest loading items for each of the LSP-R dimensions. Their goal in developing and testing the LSP-R8 was to generate a scale that would be more suitable for use in general population surveys. As they note, the reduction from 24 to 8 items shortened completion time of the scale by 60%.

Reliability

Studies have reported estimates of reliability for the LSP-R. Internal consistency, as estimated by Cronbach's alpha, is consistently above .80 for all subscales (see Table P36.1). In their two-wave panel, Rinke and Moy (2015) reported lower internal consistency

Table P36.1 Standardized Cronbach's alpha values and standardized regression weights for published uses of the Revised Listening Styles Profile (LSP-R)

	Bodie *et al*. (2013)			Gearhart, Denham, and Bodie (2014)		Keaton, Keteyian, and Bodie (2014)	
Item	**α T1**	**α T2**	**λ (T1/T2)**	**α**	**λ**	**α**	**λ**
RL	.82	.86		.88		.81	
1			.67/.66		.69		.68
2			.69/.67		.80		.57
3			.75/.83		.84		.68
4			.51/.67		.57		.62
5			.64/.74		.68		.66
6			.68/.70		.73		.68
AL	.91	.91		.93		.86	
7			.81/.82		.84		.78
8			.86/.66		.78		.70
9			.83/.82		.77		.66
10			.62/68		.78		.69
11			.88/.88		.84		.76
12			.77/.88		.79		.68
TOL	.88	.89		.88		.79	
13			.59/.55		.75		.71
14			.74/.78		.70		.58
15			.71/.76		.74		.60
16			.80/.83		.75		.63
17			.71/.81		.73		.72
18			.67/.56		.74		.58
CL	.86	.85		.91		.82	
19			.72/.69		.74		.67
20			.79/.75		.87		.73
21			.74/.74		.84		.73
22			.76/80		.86		.67
23			.66/.72		.74		.72
24			.74/.79		.69		.49

estimates on some dimensions between αt1 and αt2: analytical listening = .90, .88; critical listening = .84, .82; relational listening = .79, .78; and task-oriented listening = .71, .75. Bodie *et al.* (2013) also reported evidence of test–retest reliability, $r > .71$, using a 14–45-day interval between scale administrations. Test–retest reliability estimates (at a 4-week interval) of the LSP-R8 were lower: AL = .59; CL = .76; RL = .67; TOL = .65 (Rinke & Moy, 2015).

Validity

Several studies have been conducted that build a validity portfolio for the LSP-R. As seen in Table P36.1, lambda estimates obtained from confirmatory factor analyses provide evidence that items are valid indicators of the appropriate factor. Model fit statistics reported in these articles also are all within acceptable ranges (CFI > .90, SRMR <. 08, RMSEA < .08), providing evidence of construct validity.

Bodie *et al.* (2013) reported evidence of convergent and discriminant validity by showing the LSP-R factors correlated appropriately with theoretically relevant variables, among them need for cognition, emotional contagion, need to evaluate, empathy, extraversion, and normative information processing. Bivariate relations between the revised listening styles and these other self-report scales were in predicted directions, and the association between the LSP-R and other listening constructions refuted suggestions of construct proliferation (see Chapter 5). In addition, the four goals measured by the LSP-R correlated as expected with the LSP-16 orientations (see Table P36.2).

Sample Studies

A number of studies into individual listening style have been conducted. Most researchers conducting these early studies focused on examining the relation between listening styles (as measured by the LSP-16) and personality-related characteristics, among

Table P36.2 Bivariate correlations between LSP-16 and LSP-R reported in Bodie *et al.* (2010)

	1	2	3	4	5	6	7
1. People	–						
2. Action	.04	–					
3. Content	.24**	.29**	–				
4. Time	.00	.32**	.09	–			
5. Relational	.71**	−.05	.19**	−.04	–		
6. Analytic	.32**	−.01	.30**	−.03	.39**	–	
7. Critical	.11*	.57**	.41**	.06	−.01	.13**	–
8. Task-oriented	.05	.65**	.09	.35**	−.13**	−.04	.29**

Note: * $p < .05$.
** $p < .01$ level; $N = 409$.
Source: Bodie and Worthington (2010). Reproduced with permission of Taylor & Francis.

them: communication anxiety, conversational sensitivity, empathic response style, extroversion, psychoticism, neuroticism, need for cognition, receiver apprehension, temperament, and type-A personality (Chesebro, 1999; Salisbury & Chen, 2007; Sargent, Fitch-Hauser, & Weaver, 1997; Villaume & Bodie, 2007; Weaver, 1998; Weaver & Kirtley, 1995; Weaver, Watson, & Barker, 1996).

A few studies, however, have examined individual listening style in specified situations. For example, in her study of listening and juror decision making, Worthington (2001) reported that highly people-oriented mock jurors found plaintiffs less at fault, whereas time-oriented listeners awarded higher damages. Imhof (2004) investigated the stability of individual listening profiles across four listening contexts (studying, family, friends, and work) with differing goals. Her study suggested that individual listening, as measured by the LSP-16, varied with the context. This study suggests that listening styles are more akin to goals rather than habitual responses, a finding recently confirmed with the LSP-R (see below).

Although these studies contributed to our understanding of listening style, their findings must be viewed with caution as the LSP-16 has consistently exhibited problems with its psychometric properties. It was these problems that led to a reexamination and revision of the scale (Bodie & Worthington, 2010; Bodie *et al.*, 2013). Although some overlap between the two scales has been reported, they are distinctly different—the LSP-R offers an improved framework for conceptualizing listening preferences (see Bodie *et al.*, 2013). Because of its newness, the LSP-R has not been used as extensively, as initial studies have primarily focused on building a validity portfolio for the measure.

An important piece of the validity portfolio for the LSP-R are findings reported by Gearhart, Denham, and Bodie (2014). The authors, reflecting Imhof's (2004) findings that used the LSP-16, provided evidence that the LSP-R orientations are best conceptualized as situationally variable goals rather than habitualized orientations. In their study, US undergraduate students were asked to respond to the LSP-R and then recall a conversation in which they were primarily enacting a listening role. After describing their interaction, participants answered two sets of items. The first set included 12 items that reflected each of the LSP-R factors (three items for each orientation). For instance, participants were asked how important it was (from 1 = *not at all* to 7 = *very important*) in the recalled conversation to "understand how others were feeling" (RL), "remain nonjudgmental" (AL), "help others to get to the point quickly" (TOL), and "find mistakes" (CL). The second set included 43 items that assessed various characteristics of the recalled conversation. Not only did participants "switch" their "primary listening style" after recalling a particular conversation, but a discriminant function analysis showed that the measures of conversational characteristics more accurately classified participants' choice of situational listening style than did their general LSP-R scores. Their study also questioned whether individuals can be categorized as holding a primary style at all (something that was questionable from the beginning of the LSP-16). Approximately 50% of participants in the Gearhart *et al.* study held multiple, primary LSP-R listening styles, and the remaining participants split between holding one primary style or no primary style. Specifically, only 105 out of 382 participants (27.5%) could be said to have a primary "listening style": RL ($n = 33$; 8.6%), AL ($n = 20$; 5.2%), TOL ($n = 33$; 8.6%), and CL ($n = 19$; 5%). A majority ($n = 184$; 48.2%) were classified as reporting multiple primary LSP-R listening styles, and 93 (24.3%) reported no primary style.

Keaton, Keteyian, and Bodie (2014) found further validity evidence in their study of the relationship between the LSP-R and communication preferences (as measured by the Communication Components inventory). For example, bivariate correlations revealed that relational listening was associated with interpersonal (.61) and linguistic communicator component preferences (.48), analytical listening was positively related to auditory (.41) and logical components (.31), critical listening had significant correlations with logical (.42) and visual (.42) components, and task-oriented listening was related to logical (.35) and linguistic (.24) components.

Critique

The LSP-R offers two significant contributions to listening research. First, as mentioned, it provides an improved framework for investigating individual listening preferences. At the same time, initial research with the measure reaffirms the importance of contextual and situational cues on preference choice. Bodie *et al.* (2013) claim that the LSP-R taps into "various goals that listeners have when engaged in situations that call them to be a particular kind of listener" (p. 86). They question the long-held belief by many users of the LSP-16 that listening style is trait-like and habitual—and, subsequently, relatively invariant.

Despite its relative newness, there is a strong validity and reliability portfolio for the measure. However, research in this area can and should be expanded; additional studies examining the nature and extent of an individual's willingness to shift preferences are especially needed. Finally, research into potential cultural differences is also needed. Social expectations of a listening situation may lead individuals from differing cultures and countries to prefer different styles of listening. Understanding these differences may help researchers better identify the trait and state characteristics of listening style preferences.

References

Bodie, G. D., & Worthington, D. L. (2010). Revisiting the Listening Styles Profile (LSP-16): A confirmatory factor analytic approach to scale validation and reliability estimation. *International Journal of Listening*, *24*, 69–88. doi:10.1080/10904011003744516

Bodie, G. D., Worthington, D. L., & Gearhart, C. C. (2013). The Revised Listening Styles Profile (LSP-R): Development and validation. *Communication Quarterly*, *61*, 72–90. doi:10.1080/01463373.2012.720343

Chesebro, J. L. (1999). The relationship between listening styles and conversational sensitivity. *Communication Research Reports*, *16*, 233–238. doi:10.1080/08824099909388722

Gearhart, C. C., Denham, J. D., & Bodie, G. D. (2014). Listening as a goal directed activity. *Western Journal of Communication*, *78*, 668–684. doi:10.1080/10570314.2014.910888

Imhof, M. (2004). Who are we as we listen? Individual listening profiles in varying contexts. *International Journal of Listening*, *18*, 36–46. doi:0.1080/10904018.2004.10499061

Keaton, S. A., Keteyian, R. V., & Bodie, G. D. (2014). Bivariate and multivariate associations between trait listening goals and trait communicator preferences. *International Journal of Listening*, *28*, 82–97. doi:10.1080/10904018.2014.880868

Rinke, E., & Moy, P. (2015, Sept. 14). Proposal for the 2016 pilot study: Individual listening styles: Short Form of Listening Styles Profile-Revised (LSP-R8). American National Election

Studies forum. Retrieved from http://forum.electionstudies.org/proposal-for-the-2016-pilot-study-individual-listening-styles-short-form-of-listening-styles-profile-revised-lsp-r8/

Salibury, J. R., & Chen, G.-M. (2007). An examination of the relationship between conversation sensitivity and listening styles. *Intercultural Communication Studies, 16*, 251–262.

Sargent, S. L., Fitch-Hauser, M., & Weaver, J. B., III. (1997). A listening styles profile of the type-A personality. *International Journal of Listening, 11*, 1–14.

Villaume, W. A., & Bodie, G. D. (2007). Discovering the listener within us: The impact of trait-like personality variables and communicator styles on preferences for listening style. *International Journal of Listening, 21*, 102–123.

Watson, K. W., Barker, L. L., & Weaver, J. B., III. (1995). The listening styles profile (LSP-16): Development and validation of an instrument to assess four listening styles. *International Journal of Listening, 9*, 1–13. doi:10.1080/10904018.1995.10499138

Weaver, J. B., III. (1998). Personality and self-perceptions about communication. In J. C. McCroskey, J. A. Daly, M. Martin, & M. J. Beatty (Eds.), *Communication and personality* (pp. 95–117). Cresskill, NJ: Hampton.

Weaver, J. B., III, & Kirtley, M. D. (1995). Listening styles and empathy. *Southern Journal of Speech Communication*, *60*, 131–140. doi:10.1080/10417949509372970

Weaver, J. B., III, Watson, K. W., & Barker, L. L. (1996). Individual differences in listening style: Do you hear what I hear? *Personality and Individual Differences*, *20*, 381–387.

Worthington, D. L. (2001). Exploring juror's listening processes: The effect of listening style preference on juror decision making. *International Journal of Listening, 15*, 20–35. doi:10.1080/10904018.2001.10499043

Further Reading

Worthington, D. L. (2003). Exploring the relationship between listening style preference and personality traits: Temperament. *International Journal of Listening, 17*, 68–87. doi:10.1080/10904018.2003.10499056

Worthington, D. L. (2005). Exploring the relationship between listening style preference and personality traits: Verbal aggressiveness. *International Journal of Listening. 19*, 3–11. doi:10.1080/10904018.2005.10499069

Worthington, D. L. (2008). Exploring the relationship between listening style preference and need for cognition. *International Journal of Listening, 22*, 46–58. doi:10.1080/10904010701802154

Scale

The Listening Styles Profile Revised (LSP-R) (Bodie, Worthington, & Gearhart, 2013; Watson, Barker, & Weaver, 1995)

Below are several items that people use to describe themselves as a listener. We would like you to assess how each statement applies to you by marking your level of agreement/disagreement with each item. The stronger you disagree with a statement, the lower the number you will circle. The stronger you agree with a statement, the higher the number you will circle. Please do not think of any specific listening situation but of your general ways of listening—how you typically listen in most situations.

1 = Strongly Disagree
2 = Disagree
3 = Somewhat Disagree
4 = Unsure
5 = Somewhat Agree
6 = Agree
7 = Strongly Agree

<u>RL = Relational Listening</u>

1) When listening to others, I am mainly concerned with how they are feeling.
2) I listen to understand the emotions and mood of the speaker.
3) When listening to others, it is important to understand the feelings of the speaker.
4) I listen primarily to build and maintain relationships with others.
5) I enjoy listening to others because it allows me to connect with them.
6) When listening to others, I focus on understanding the feelings behind words.

<u>AL = Analytical Listening</u>

7) I wait until all the facts are presented before forming judgments and opinions.
8) I fully listen to what a person has to say before forming any opinions.
9) I tend to withhold judgment about another's ideas until I have heard everything they have to say.
10) When listening to others, I attempt to withhold making an opinion until I've heard their entire message.
11) When listening to others, I consider all sides of the issue before responding.
12) To be fair to others, I fully listen to what they have to say before making judgments.

<u>TOL = Task-Oriented Listening</u>

13) I am impatient with people who ramble on during conversations.
14) I find it difficult to listen to people who take too long to get their ideas across.
15) I get frustrated when people get off topic during a conversation.
16) When listening to others, I become impatient when they appear to be wasting time.
17) I prefer speakers who quickly get to the point.
18) When listening to others, I appreciate speakers who give brief, to-the-point presentations.

<u>CL = Critical Listening</u>

19) I often catch errors in other speakers' logic.
20) I tend to naturally notice errors in what other speakers' say.
21) When listening to others, I focus on any inconsistencies and/or errors in what's being said.
22) I have a talent for catching inconsistencies in what a speaker says.
23) When listening to others, I notice contradictions in what they say.
24) Good listeners catch discrepancies in what people say.

Note: Labels should be removed and items randomized prior to administration. The first two items under each subscale comprise the LSP-8.

Profile 37

Medical Communication Competence Scale (MCCS)

(Cegala, Coleman, & Turner, 1998)

Profiled by: Danielle Blanch-Hartigan, PhD, MPH

Bentley University

Construct

The Medical Communication Competence Scale (MCCS) was developed to measure patients' and physicians' perceptions of self- and other-communication competence in clinical interactions.

Instrument Type

Self- and Other-Report

Description

The MCCS measures patients' and physicians' perceptions of self- and other-communication competence. The instrument is not specific to any disease, provider type, or care setting. It can be given to either the physician or the patient for individual assessment, or it can be provided to both the patient and the physician for direct comparison.

There are two versions of this instrument—a physician version and a patient version. Although the number of items has varied with administration, the current physician version contains 36 items, and the current patient version contains 39 items. In the physician version, 23 items assess physicians' own communication competence (e.g., "I did a good job of …"), and 13 items ask them to assess perceptions of the patient's competence (e.g., "The patient did a good job of …"). In the patient version, 16 items assess patients' own communication competence (e.g., "I did a good job of …"), and 23 items ask them to assess perceptions of the physician's competence (e.g., "The doctor did a good job of …"). Items

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

on each version were written to assess four distinct competencies: information giving, information verifying, information seeking, and socioemotional communication. Responses to all items are recorded using 7-point Likert scaling.

Administration

The MCCS is a self-administered survey that is designed to be completed after a physician visit and refers specifically to that visit. The survey can be administered immediately following the visit or after a predetermined delay. Surveys can be taken in person, using paper or computer, or online.

Scoring

The MCCS was intended to produce up to four composite scores: two for physician competence (self- and patient-reported) and two for patient competence (self- and physician-reported). For example, for the physician version, the 23 items referencing physicians' assessments of their own competence are averaged to form a score that reflects physician self-reported competence; the 16 items referencing physicians' assessment of patients' competence are averaged to form a score that reflects physicians' perceptions of patient competence. A similar procedure is followed for the patient version of the instrument. Each of these scores ranges from 1 (*strongly disagree*) to 7 (*strongly agree*).

Development

The MCCS was developed to assess two broad medical communication competencies, information exchange and socioemotional communication, from the perspective of the physician and patient immediately following a medical consultation (Cegala, Coleman, & Turner, 1998; Cegala, McGee, & McNeilis, 1996). *Information exchange* was conceptually defined as seeking, giving, and verifying information. *Socioemotional communication* was conceptually defined as abilities to convey warmth, empathy, understanding, friendliness, and trust. These two superordinate competencies are thought to represent the "major goals" of a medical consultation (Cegala *et al.*, 1998, p. 263).

In the initial stage of item development, 52 patients and 65 physicians sampled from "different locations within the state of Ohio" (Cegala *et al.*, 1998, p. 266) were asked immediately after their consultations to rate self- and other-competence and provide descriptions of the behaviors that prompted their ratings. Descriptive data were content analyzed based on the two broad competencies described here. Descriptive data also conformed to the three subordinate information exchange competencies (information seeking, giving, and verifying) and the three subordinate socioemotional communication competencies (creation of a friendly and trusting atmosphere; demonstration of care, concern, and interest; and display of affective support).

The 56 most frequently mentioned competence-related behaviors were used to craft scale items that were administered to 6 physicians and 13 patients who rated them on a scale from 1 (*unimportant*) to 5 (*important*). Items identified as *important* by 80% of the physicians and patients were included in the pilot MCCS. These 37 physician items and 41 patient items were then administered to 117 total respondents, 65 of whom were physicians (mean age = 45, range = 28–83, 75% male, 89% White); the remaining 52 were patients (mean age = 49, range = 19–89, 35% male, 85% White). Most of the patient participants had an established relationship with the physician ($n = 36$, mean = 8.4 years). The pilot data were subjected to cluster analyses using Ward's method. Based on these analyses, one item ("Answering the patient's [my] questions thoroughly") was dropped from the final MCCS. Over time, some items have been removed and others altered to create a 36-item physician version and a 39-item patient version.

A French version that includes 40 items for patients to assess their own and physician competence was developed using forward-backward translation by certified translators (Trudel, Leduc, & Dumont, 2014).

Reliability

Internal consistency estimates reported for the instrument have been high. The reliability coefficients reported by Cegala *et al.* (1998) for the doctor scale were .86 for the information giving factor, .75 for information seeking, .78 for information verifying, and .90 for socioemotional communication. For the patient scale, estimates were reported as .79, .76, .85, and .92, respectively.

Avtgis and Polack (2007) reported high Cronbach's alphas for the four dimensions of the MCCS that assess patient perceptions of physician competence (.78–.90). The French version of the patient items demonstrated adequate reliability, with Cronbach's alphas ranging from .65 to .97 (Trudel *et al.*, 2014).

Validity

Face validity was demonstrated by asking both physicians and patients to rate the importance of initially proposed items during the measure's development and selection phase (Cegala *et al.*, 1996). Construct validity was demonstrated through the SPSS-X Proximities and Cluster Program, which computed Euclidean distances among the standardized item responses and then identified clusters using Ward's method. Items fell into the four clusters as hypothesized. Those clusters were: (1) information giving, (2) information verifying, (3) information seeking, and (4) socioemotional communication. A separate report presented data from 90 physicians and 202 patients sampled in previous studies (Cegala, 2007). These data were submitted to common factor analysis with oblimin rotation that provided added support for the four dimensions.

Cegala and colleagues (1998) also presented results of a number of cross-sectional studies that provide evidence of validity. As in previous doctor–patient communication literature, physicians and patients rated their own socioemotional competence higher than the information-giving, seeking, or verifying subscales. Physicians rated patients' information competence (giving, seeking, and verifying) lower than patients' self-rated competence. Additional validity evidence comes from the finding that patients who said they want more information were self- and physician-rated as higher in information-seeking competence.

This instrument has been featured in a number of reviews of instruments for patient–provider communication (Boon & Stewart, 1998; Hudon, Fortin, Haggerty, Lambert, & Poitras, 2011; Légaré, Moher, Elwyn, LeBlanc, & Gravel, 2007; Zill *et al*, 2014). One of these reviews systematically evaluated the quality of articles describing the 13 instruments designed to assess perceptions of patient-centered care on a modified version of the Standards for Reporting of Diagnostic Accuracy (STARD) (Hudon *et al.*, 2011). The MCCS scored a 10 out of 15 possible quality points, falling at the lower end of the range of quality compared to the other instruments. Points were lost for not adequately describing recruitment and participants in the study upon which the reliability and validity data were based and for not describing how outliers and missing data were handled.

The most recent review by Zill and colleagues (2014) went even further to evaluate psychometric properties of a number of patient–provider communication instruments using two approaches, the COSMIN (Consensus-based Standards for the Selection of Health Status Measurement Instruments) checklist and the quality criteria of Terwee and colleagues (Zill *et al.*, 2014). The MCCS was fair in content validity but poor in structural validity and hypothesis testing. Internal consistency of MCCS items was rated as poor in one assessment approach but intermediate in a second. Construct validity received a positive score on this second approach.

Availability

Cegala *et al.* (1998) introduced and presented the MCSS measure in their 1998 article published in *Health Communication*. The scale has been modified since that time. Presented in the Scale section are the 36- and 39-item physician and patient versions. They are included here with permission. Both are free to use with appropriate citation.

Sample Studies

The instrument has been used in studies assessing the impact of physician training programs on communication behavior (Cegala, McClure, Marinelli, & Post, 2000) and comparing physicians' and patients' perceptions of competent communication (Cegala, Gade, Lenzmeier Broz, & McClure, 2004).

Avtgis and Polack (2007), using 16 items to assess patient perceptions of physician competence, reported that patients' perceptions of relationship quality predict their perceptions of physician communication competence.

A Canadian study translated the patient MCCS into French to assess the relationship between perceived communication and health-related quality of life (HRQOL) in 120 women with early-stage breast cancer (Trudel *et al.*, 2014). Results indicated that patients' perception of their own communication competence was more strongly related to HRQOL than their perception of their physicians' communication competence. A subsequent study in France, using the same translation of the patient MCCS, demonstrated that older breast cancer survivors who reported greater information-seeking needs were more likely to report higher competence in information giving, lower competence in information seeking, and lower satisfaction with physicians' information giving (Brédart, Kop, Fiszer, Sigal-Zafrani, & Dolbeault, 2015).

Critique

The MCCS has shown adequate evidence for reliability and validity in a number of publications. The development process has been extremely well described.

As noted in a 1998 review of patient–provider communication measures, a real strength of the MCCS is the ability to directly compare patient and physician perceptions of the interaction (Boon & Stewart, 1998). The only other widely used instrument with validity evidence that allows for a direct comparison between patient and providers is the Patient-Provider Orientation Scale (PPOS), but this scale has more to do with their preferred interaction style than perceived communication skills (Krupat *et al.*, 2000).

The ability to self-assess communication competence has itself been questioned. Many studies have demonstrated that providers often overestimate their ability to communicate with patients. Likewise, patients' perceptions of physician communication quality are often influenced by factors such as patient and provider sex. A review of communication competence rating tools for patient–provider interactions did not include the MCCS, but found that, in general, communication competence instruments vary greatly in usability and psychometric properties (Schirmer *et al.*, 2005).

It should be noted that although the scale is intended as unidimensional within perspective (e.g., one score for physician self-reported competence), the original report and subsequent descriptions of the scale suggest that items cluster into subordinate and superordinate dimensions. Information exchange is said to comprise information seeking, giving, and verifying. Although socioemotional communication was originally conceptualized along three dimensions, it is treated as a single cluster of items. As such, the measurement model for the scale is a second-order latent factor structure, with information exchange and socioemotional communication as first-order latent factors. The information exchange factor is further divided into three dimensions. Items should load on their respective first-order factors appropriately with no cross-loadings. This measurement model has not yet been tested against alternative models to date. If medical competence does indeed depend on these distinct skill sets and this instrument has a validity portfolio that stands increased scrutiny, then training programs can be designed and tailored around reported areas of weakness.

Finally, although the MCCS contains items that reflect listening behavior, listening is not a direct focus of the scale. Instead, the socioemotional factor contains elements closely related to listening. More work should be done to empirically examine the MCCS with more direct measures of listening behavior and attributes.

References

Avtgis, T. A., & Polack, E. P. (2007). Predicting physician communication competence by patient perceived information exchange and health locus of control. *Human Communication, 21*, 136–144. Retrieved from http://www.uab.edu/Communicationstudies/humancommunication/avtgis.pdf

Boon, H., & Stewart, M. (1998). Patient-physician communication assessment instruments: 1986 to 1996 in review. *Patient Education and Counseling, 35*, 161–176. PII:S0738-3991(98)00063-9

Brédart, A., Kop, J.-L., Fiszer, C., Sigal-Zafrani, B., & Dolbeault, S. (2015), Breast cancer survivors' perceived medical communication competence and satisfaction with care at the end of treatment. *Psycho-Oncology, 24.* Advance online publication. doi:10.1002/pon.3836.

Cegala, D. (2007, November). *A validity study of the Medical Communication Competence Scale (MCCS).* Paper presented at the meeting of the National Communication Association convention, Chicago, IL. Retrieved from http://citation.allacademic.com/meta/p_mla_apa_research_citation/1/9/0/6/0/pages190601/p190601-1.php

Cegala, D. J., Coleman, M. T., & Turner, J. W. (1998). The development and partial assessment of the medical communication competence scale. *Health Communication, 10,* 261–288. doi:10.1207/s15327027hc1003_5

Cegala D. J., Gade, C., Lenzmeier Broz S., & McClure, L. (2004). Physicians' and patients' perceptions of patients' communication competence in a primary care medical interview. *Health Communication, 16,* 289–304. doi:10.1207/S15327027HC1603_2

Cegala, D. J., McClure, L., Marinelli, T. M., & Post, D. M. (2000). The effects of communication skills training on patients' participation during medical interviews. *Patient Education and Counseling, 41,* 209–222. PMID:12024545

Cegala, D. J., McGee, D. S., & McNeilis, K. S. (1996). Components of patients' and doctors' perceptions of communication competence during a primary care medical interview. *Health Communication, 8,* 1–27. doi:10.1207/s15327027hc0801_1

Hudon, C., Fortin, M., Haggerty, J. L., Lambert, M., & Poitras, M. E. (2011). Measuring patients' perceptions of patient-centered care: A systematic review of tools for family medicine. *The Annals of Family Medicine, 9,* 155–164. doi:10.1370/afm.1226

Krupat, E., Rosenkranz, S. L., Yeager, C. M., Barnard, K., Putnam, S. M., & Inui, T. S. (2000). The practice orientations of physicians and patients: The effect of doctor–patient congruence on satisfaction. *Patient Education and Counseling, 39,* 49–59. PMID:11013547

Légaré, F., Moher, D., Elwyn, G., LeBlanc, A., & Gravel, K. (2007). Instruments to assess the perception of physicians in the decision-making process of specific clinical encounters: A systematic review. *BMC Medical Informatics and Decision Making, 7*(1), 30.

Schirmer, J. M., Mauksch, L., Lang, F., Marvel, M. K., Zoppi, K., Epstein, R. M., … & Pryzbylski, M. (2005). Assessing communication competence: A review of current tools. *Family Medicine, 37,* 184–192. Retrieved from http://www.stfm.org/fmhub/fm2005/March/Julie184.pdf

Trudel, J. G., Leduc, N., & Dumont, S. (2014). Perceived communication between physicians and breast cancer patients as a predicting factor of patients' health-related quality of life: A longitudinal analysis. *Psycho-Oncology, 23,* 531–538. doi:10.1002/pon.3442

Zill, J. M., Christalle, E., Müller, E., Härter, M., Dirmaier, J., & Scholl, I. (2014). Measurement of physician-patient communication—A systematic review. *PloS One, 9,* e112637. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4273948/pdf/pone.0112637.pdf

Scale

The Medical Communication Competence Scale (Cegala, Coleman, & Turner, 1998; Cegala, 2007)

Source: Cegala *et al*. (1998) and Cegala (2007).

Doctor's Version

Directions

The purpose of this questionnaire is to obtain your views about communication during the interview you just had. There are two parts to the questionnaire. In the first part, you are asked to make judgments of your own communication. In the second part, you are asked to make judgments of the patient's communication. For each item, please circle the most appropriate alternative. If you do not believe an item is relevant to this particular interview, please write NA next to the item.

Response Options

Strongly Disagree	Disagree	Slightly Disagree	Not Sure	Slightly Agree	Agree	Strongly Agree
1	2	3	4	5	6	7

Doctor's Self-Competence Items

I provided good explanations of the following to the patient:

1) The diagnosis of his/her medical problem.
2) The causes of his/her medical problem.
3) The treatment for his/her medical problem.
4) The advantages and disadvantages of treatment options.
5) The purpose of any tests that were needed.
6) How prescribed medication will help his/her problem.
7) How to take prescribed medication.
8) The possible side effects of the medication.
9) The long-term consequences of his/her medical problem.

I did a good job of

10) Reviewing or repeating important information for the patient.
11) Making sure the patient understood my explanations.
12) Making sure the patient understood my directions.
13) Checking my understanding of information the patient provided.
14) Using language the patient could understand.
15) Asking the patient the right questions.
16) Asking questions in a clear, understandable manner.
17) Using open-ended questions.
18) Being warm and friendly.
19) Contributing to a trusting relationship.
20) Showing the patient I cared about him/her.

21) Making the patient feel relaxed or comfortable.
22) Showing compassion.
23) Being open and honest.

Doctor's Other-Competence Items:

The patient did a good job of

24) Providing relevant history associated with his/her medical problem.
25) Explaining symptoms associated with his/her medical problem.
26) Identifying what medications he/she is taking.
27) Answering my questions thoroughly.
28) Answering my questions honestly.
29) Letting me know when he/she didn't understand something.
30) Asking me to explain terms he/she didn't understand.
31) Letting me know when I needed to repeat something.
32) Asking questions about his/her medical problem.
33) Pursuing answers to his/her questions.
34) Asking appropriate questions.
35) Contributing to a trusting relationship.
36) Being open and honest.

Patient's Version

Directions

The purpose of this questionnaire is to obtain your views about communication during the interview you just had. There are two parts to the questionnaire. In the first part, you are asked to make judgments of your own communication. In the second part, you are asked to make judgments of the doctor's communication. For each item, please circle the alternative that best describes how you feel. If you do not believe an item applies to this particular interview, please write NA next to the item.

Response Options

Strongly Disagree	Disagree	Slightly Disagree	Not Sure	Slightly Agree	Agree	Strongly Agree
1	2	3	4	5	6	7

Patient's Self-Competence Items:

I did a good job of

1) Presenting important history that has to do with my medical problem.
2) Describing the symptoms of my medical problem.
3) Explaining my medical problem.
4) Identifying the medications I am taking.
5) Answering the doctor's questions thoroughly.
6) Answering the doctor's questions honestly.
7) Letting the doctor know when I didn't understand something.
8) Letting the doctor know when I needed him/her to repeat something.
9) Making sure I understood the doctor's directions.

10) Repeating important information to make sure I understood correctly.
11) Asking the doctor to explain terms I didn't understand.
12) Asking the doctor all the questions that I had.
13) Getting the answers to my questions.
14) Getting all the information I needed.
15) Contributing to a trusting relationship.
16) Being open and honest.

<u>Patient's Other-Competence Items</u>

The doctor explained the following to my satisfaction:

17) What my medical problem was.
18) The causes of my medical problem.
19) What I could do to get better.
20) The benefits and disadvantages of treatment choices (that is, choices about what I could do to get better).
21) The purpose of any tests that were needed.
22) How prescribed medicine would help my problem.
23) How to take prescribed medicine.
24) The possible side effects from the medicine.
25) The long-term consequences of my medical problem.

The doctor did a good job of

26) Reviewing or repeating important information.
27) Making sure I understood his/her explanations.
28) Making sure I understood his/her directions.
29) Using language I could understand.
30) Checking his/her understanding of what I said.
31) Asking me questions related to my medical problem.
32) Asking me questions in a clear, understandable manner.
33) Asking questions that allowed me to elaborate on details.
34) Being warm and friendly.
35) Contributing to a trusting relationship.
36) Showing he/she cared about me.
37) Making me feel relaxed or comfortable.
38) Showing compassion.
39) Being open and honest.

Note: Labels should be removed and items randomized prior to administration.

Profile 38

Memory for Conversation

Profiled by: Jonathon Frost, MA and Brock Bybee, MA

Louisiana State University and Agricultural & Mechanical College

Construct

Memory for conversation, or conversational memory, refers to how much information (and what type of information) individuals remember after engaging in a conversation. Conversational memory is used as an indicator of "how actively individuals attend to and process information" (Sillars, Weisberg, Burggraf, & Zietlow, 1990, p. 505).

Instrument Type

Cognitive Assessment

Description

Memory for conversation is grounded in research on memory retention, with a specific focus on what people remember from natural conversations. Most generally, measures of conversational memory aim to provide an estimate of long-term memory storage or capacity rather than short-term memory (see Profile 12, Conversational Listening Span [CLS], and Profile 35, Listening Span Tests). The emphasis on reflecting a natural setting has led researchers to develop various methods of testing in an attempt to more accurately demonstrate natural conversational memory. The three most popular methods, free recall, cued recall, and recognition, are covered in this profile.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

In order to administer a measure of conversational memory, participants first have to participate in some form of dyadic interaction, usually short in duration. In some instances, observers are instructed to watch a conversation between two people. Following the interaction, interlocutors and/or observers (participants) are tasked with a disruptive activity (often in the form of a short, irrelevant video) to interrupt their short-term memory. Participants are then evaluated using one of three methods to determine memory retention.

The first method, free recall, involves asking participants to recall the engaged or observed conversation without any assistance. The open-ended method involves asking a simple question about the participants' ability to recall details from a previous conversation. Responses are then coded following a standard rubric. We have included the rubric used by Stafford and Daly (1984) at the end of this profile.

The second method, cued recall, involves presenting the participant with a cue (typically a description of conversation patterns or specific information; see below for example) and asking them to recall relevant information from their conversation about that cue. In many ways, cued recall as a measure of conversational memory is similar to standard measures of listening comprehension that use multiple-choice items (see Watson-Barker Listening Test, Profile 64).

The third method, recognition, involves presenting participants with a specific item and asking them to either (a) indicate whether the item did or did not occur in the conversation, or (b) report how many times the item occurred during the conversation.

Scoring

Scoring methods vary due to the fact that different methods are employed to measure what people remember. Researchers typically record dyadic interactions between participants using audio or video surveillance, or alternatively, present an observer with a pre-recorded conversation. Researchers identify each unique piece of information discussed in the conversation or create a unique coding system based on the transcript of the conversation.

For free recall, information from recorded conversations is coded into smaller units of meaning such as thought units that are then placed into larger categories. For instance, recall attempts were coded by Stafford and colleagues (Stafford, Burggraf, & Sharkey, 1987; Stafford & Daly, 1984) into seven categories the authors labeled recall units: *reproductions*, which preserved the original idea of a message; *redundant reproductions*, in which participants recalled an idea they had already identified; *themes*, or summary statements that captured the overall essence of a discussion but did not tie to any specific conversation piece; *elaborations*, which are ideas not specifically discussed during the conversation but are plausible in the context and/or logical extensions of the conversation; *descriptions*, which commented on the roles of the participants or the overall experience of the conversation; *evaluations*, which are descriptions that reflect a positive or negative interpretation or bias; and *errors*, or information that was inaccurate or inconsistent with the conversation.

For cued recall, researchers develop a unique set of questions based on the transcript of the conversation. Cues can be *sequential* ("What happened after X?") or *semantic* ("What did they say about Y?"), and questions can ask about both verbal and nonverbal behavior. Responses to questions are then measured against the conversation transcript. If an item was left blank by the participant, or if the response does not accurately reflect the transcript, the item is recorded as lack of recall (Benoit & Benoit, 1988a).

Recognition can be measured using three different question types (Benoit & Benoit, 1990; see example at end of profile): *actual recognition questions* (users are asked if they recognize information that was actually in the conversation), *spurious recognition* (users are presented with information that was not in the conversation), and *frequency estimates* (participants are asked how many times either nonverbal or verbal communication patterns occurred during the conversation). Alternatively, participants can be given a list of information and asked to select all pieces of information discussed during a task or conversation (see Knutsen & Le Bigot, 2015).

Development

The study of conversational memory developed out of memory research from the 1970s and 1980s. Conversational memory studies have developed along two paths: studies that seek to determine the quantity of information people can remember from conversation (i.e., how much information people remember), and studies that seek to determine the quality of information retained from conversation (what people remember), although the two goals are often intertwined (Stafford, Waldron, & Infield, 1989).

Most research in conversational memory can be traced back to Stafford and Daly's (1984) study, which used a free recall method. Their study ignited a long-standing debate over the optimal way to study conversational memory. Early studies of conversational memory employed recognition as the primary form of elicitation (Hjelmquist, 1984; Keenan, MacWhinney, & Mayhew, 1977; MacWhinney, Keenan, & Reinke, 1982). Stafford and Daly (1984), however, criticized the validity of recognition as an accurate representation of natural conversational memory and instead employed free recall as an elicitation method. Benoit and Benoit (1988a, 1988b), as a rebuttal, argued that free recall was not a more accurate representation and offered cued recall as a more effective approach to studying conversational memory.

The debate is far from settled, as there remains some dispute over which method more accurately represents natural conversational memory, and all three are still employed for various purposes.[1] Recent research focuses more on factors that influence how we remember rather than simply looking at the amount we remember. Researchers are advised to carefully consider the construct of interest in any given study and to choose the most appropriate operational method for their specific purpose.

1 For a recent study using free recall, see Lawson and London (2015); for cued recall, see Hedrick, Haden, and Ornstein (2009); for recognition, see Knutsen and Le Bigot (2015).

Reliability

All three methods require researchers to code conversational data into units, necessitating the employment of multiple coders to ensure adequate intercoder reliability. Cued recall generally produces higher levels of interrater agreement than free recall and recognition strategies. Benoit and Benoit (1988a) reported intercoder reliability at .84 for free recall and .99 for cued recall, and Benoit and Benoit (1990) reported intercoder reliability ranging from .79 to .92 for recognition measures. Interrater agreement is also likely a function of conversational characteristics such as duration, complexity, and topic familiarity.

Validity

As noted here, there is some disagreement over which method (free recall, cued recall, or recognition) most accurately reflects conversational memory in a natural setting, a question of construct validity (see Benoit & Benoit, 1988a, 1988b, 1990; Benoit, Benoit, & Wilkie, 1996; Stafford & Daly, 1984). Early measures of conversational memory were criticized for not accurately reflecting a natural conversational setting due to (a) the focus on recognition as the primary measure of conversational memory and (b) having participants recall information about an observed conversation rather than actively participating in the conversation themselves (Stafford & Daly, 1984). Recognition measures also have a distinct flaw in the form of spurious recognition; because people sometimes make spurious recognitions, it is impossible to judge whether a documented recognition is an accurate remembrance or simply chance.

Other scholars, however, have argued that cued recall (and to a lesser extent recognition) is a better measure of conversational memory than free recall because individuals in natural settings will typically remember information from previous conversations based on cues present in their current conversations (Benoit & Benoit, 1988b). It is, however, important to note that cued recall measures do not assess 100% of the information discussed in a conversation, as the cues used must themselves contain conversational information, making it impossible to test for 100% retention.

Availability

Stafford and Daly's (1984) coding system for free recall is reproduced at the end of this document with permission. Sample questions for both cued recall and recognition also are provided, with permission, from Benoit and Benoit (1988a) and Benoit and Benoit (1990), respectively. As questions for cued recall and recognition are catered to each unique conversation, researchers should use these as guides when developing measures appropriate to their study's domain.

Sample Studies

Conversational memory has been studied from a variety of perspectives. A common theme is the goal of quantifying how much people can remember from conversations. Free recall has been shown to elicit as little as 10% memory retention (Stafford & Daly, 1984; compare to Benoit [1988a], who reported 16%). Cued recall and recognition have yielded substantially higher rates. Benoit and Benoit (1988a) reported a 79% participant success rate using cued recall, and Benoit and Benoit (1988b) reported that participants were able to accurately recognize 86% of the information presented to them.

In addition, studies have investigated various factors that might influence conversational memory, including interactional content (Keenan *et al.*, 1977), relational satisfaction and emotional content (Sillars *et al.*, 1990), communication anxiety (Stafford & Daly, 1984), interpersonal competence (Miller, 1996), communication competence (Miller & deWinstanley, 2002), partner familiarity and resistance (Samp & Humphreys, 2007), reuse of information (Knutsen & Le Bigot, 2015), and more.

For example, Sillars *et al.* (1990) found that couples were more likely to remember conversations more negatively if they were less satisfied in their relationship; conversely, couples are more likely to remember conversations more positively if they had higher levels of relational satisfaction.

Keenan *et al.* (1977) found that participants were more likely to remember information from a conversation if they were required to in some way interact with that information afterward, and Knutsen and Le Bigot (2015) likewise found that participants remembered information that was reused. Higher levels of interpersonal competence and communication competence also increase the amount of information people will remember (Miller, 1996; Miller & deWinstanley, 2002).

Baroni and Nicolini (1995) were not able to find a significant difference in general memory patterns for men and women; however, they did find that men remembered more information about their partner's speaking utterances if their partner was also a male.

Samp and Humphreys (2007) found that people are more likely to accurately remember information from conversations with friends than with strangers. However, when encountered with partner resistance, conversations with friends were remembered less accurately in terms of positive information than conversations with strangers.

Critique

Possibly the most crucial factor to consider when attempting to measure conversational memory is which method to use. Memory retention rates will vary as a function of method, and there is very little consensus on which measure more accurately captures the construct. Further complicating this issue is the fact that researchers cannot easily make generalizations across methods. Free recall is measured against everything that happened in a conversation (and as such is a little less reliable and a lot less effective at eliciting memory), but cued recall and recognition are only measured against the information about which researchers ask. Recognition measures have the additional complication of spurious recognition. As such, it is impossible to accurately gauge how much a participant is honestly recognizing and how much is chance.

Because of their flaws, it is difficult to determine which method is superior, and to say if any of the methods are accurate measures of natural conversational memory. People are rarely (if ever) demanded to recall 100% of a conversation. Conversationalists also do not always have cues to spark memories, and they must rely on free recall. Future work in this area should seek to create a clear distinction between information that can be readily recalled without external stimulus and information stored in memory that requires external stimulus to access.

Irrespective of flaws, measures of conversational memory assist in the discovery of conversational phenomena. Listening scholars are encouraged to measure various facets of how we remember certain types of information from certain classes of conversations with particular people and with various grades of specificity.

References

Baroni, M. R., & Nicolini, C. (1995). Natural conversations in males and females: Conversational styles, content recall and quality of interaction. *Pragmatics*, *5*, 407–426. doi:10.1075/prag.5.4.01bar

Benoit, P. J., & Benoit, W. L. (1988a). Conversational memory employing cued and free recall. *Central States Speech Journal*, *39*, 18–27. doi:10.1080/10510978809363231

Benoit, W. L., & Benoit, P. J. (1988b). Factors influencing the accuracy of verbal reports of conversational behavior. *Central States Speech Journal*, *39*, 218–232. doi:10.1080/10510978809363251

Benoit, W. L., & Benoit, P. J. (1990). Memory for conversational behavior. *Southern Communication Journal*, *56*, 24–34. doi:10.1080/10417949009372813

Benoit, W. L., Benoit, P. J., & Wilkie, J. (1996). Participants' and observers' memory for conversational behavior. *Southern Journal of Communication*, *61*, 139–154.

Hedrick, A. M., Haden, C. A., & Ornstein, P. A. (2009). Elaborative talk during and after an event: Conversational style influences children's memory reports. *Journal of Cognition & Development*, *10*, 188–209. doi:10.1080/15248370903155841

Hjelmquist, E. (1984). Memory for conversations. *Discourse Processes*, *7*, 321–336. doi:10.1080/01638538409544595

Keenan, J. M., MacWhinney, B., & Mayhew, D. (1977). Pragmatics in memory: A study of natural conversion. *Journal of Verbal Learning & Verbal Behavior*, *16*, 549–560. doi:10.1016/S0022-5371(77)80018-2

Knutsen, D., & Le Bigot, L. (2015). The influence of reference acceptance and reuse on conversational memory traces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 574–585. doi:10.1037/xlm0000036

Lawson, M., & London, K. (2015). Tell me everything you discussed: Children's memory for dyadic conversations after a 1-week or a 3-week delay. *Behavioral Sciences & The Law*, *33*, 429–445. doi:10.1002/bsl.2184

MacWhinney, B., Keenan, J. M., & Reinke, P. (1982). The role of arousal in memory for conversation. *Memory and Cognition*, *10*, 308–317. doi:10.3758/bf03202422

Miller, J. B. (1996). Memory for conversation. *Memory*, *4*, 615–631. doi:10.1080/741940999

Miller, J. B., & deWinstanley, P. A. (2002). The role of interpersonal competence in memory for conversation. *Personality & Social Psychology Bulletin*, *28*, 78–89. doi:10.1177/0146167202281007

Samp, J. A., & Humphreys, L. R. (2007). "I said what?" Partner familiarity, resistance, and the accuracy of conversational recall. *Communication Monographs*, *74*, 561–581. doi:10.1080/03637750701716610

Sillars, A. L., Weisberg, J., Burggraf, C. S., & Zietlow, P. H. (1990). Communication and understanding revisited: Married couples' understanding and recall of conversations. *Communication Research*, *17*, 500–522. doi:10.1177/009365090017004006

Stafford, L., Burggraf, C. S., & Sharkey, W. F. (1987). Conversational memory: The effects of time, recall, mode, and memory expectancies on remembrances of natural conversations. *Human Communication Research*, *14*, 203–229. doi:10.1111/j.1468-2958.1987.tb00127.x
Stafford, L., & Daly, J. A. (1984). Conversational memory. *Human Communication Research*, *10*, 379–402. doi:10.1111/j.1468-2958.1984.tb00024.x
Stafford, L., Waldron, V. R., & Infield, L. L. (1989). Actor-observer differences in conversational memory. *Human Communication Research*, *15*, 590-–611. doi:10.1111/j.1468-2958.1989.tb00200.x

Scales

Sample Free Recall Coding Scheme (Stafford & Daly, 1984)

Source: Stafford, "Memory for Conversation" MA thesis. Reproduced with permission of Laura Stafford.

Guidelines for abstracting a discourse base

1) Conversations are divided into units.
2) An idea unit (IU) equals the smallest unit of meaning that has informational or affective value.
3) Idea units are numbered sequentially.
4) Include question marks when applicable.
5) Incomplete idea units are labeled fragments (Fr). Fragments are included on the coding sheet in the order they occur but are numbered separately from IU's. Fragments are not deleted because they many have potential informational or affective value to the conversants.
6) Idea units are normally expressed by the conversants. The coder does not fill in incomplete sentences or add implied statements.
7) Shared referents are indicated by placing parentheses around the antecedent.

   For example, in the idea unit "The instructor is crazy," if the instructor had been identified as Dr. Daly, the idea unit would be written "The instructor (Dr. Daly) is crazy." This is done to aid the coder in scoring, since a recall unit using either label (instructor or Dr. Daly) would be scored as correct.
8) An idea unit may be embedded in another idea unit and is indicated using parentheses.

   For example, if in the conversation one interactant states, "I'm going to Nashville to see my boyfriend," the statement would be coded as two sequential idea units; e.g. 17. I'm going to Nashville and 18. (17) to see my boyfriend.
9) Redundancies by the same conversant are all counted as one idea unit. But all terms used to refer to the same idea are indicated.

   For example, the statements "It is a nice day" and "It is a pretty day" would be coded as one idea unit because the gist of the two statements is the same. On the coding sheet the idea unit would be expressed as 21. It is a nice day. (It is a pretty day.)
10) When one conversant is redundant with another conversant, the redundancies are coded as separate idea units.

    For example, if Person A says "It is a nice day" and Person B says "It is a pretty day," the statements would be coded as two separate idea units.

An exception to this occurs when the redundancy is simply a back channel response. Such responses are deleted.

11) Proper names and their referents do not form separate idea units.
12) Back channel responses, such as "uh huhs," are deleted.
13) Responses that make sense in a sentence form in reference to the statement preceding it and could be interpreted as having informational or affective value are coded as idea units.

    For example, if Person A says "I am a minister" and Person B says, "Neat" the statements would be coded as: "I am a minister" and "It is neat that you are a minister."
14) Anytime a response is deleted by a conversant, the coder must check the original tape to see if the response was nonverbal. If so, the nonverbal response must be coded as an idea unit.

    For example, if Person A says "Are you a senior?" and there is no response from Person B, then Person A says "I'm a freshman," the coder must check the video tape. The response would then be coded as an idea unit with the symbol NV. The above exchange would be coded as: Person A "Are you a senior?" Person B (NV) "Yes, I am." Person A "I'm a freshman."
15) On the coding sheet each proposition is numbered, labeled, and placed under the appropriate speaker (A or B).
16) A checkmark is placed in the appropriate column for each proper name mentioned.
17) When in doubt if some information is one or two idea units, the guiding rule is intuitive. The coder must ask himself, "Could a part of this information be remembered without the other part?"
18) A conversation and its recall protocols should be coded in one session.
19) The total number of idea units is tallied.
20) The total number of proper names is tallied.

Categories, Rules, and Examples

Categories of Recall Units

Reproductions (Rep): Recall units that preserve the gist of the original idea units. These may be verbatim or paraphrases. A liberal interpretation or "gist" is used.

1) Redundant Reproductions (RR): The subject repeats a reproduction.
2) Topics (T): A mention of a part of the conversation, but no information about the reference is given.
3) Elaborations (el): Addition of plausible details or normal properties. This occurs when subjects think something was said and wasn't—but could have been. Subjects usually elaborate on what they think they told their partner.
4) Global Representations (gr): Creating an IU unrelated to to any single sentence but that captures the global situation being described. Gr's add some information other than just a mention of a topic.
5) Inferences (I): Represent shared world knowledge not explicitly stated in the text base. A inference can be ascertained from the text.
6) Evaluations: Expression of a positive or negative bias. There are four types of evaluation. One could make judgments about the other person (ep) experimental setting (e ex), conversation (e c), or unclassifiable evaluative comments (e unc).

7) Descriptions: A recall unit intended to give a mental picture of something with no positive or negative bias attached. Four types of descriptions: Person (d p), experimental setting (d ex), conversation (d c), and unclassifiable descriptive comments (d unc).
8) Errors (Er): Incorrect recall of one of more idea units. (This is not to be confused with the addition of plausible information.)
9) Unclassifiables (Unc): Any recall unit that cannot be classified in one of the above categories.

Rules for Scoring Recall Protocols

1) Recall units are any memory of, any implication from, any description, any evaluation, or any elaboration of the conversation, the conversants, or the experience.
2) Each protocol is divided into recall units by placing parentheses around each unit.
3) Information that is not part of the recall is deleted (e.g., "I can't think of anything else to say," or "I don't like talking to a tape recorder.")
4) Each recall unit is identified as belonging to one of the categories described above.
5) The abbreviation for the category is places above the recall unit.
6) The number of the corresponding idea unit is placed above each reproduction.
7) A checkmark is placed by each idea unit recalled in the appropriate column of the discourse coding sheet.
8) The total number of recall units is tallied.
9) The total number of recall units in each category is tallied.
10) The total number of reproductions is also tallied in two groups: Those reproductions that the subject originally said and those originally spoken by his partner.

An example coding of a fictitious conversation.

A Fictitious Conversation Transcript

Imagine two people in conversation, Person A and Person B. The numbers beside utterances are speaking turns (e.g., Person A says, "It is a nice day" followed by Person B responding with "Yes, it is."). They have been transcribed based on a set of rules that are not specified here (e.g., when a person states two ideas in a compound form like "I play softball and basketball" these ideas are separated into two thought units with full subject and verb phrases).

Person A	Person B
1. It is a nice day.	2. Yes, it is.
3. Where are you from?	4. Houston.
5. What religion are you?	6. Baptist.
7. My husband's a preacher.	8. I go to church.
9. I'm Methodist.	10. Do you like sports?
11. Yes.	
12. I play softball.	
13. I play basketball.	

B's Recall Protocol

Imagine now that after this conversation, Person B is asked to recall what she can. On a blank sheet of paper, Person B writes the following:

1) She said it was a pretty day.
2) I told her I was Baptist and I go to the Second Baptist church.
3) We talked about sports.
4) She plays a lot of sports.
5) It was interesting.
6) He seemed nice.
7) That's all I remember.
8) Oh yea, she said she was married.
9) I told her I was Baptist.
10) She said she was from Dallas.

Now, these responses can be coded according to the rubric developed, as follows:

Rep 1 Rep 6

(She said it was a pretty day.) (I told her I was Baptist)

el Rep 8

and (I go to the Second Baptist church).

T gr

(We talked about sports.) (She plays a lot of sports).

E unc e p

(t was interesting.) (He seemed nice.) That's all I remember.

i

Oh yeah. (she said she was married.)

Rr er

(I told her I was a Baptist.) (She said she was from Dallas.)

Sample Cued Recall Questionnaire Items (Benoit & Benoit, 1988a)

Source: Benoit and Benoit (1988a). Reproduced with permission of Taylor & Francis.

"*Two different kinds of cues were employed in the questionnaire. First, some utterances were elicited by providing the previous utterance as a cue, and subjects were asked to supply the subsequent (target) utterance*. Examples of previous utterance cues include:

What did Subject A say after Subject B said: 'Yesterday I did my laundry'?
What did Subject B say after Subject A said: 'There's still some things I can't figure out'?
What did Subject A say after Subject B said: 'How did you decide to pick communication?'
What did Subject B say after Subject A said: 'I've never lived up here in the summer'?

What did Subject A say after Subject B said: 'Usually I'm very punctual'?

The actual names of the subjects replaced 'Subject A' or 'Subject B' in the questionnaire. Second, some comments were elicited with a semantic cue describing the topic of the target utterance. Examples of semantic cues include:

What did Subject A say about 'courses where you don't know anyone'?
What did Subject B say about 'Texas'?
What did Subject A say about 'the kind of places that make good vacations'?
What did Subject B say about 'Greek identification cards'?
What did Subject A say about 'the kind of place that is good to work at'?" (pp. 21–22)

Sample Recognition Questionnaire Items (Benoit & Benoit, 1990)

Source: Benoit and Benoit (1990). Reproduced with permission of Taylor & Francis.

"*Actual specific verbal recognition item:* Did Judy say something like, 'Are you taking speech 75 now?'

Spurious specific verbal recognition item: Did John say something like, 'So, what did you do over break?'

Actual specific nonverbal recognition item: Did Judy smile while saying something like, 'I had Econ 1 and 2?'

Spurious specific nonverbal recognition item: Did John laugh while saying something like, 'What do you think of this weather?'

Verbal frequency estimate: ______ How many times did Judy refer to John by name?

Nonverbal frequency estimate: ______ How many times did John smile?" (p. 34)

Profile 39

Metacognitive Awareness Listening Questionnaire (MALQ)

(Vandergrift, Goh, Mareschal, & Tafaghodtari, 2006)

Profiled by: Christine C. M. Goh, PhD[1]

National Institute of Education, Nanyang Technological University, Singapore

Construct

The Metacognitive Awareness of Listening Questionnaire (MALQ) is an instrument for eliciting second language (L2) learners' metacognitive awareness about listening (Vandergrift, Goh, Mareschal, & Tafaghodtari, 2006). The construct consists of learners' reported use of strategies in a listening event or task as well as metacognitive knowledge about themselves as L2 listeners, the nature of listening, and listening strategies (Goh, 1997; Vandergrift & Goh, 2012). The MALQ examines four areas of strategy use—problem-solving, planning-evaluation, mental translation, and directed attention—as well as learners' person knowledge.

Instrument Type

Self-report

Description

The MALQ consists of 21 items that represent five areas of metacognitive awareness about L2 listening. These comprise four groups of strategies—*problem-solving strategies* (making and monitoring inferences), *planning and evaluation strategies* (preparing for listening and evaluating results of listening efforts), *mental translation strategies*

1 In memory of Larry Vandergrift (1946–2015), a great scholar and a wonderful colleague.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

(finding equivalents in the learners' own language), and *directed attention strategies* (concentrating and staying on task)—and one dimension of metacognitive knowledge labeled *person knowledge* (perceptions of challenges, anxiety, and self-efficacy). The five factors consist of different numbers of items each: three items (mental translation and person knowledge), four items (directed attention), five items (planning and evaluation), and six items (problem-solving). All items are listed at the end of this profile and are organized by factor. Participants respond to the items along 6-point Likert scaling.

The MALQ was designed to be used in various language instructional settings for three purposes: assessing L2 learners' metacognitive awareness at any point in time of their language development, tracking their metacognitive development in L2 listening at different points of their language learning process, and guiding learners to reflect on their own strategy use and person knowledge. Researchers also can use the MALQ to identify relations between learners' metacognitive awareness and their listening comprehension performance. The MALQ has been translated from English into a number of other languages (e.g., Mandarin, Persian, and Turkish).

Administration

The MALQ is administered after language learners have completed a listening task. An information cover page, which is provided to participants, states that there are no right or wrong answers and that confidentiality will be assured. Administrators of the questionnaire repeat this information and also iterate that the researchers/instructors are only interested in the students' own honest appraisal of how they have attempted to understand the oral texts. The cover page also includes instructions with a sample response item illustrating how to use the answer scales. The MALQ takes about 15–20 minutes to complete.

Scoring

After reverse scoring responses to six items (see Scale section), items are averaged within each subscale to generate five scores per participant. Scoring can be done by the researcher, or participants can be given a guided scoring sheet for self-evaluation. Higher scores for the factors of problem solving, planning and evaluation, and directed attention are desirable as they indicate use of facilitative listening strategies. Lower scores for mental translation and person knowledge (after reverse coding) would suggest less frequent use of translation, which may take up more processing time, and lower listening anxiety. A more nuanced understanding of the interpretation of the scores for the five factors is offered by Goh and Hu (2014).

Development

L2 strategy researchers in the past three decades have argued that learners' use of strategies and awareness of other variables in learning can positively influence learning development (e.g., Cohen & Macaro, 2007; Wenden, 1998). Attention in L2 listening research also has increasingly been directed at learners' self-reports of their understanding and awareness of listening comprehension processes (Goh, 1997; Mendelsohn & Rubin, 1995). At the same time, several questionnaires have been developed to elicit learners' self-reports about L2 listening (see Chamot, Barnhardt, El Dinary, & Robbins, 1999; Goh, 2002; Hasan, 2000; Vandergrift, 2005; Vogely, 1995).

The MALQ is founded on research from the areas of metacognition, listening comprehension, and self-regulation. After reviewing existing instruments assessing L2 listening and reading comprehension processes, an initial list of 88 items was generated. Further assessment (e.g., content validity, clarity, readability, and redundancy) resulted in a preliminary set of 51 items. This draft version of the MALQ was eventually field-tested with a large sample ($N = 966$) in three countries: Canada ($n = 725$), Singapore ($n = 193$), and the Netherlands ($n = 48$). The collected data were subjected to an exploratory factor analysis (EFA) and a principal component analysis (PCA) to find a parsimonious set of items. An initial 13-factor solution resulted, but for reasons of parsimony and meaningfulness, additional analyses were conducted including principal factor analyses (PFAs) with Promax rotation with Kaiser normalization, resulting in a final five-factor solution.

The questionnaire was next examined for the presence of unsatisfactory items due to low factor loadings, complex loadings, and reduced internal consistency estimates—12 items were deleted. Using an iterative process, the remaining 39 items were submitted to additional PFAs with Promax rotation, confirming the five-factor model (information on the additional PFAs was not provided).[2] After each analysis, the items were reevaluated for unsatisfactory items (i.e., low factor loadings, complex loadings, and reduced reliability), and this resulted in the retention of 21 items. Another round of data were collected from a second sample ($N = 512$) using the revised MALQ; these data were submitted to a confirmatory factor analysis (CFA). The hypothesized five-factor model of metacognitive awareness generated from the EFA results was compared with two potential rival models comprising four factors and six factors. The five-factor model comprising 21 items was finally accepted and selected as the model for the MALQ (CFI = .91, RMSEA = .07, TLI = .90).

Reliability

Vandergrift *et al.* (2006) reported internal consistency estimates of sample data ranging from .68 to .78 (Cronbach's alphas: problem solving = .74; planning and evaluation = .75; translation = .78; person knowledge = .74; and directed attention = .68). It is more common for studies to present overall reliability estimates of the MALQ. For example, Cronbach's alpha values of .86 (Rahimia & Katala, 2012) and .76 and .74 (Alamdari & Fahim, 2015) have been reported. Of course, only reporting overall reliability estimates may mask potential issues with subscales, and computing an overall score assumes a second-order latent construct that seems questionable in the case of the MALQ. It is recommended that researchers both confirm the established factor structure as well as assess score reliability (and report the results).

2 The original article does not specify the number of PFAs performed.

Validity

In support of construct validity, the MALQ was submitted to both EFA and CFA during the entire process of its development (as described in this profile) to ensure strong psychometric properties. The procedure for administering the MALQ further aims to address a criticism of learner self-reports in L2 strategy research—that learners do not report accurately when they are asked about the strategies they use, thereby threatening the validity of the self-reports. Students complete the MALQ immediately following an authentic listening activity in class, basing their responses on this activity. This form of immediate verbal reporting, which taps into information still fresh in memory, strengthens the validity of self-reports (Ericsson & Simon, 1987). Al-Alwan, Asassfeh, and Al-Shboul (2013) found that three dimensions of the MALQ, problem solving (24%), planning and evaluation (17%), and directed attention (15%), explained 56% of the total variance in students' performance on a listening comprehension test developed by the authors (the Cronbach's alpha for their listening comprehension scale was .77).

Availability

The original version of the MALQ was first published in *Language Learning* (Vandergrift *et al.*, 2006), and has since been republished with permission in Goh (2008) and Vandergrift and Goh (2012). The measure has been adapted depending on the language and context under study. The MALQ is provided, with permission, at the end of this profile and is free to use for research purposes. An unpublished guide for scoring and interpreting MALQ scores is available from the author of this profile.

Sample Studies

The MALQ has been used to elicit L2 listeners' metacognitive awareness as well as to examine the relationships between metacognitive awareness and L2 listening comprehension. Using a sample of 341 Canadian and Iranian participants, Vandergrift *et al.* (2006) reported a significant association ($r = .36$) between metacognitive awareness of listening processes and comprehension ability. A regression analysis further suggested that metacognition significantly predicted participants' listening scores, indicating that 13% of the variance in listening performance could be explained by metacognition. Zeng (2012) reported that metacognition accounted for about 15% of the variance in the listening performance of his EFL Chinese undergraduates, and an even higher percentage of variance of 22% was reported by Goh and Hu (2014) in their study of 113 ESL learners in Singapore. Their study was also the first to examine intrapersonal variation in different aspects of metacognitive awareness using the MALQ.

The MALQ also has been used for comparing metacognitive awareness development before and after metacognitive instruction in listening. In Vandergrift and Tafaghodtari (2010), L2 learners demonstrated gains in listening performance when they were taught listening strategies; the learners also experienced an increase in metacognitive knowledge at the end of the metacognitive instruction. In another intervention study that assessed changes through the MALQ, O'Bryan and Hegelheimer (2009) found some development in their participants' metacognitive awareness but not in all factors. Greater metacognitive developments were found in the participants in the respective studies by Mareschal (2007) and Zeng (2012) and this was likely due to the longer period of metacognitive intervention in L2 listening.

The MALQ has been adapted to the L1 context (see Metacognitive Listening Instrument, Profile 40).

Critique

Since its publication, Vandergrift *et al.*'s (2006) article on the development of the MALQ has been cited in over 250 studies and scholarly discussions about L2 listening. The questionnaire, which has undergone rigorous validation procedures to ensure robust psychometric properties, has been used in a number of published studies as well as unpublished master's and doctoral dissertations involving L2 learners of various languages, such as French, English, and Arabic, from different parts of the world. Further applications of MALQ can include intrapersonal variations in perceived strategy use and person knowledge for different kinds of listening tasks. One criticism may be that the MALQ may not be fully comprehensive (i.e., that it does not address some strategies that are listed in the L2 listening literature). Although a questionnaire that includes an "exhaustive" list of items may offer a greater degree of face validity, the validation procedures have indicated that many of the original 88 items were problematic as explained in this profile and in the article. Nevertheless, the small number of items for some factors in the MALQ may merit further consideration, and one possible enhancement to the MALQ that is being considered is adding further items to these factors and submitting data to further CFAs. This is especially true given that decisions on inclusion and exclusion of items and factors in the original development studies were made on statistical bases rather than strictly theoretical ones. Other items might tap into additional factors. Perhaps these five factors are the most important, and perhaps they are not; only additional research can determine this. Finally, some work has treated metacognitive listening strategies as a single score (e.g., in reports of internal consistency). The implication is that one can average across all strategies to generate a total "master strategy." This assumption seems questionable on theoretical and empirical grounds. Researchers are encouraged to submit their own data to CFA to test whether a second-order latent factor explains associations among the first-order constructs.

References

Al-Alwan, A., Asassfeh, S., & Al-Shboul, Y. (2013). EFL learners' listening comprehension and awareness of metacognitive strategies: How are they related? *International Education Studies*, *6*, 31–39. doi:10.5539/ies.v6n9p31

Alamdari, E. F., & Fahim, M. (2015). The effect of metacognitive instruction and dialogic interactions on Iranian EFL learners' metacognitive awareness in listening. *Indian Journal of Fundamental and Applied Life Sciences*, *5*(S2), 1368–1376.

Chamot, A. U., Barnhardt, S., El-Dinary, P. B., & Robbins, J. (1999). *The learning strategies handbook*. White Plains, NJ: Longman.

Cohen, A. D., & Macaro, E. (2007). *Language learner strategies: 30 years of research and practice*. Oxford: Oxford University Press.

Ericsson, K. A., & Simon, H. (1987). Verbal reports on thinking. In C. Færch & G. Kasper (Eds.), *Introspection in second language research* (pp. 24–53). Clevedon, UK: Multilingual Matters.

Goh, C. C. M. (1997). Metacognitive awareness and second language listeners. *ELT Journal*, *51*, 361–369. doi:10.1093/elt/51.4.361

Goh, C. C. M. (2002). Learners' self-reports on comprehension and learning strategies for listening. *Asian Journal of English Language Teaching*, *12*, 46–68.

Goh, C. C. M. (2008). Metacognitive instruction for second language listening development: Theory, practice and research implications. *RELC Journal*, *39*, 188–213. doi:10.1177/0033688208092184

Goh, C. C.M., & Hu, G. (2014). Exploring the relationship between metacognitive awareness and listening performance with questionnaire data. *Language Awareness*, *23*, 255–274. doi:10.1080/09658416.2013.769558

Hasan, A. (2000). Learners' perceptions of listening comprehension problems. *Language, Culture and Curriculum*, *13*, 137–153. doi:10.1080/07908310008666595

Mareschal, C. (2007). *Student perceptions of a self-regulatory approach to second language listening comprehension development*. Unpublished doctoral dissertation. University of Ottawa, Canada.

Mendelsohn, D. J., & J. Rubin (1995). *A guide for the teaching of second language listening.* San Diego, CA: Dominie Press.

O'Bryan, A., & Hegelheimer, V. (2009). Using a mixed methods approach to explore strategies, metacognitive awareness and the effects of task design on listening development. *Canadian Journal of Applied Linguistics*, *12*, 9–37.

Rahimi, M., & Katal, M. (2012). Metacognitive listening strategies awareness in learning English as a foreign language: a comparison between university and high-school. *Procedia—Social and Behavioral Sciences*, *31*, 82–89. doi:10.1016/j.sbspro.2011.12.020

Vandergrift, L. (2005). Relationships among motivation orientations, metacognitive awareness and proficiency in L2 listening. *Applied Linguistics*, *26*, 70–89. doi:10.1093/applin/amh039

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. New York: Routledge.

Vandergrift, L., Goh, C. C. M., Mareschal, C., & Tafaghodtari, M. H. (2006). The metacognitive awareness listening questionnaire (MALQ): Development and validation. *Language Learning*, *56*, 431–462. doi:10.1111/j.1467-9922.2006.00373.x

Vandergrift, L., & Tafaghodtari, M. H. (2010). Teaching L2 learners how to listen does make a difference: An empirical study. *Language Learning*, *60*, 470–497. doi:10.1111/j.1467-9922.2009.00559.x

Vogely, A. (1995). Perceived strategy use during performance on three authentic listening comprehension tasks. *The modern language journal*, *79*, 41–56. doi:10.1111/j.1540-4781.1995.tb05414.x

Wenden, A. (1998). Metacognitive knowledge and language learning. *Applied Linguistics*, *19*, 515–37. doi:10.1093/applin/19.4.515

Zeng, Y. (2012). *Metacognition and self-regulated learning (SRL) for Chinese EFL listening development*. Unpublished doctoral dissertation. Nanyang Technological University, Singapore.

Scale

Metacognitive Awareness Listening Questionnaire (MALQ)
(Vandergrift *et al.*, 2006)

Source: Vandergrift *et al.* (2006). Reproduced with permission of John Wiley & Sons.

The statements below describe some strategies for listening comprehension and how you feel about listening in the language you are learning. Do you agree with them?

This is not a test, so there are no "right" or "wrong" answers. By responding to these statements, you can help yourself and your teacher understand your progress in learning to listen.

Please indicate your opinion after each statement. Circle the number which best shows your level of agreement with the statement. For example:

	Strongly disagree	Disagree	Slightly disagree	Partly agree	Agree	Strongly agree
I like learning another language	1	2	3	4	5	6

Please circle only **ONE** number for each statement

Strategy Groups

Problem Solving

1) I use the words I understand to guess the meaning of the words I don't understand.
2) As I listen, I compare what I understand with what I know about the topic.
3) I use my experience and knowledge to help me understand.
4) As I listen, I quickly adjust my interpretation if I realize that it is not correct.
5) I use the general idea of the text to help me guess the meaning of the words that I don't understand.
6) When I guess the meaning of a word, I think back to everything else that I have heard, to see if my guess makes sense.

Planning-Evaluation

7) Before I start to listen, I have a plan in my head for how I am going to listen.
8) Before listening, I think of similar texts that I may have listened to.
9) After listening, I think back to how I listened, and about what I might do differently next time.
10) As I listen, I periodically ask myself if I am satisfied with my level of comprehension.
11) I have a goal in mind as I listen.

Mental Translation

12) I translate in my head as I listen.*
13) I translate key words as I listen.*
14) I translate word by word, as I listen.*

Directed Attention

15) I focus harder on the text when I have trouble understanding.
16) When my mind wanders, I recover my concentration right away.
17) I try to get back on track when I lose concentration.
18) When I have difficulty understanding what I hear, I give up and stop listening.*

Metacognitive Knowledge

Person Knowledge

19) I find that listening in ________ is more difficult than reading, speaking, or writing in _________.*
20) I feel that listening comprehension in __________ is a challenge for me.*
21) I don't feel nervous when I listen to __________.

Note: The blank spaces in items 19–21 should be completed by the researcher and reference the respondents' first language. Labels should be removed and Items randomized prior to administration. Item marked with an asterisk (*) should be reverse-coded prior to scoring. Participants can be given a scoring sheet for self-scoring purposes.

Profile 40

Metacognitive Listening Strategies Instrument (MLSI)

(Janusik & Keaton, 2011, 2015)

Profiled by: Laura A. Janusik, PhD

Rockhurst University

Construct

Measures of metacognitive listening strategies are designed to assess listener awareness of, and ability to regulate, his or her own listening comprehension processes (Goh, 2008).

Instrument Type

Self-Report

Description

Measures of metacognitive listening strategies assess the extent to which listeners are aware of their listening comprehension processes, as well as their ability to regulate those processes via self-appraisal and self-regulation (Goh, 2008; Paris & Winograd, 1990). *Self-appraisal* is recognizing that comprehension is not present, and *self-regulation* is adapting and finding something that will assist with comprehension.

The study of metacognitions related to listening began in the listening to a second language (L2) context. The most widely used instrument, the Metacognitive Awareness Listening Questionnaire (MALQ) (Vandergrift, Goh, Mareschal, & Tafaghodtari, 2006), is profiled in this volume. Briefly, the 21 item MALQ assesses self-perceived listening metacognitions based on five factors: Planning-Evaluation, Problem Solving, Directed Attention, Person Knowledge, and Mental Translation. Janusik and Keaton (2011) attempted to adapt the MALQ to first language (L1) listening. After employing factor-analytic procedures, Janusik and Keaton were able to fit an 11-item, 3-factor model consisting of Planning-Evaluation, Problem Solving, and Directed Attention (Janusik &

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Keaton, 2011). This three-factor version is known as the Metacognitive Listening Strategies Instrument (MLSI). The most recent revised version—the MLSI-R—is a six-item, unidimensional instrument (Janusik & Keaton, 2015).

Administration

The MLSI and MLSI-R may be administered as paper-and-pencil assessments or using online survey software. The MLSI takes between 5 and 15 minutes to complete, and the MLSI-R can be completed in 5 or fewer minutes.

Scoring

Both the MLSI and the MLSI-R use 5-point Likert scaling. For the MLSI, responses are averaged within each factor to generate three scores per respondent. The MLSI-R is scored as a single factor.

Development

The metacognitive model has a long history in the education literature. The model is based on "a construct that refers to thinking about one's thinking or the human ability to be conscious of one's mental processes" (Vandergrift *et al.*, 2006, pp. 432–433). Two broad types of learning strategies have been identified (Nyikos & Oxford, 1993). The first, *cognitive process strategies*, are generally automatic and unconscious, and the second, *metacognitive strategies*, places consciousness, management, and control in the learner's hands (McLaughlin, Rossman, & McLeod, 1983). The MALQ is based on the theoretical concepts of self-appraisal and self-management that come from Information Processing Theory (Vandergrift *et al.*, 2006). Self-appraisal requires a listener to evaluate whether or not she is understanding, and self-management is "metacognition in action," which requires the listener to select a strategy to enhance comprehension.

As noted, the MLSI is based on the MALQ, which was designed for students learning a second language, or L2 listening (Vandergrift *et al.*, 2006). When translating the MALQ to the first language to create the MLSI, Janusik and Keaton (2011) made four modifications. First, probes associated with switching from one language to another were dropped (i.e., two of the three items from the Mental Translation factor). Second, items were contextualized to listening in the classroom (for a rationale on the importance of contextualizing listening measures, see Imhof, 2004). Third, based on Zull's (2002) work on learning and the brain, the following probe was added: "When I listen to class lectures and discussions, I consciously try to connect what I'm hearing to things I've learned in other (not this) classes." Finally, a second probe was added that equates to the listening process: "I consciously make meaning in my head as I listen to class lectures and discussions."

Thus, the MLSI began as a 20-item instrument of listening metacognitions in a classroom setting (Janusik & Keaton, 2011). Testing on a small sample at a private Midwestern university reduced the 20 items to 11 items that loaded on 3 factors. The three factors retained reflected three of the original MALQ factors, namely Directed Attention, Planning-Evaluation, and Problem Solving.

Because the MLSI was only tested on a small sample of US students ($N = 142$), and because it represented listening in one's first language as opposed to one's second language, two questions arose (Janusik & Keaton, 2015). First, would the factor structure hold with a larger US population? Second, are the probes specific to first-language listening, regardless of language, or might one's conceptualization of listening (Imhof & Janusik, 2006) influence and determine one's metacognitive listening strategies?

Subsequently, the MLSI was tested with a sample of 876 students from 37 colleges and universities across the United States. Only data from those completing the instrument in their first language were used. Results of the expanded study found that the three-factor structure did not hold. The instrument was further reduced from 11 items to 6 items and renamed the MLSI-R.

Gathering evidence for validity and score reliability continues. A recent study by Imhof, Janusik, and Keaton (2014) tested the MLSI with Finnish, German, and Japanese first-language speakers (Finland, $n = 113$; Germany, $n = 129$; Japanese, $n = 327$). The factor structure of the MLSI did not fit; however, the factor structure of the MLSI-R did (Janusik & Keaton, 2015).[1] In addition, there was no clear pattern of relations between listening concepts and listening metacognitions.

Reliability

For comparison purposes, reliability estimates of the three iterations of the MSLI are reported here (Janusik & Keaton, 2011, 2013, 2015). The initial study of the 20-item MSLI (Janusik & Keaton, 2011) indicated that the data did not fit the MALQ factor structure in a confirmatory factor analysis. Cronbach's alphas for each factor were as follows: Planning-Evaluation ($\alpha = .64$), Problem Solving ($\alpha = .71$), Directed Attention ($\alpha = .62$), and Person Knowledge ($\alpha = .35$). Due to the lack of fit, the authors submitted data to an exploratory factor analysis, which subsequently resulted in the 11-item best-fit model of Planning-Evaluation ($n = 4$, $\alpha = .69$), Problem Solving ($n = 3$, $\alpha = .70$), and Directed Attention ($n = 4$, $\alpha = .69$).

Testing the 11-item instrument on an independent US population, Janusik and Keaton (2015) failed to find an adequate model: Evaluation ($n = 4$, $\alpha = .59$), Problem Solving ($n = 3$, $\alpha = .40$), and Directed Attention ($n = 4$, $\alpha = .56$). A follow-up exploratory factor analysis resulted in the 6-item unidimensional scale ($n = 6$, $\alpha = .78$). This factor structure also fit the first-language cultures in Finland ($\alpha = .67$), Germany ($\alpha = .68$), and Japan ($\alpha = 83$).

Validity

The only evidence of construct validity for the MLSI comes in the form of confirmatory factor analyses. In general, the original 3-factor MLSI has not been replicated, whereas the unidimensional MLSI-R seems to fit across samples in the United States as well as Germany, Finland, and Japan.

1 US data model fit: χ^2 (9) = 93.38, $p < .000$, CFI = .91, SRMR = .04, RMSEA = .10, CI90% = .09–.12. Japanese data model fit: χ^2 (9) = 46.69, $p < .000$, CFI = .93, SRMR = .05, RMSEA = .11, CI90% = .08–.14. German sample model fit: χ^2 (9) = 8.34, $p < .50$, CFI = 1.00, SRMR = .03, RMSEA = .000, CI90% = .000–.09). Finnish data model fit: χ^2 (9) = 11.93, $p < .22$, CFI = .97, SRMR = .04, RMSEA = .05, CI90% = .000–.13.

Availability

The MLSI is provided here, with permission. For the initial 20-item instrument, see Janusik and Keaton (2011). The 6-item unidimensional instrument (MLSI-R) was introduced in the *Journal of Intercultural Communication Research* (Janusik & Keaton, 2015); items that comprise the MLSI-R are noted here. For additional information on the MALQ, upon which the MLSI is based (Vandergrift *et al.*, 2006), please see the MALQ profile (Profile 39).

Sample Studies

The MLSI and its derivative, the MLSI-R, are quite new. As a result, no completed studies exist beyond those previously described (Imhof *et al.*, 2014; Janusik & Keaton, 2011, 2015; Janusik, Keaton, & Imhof, 2012).

One recent study approached US metacognitive listening strategies from grounded theory perspective (Janusik & Varner, 2015). Preliminary results supported previous findings suggesting that listening metacognitions are not intuitive (Graham, Santos, & Vanderplank, 2011; Janusik & Keaton, 2015). Preliminary results also indicated that US students rely more on physical and interactive listening strategies, such as turning off cell phones or asking another student for help, as opposed to metacognitive ones.

Critique

The MLSI and MLSI-R are based on the same theoretical framework that grounds the MALQ (Vandergrift *et al.*, 2006). This includes the model of metacognition (Flavell, 1979; Metcalfe & Shimanaura, 1994), as well as the self-appraisal and self-regulation components of Information Processing Theory.

For educational purposes, it seems relevant for students to understand available metacognitive listening strategies and to reflect on and utlize those strategies. The extensive literature from the L2 context suggests that because most metacognitive listening strategies are not intuitive (Graham *et al.*, 2011), instruction and reinforcement are necessary to increase student use of such strategies (Baleghizadeh & Rahimi, 2011; Birjandi & Rahimi, 2012; Freday, 2011; Goh & Taib, 2006; Liao, 2009; Selamat & Sidhu, 2011; Vandergrift & Tafaghodtari, 2010). However, when taught and reinforced, increased metacognitive listening strategies use leads to increased comprehension (Freday, 2011; Mareschal, 2007; Vogely, 1995).

From prior work, we know students use a restricted range of strategies to improve their listening (Imhof, 1998; Janusik & Keaton, 2011, 2015; Janusik & Varner, 2015; Stein, 1999). Given these positive results, it is important for future work to establish a sound scale for measuring the most salient metacognitive strategies for various populations. The MALQ seems to measure five such factors in a consistent fashion in the L2 context. Perhaps the strategies used by L2 listeners are not the same strategies used by L1 listeners, and thus any scale designed for the latter context should be developed anew. For scale-building purposes, future research must first focus on identifying the metacognitive listening strategies that increase student comprehension and then develop items that will reliably tap these dimensions. Subsequent research can then identify effective methods to teach the strategies.

The MLSI has yet to be replicated, thus begging the question as to whether future work should utilize these items at all. Although the MLSI-R has demonstrated adequate fit across samples, a unidimensional measure of *strategies* seems, at best, only a start. Three problem-solving items and three directed action items comprise the MLSI-R, whereas none of the planning-evaluation items made the final cut. Planning and evaluation have been measured in other contexts, and items from these more established scales (e.g., the Survey of Imagined Interactions; Honeycutt, 2009), along with exploratory work like that conducted by Janusik and Varner (2015), could be used as guides toward building a multidimensional metacognitive listening strategies instrument for the L1 context. While the work on the scale's validity portfolio continues, its originators encourage its use as a descriptive instrument to help L1 listeners develop metacognitive awareness.

References

Baleghizadeh, S., & Rahimi, A. M. (2011). The relationship among listening performance, metacognitive strategy use and motivation from a self-determination theory perspective. *Theory and Practice in Language Studies*, *1*, 61–67. doi:10.4304/tpls.1.1.61-67

Birjandi, P., & Rahimi, A. H. (2012). The effect of metacognitive strategy instruction on the listening performance of EFL students. *International Journal of Linguistics*, *4*, 495–517. doi:10.5296/ijl.v4i2.1707

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*, 906–911. doi:10.1037/0003-066X.34.10.906

Freday, M. E. (2011). *Strategic listening: Examining the efficacy of a pedagogical cycle intended to teach the listening process*. Unpublished Master's thesis. Ohio University, Athens, OH. Retrieved from https://etd.ohiolink.edu/!etd.send_file?accession=ohiou1307126560&disposition=inline

Goh, C. C. (2008). Metacognitive instruction for second language listening development theory, practice and research implications. *RELC Journal*, *39*, 188–213. doi:10.1177/0033688208092184

Goh, C., & Taib, Y. (2006). Metacognitive instruction in listening for young learners. *ELT Journal*, *60*, 222–232. doi: 10.1093/elt/ccl002

Graham, S., Santos, D., & Vanderplank, R. (2011). Exploring the relationship between listening development and strategy use. *Language Teaching Research*, *15*, 435–456. doi:10.1177/1362168811412026

Honeycutt, J. M. (Ed.) (2009). *Imagine that: Studies in imagined interaction*. Cresskill, NJ: Hampton Press.

Imhof, M. (1998). What makes a good listener? Listening behavior in instructional settings. *International Journal of Listening*, *12*, 81–105. doi:10.1080/10904018.1998.10499020

Imhof, M. (2004). Who are we as we listen? Individual listening profiles in varying contexts. *International Journal of Listening*, *18*, 36–46. doi:10.1080/10904018.2004.10499061

Imhof, M., & Janusik, L. A. (2006). Development and validation of the Imhof-Janusik Listening Concepts Inventory to measure listening conceptualization differences between cultures. *Journal of Intercultural Communication Research*, *35*, 79–98. doi:10.1080/17475750600909246

Imhof, M., Janusik, L.A., & Keaton, S. (2014, March). *Listening concepts and metacognition in a cross-cultural perspective*. Paper presented at the International Listening Association Convention, Minneapolis, MN.

Janusik, L. A., & Keaton, S. A. (2011). Listening metacognitions: Another key to teaching listening? *Listening Education*, *3*, 33–44. doi:www.listen.org/listening_education
Janusik, L., & Keaton, S. (2013, November). *Towards developing a cross-cultural metacognition scale for listening in the first language (L1) contexts: The metacognitive listening scale*. Paper presented at the National Communication Association convention, Washington DC.
Janusik, L., & Keaton, S. (2015). Towards developing a cross-cultural metacognition listening instrument for listening in first language (L^1) contexts: The (Janusik-Keaton) Metacognitive Listening Instrument. *Journal of Intercultural Communication Research*, *44*, pp. 288–306. doi:10.1080/17475759.2015.1073169
Janusik, L., Keaton, S., & Imhof, M. (2012, November). The relation between listening conceptualizations and listening metacognitions: Presage to process. Paper presented at the National Communication Association Convention, Orlando, FL.
Janusik, L. A., & Varner, T. (2015, November). (Re)discovering metacognitive listening strategies. Paper presented at the annual meeting of the National Communication Association, Las Vegas, NV.
Liao, K. C. (2009). *A study of metacognitive awareness about English listening of Taiwanese English majors*. Unpublished master's thesis. National Pingtung University of Education, Pingtung, Taiwan. Retrieved from http://140.127.82.35/ETD-db/ETD-search/view_etd?URN=etd-0721109-113325
Mareschal, C. (2007). *Student perceptions of a self-regulatory approach to second language listening comprehension development*. Unpublished doctoral dissertation. University of Ottawa, Ontario, Canada. Retrieved from http://www.ruor.uottawa.ca/handle/10393/29447
McLaughlin, B., Rossman, T., & McLeod, B. (1983). Second language learning: An information-processing perspective. *Language Learning*, *33*, 135–158. doi:10.1111/j.1467-1770.1983.tb00532.x
Metcalfe, J., & Shimamura, A. P. (Eds.). (1994). *Metacognition*. Cambridge, MA: MIT Press.
Nyikos, M., & Oxford, R. (1993). A factor analytic study of language-learning strategy use: Interpretations from information-processing theory and social psychology. *The Modern Language Journal*, *77*, 11–22. doi:10.1111/j.1540-4781.1993.tb01940.x
Paris, S. G., & Winograd, P. (1990). Promoting metacognition and motivation of exceptional children. *Remedial and Special Education*, *11*, 7–15. doi:10.1177/074193259001100604
Selamat, S., & Sidhu, G. K. (2011). Student perceptions of meta-cognitive strategy use in lecture listening comprehension. *Language Education in Asia*, *2*, 185–198. doi:org/10.5746/LEiA/11/V2/I2/A02/Selamat_Sidhu
Stein, S. K. (1999). *Uncovering listening strategies: Protocol analysis as a means to investigate student listening in the basic communication course*. Unpublished doctoral dissertation. University of Maryland, College Park, MD.
Vandergrift, L., Goh, C. C., Mareschal, C. J., & Tafaghodtari, M. H. (2006). The metacognitive awareness listening questionnaire: Development and validation. *Language Learning*, *56*, 431–462. doi:10.1111/j.1467-9922.2006.00373.x
Vandergrift, L., & Tafaghodtari, M. H. (2010). Teaching L2 learners how to listen does make a difference: An empirical study. *Language Learning*, *60*, 470–497. doi:10.1111/j.1467-9922.2009.00559.x
Vogely, A. (1995). Perceived strategy use during performance on three authentic listening comprehension tasks. *The Modern Language Journal*, *79*, 41–56. doi:10.1111/j.1540-4781.1995.tb05414.x
Zull, J. E. (2002). *The art of changing the brain. Enriching the practice of teaching by exploring the biology of learning*. Sterling, VA: Stylus.

Scale

Metacognitive Listening Strategies Instrument (MLSI) (Janusik & Keton, 2011)

Source: Janusik and Keton (2011). Reproduced with permission of the International Listening Association.

Instructions: Think of the **least** favorite class that you're taking this school term. Answer each of these questions concerning this class only. There are no right or wrong answers, and your honesty will help us better understand how to help you. Using the following scale, mark your level of agreement/disagreement with each item:

1 = Strongly Disagree
2 = Disagree
3 = Neither Agree nor Disagree
4 = Agree
5 = Strongly Agree

<u>Problem-Solving</u>

1) I use the words I understand to guess the meaning of the words I don't understand when listening to class lectures and discussions.
2) I use the general idea of the lecture or discussion to help me guess the meaning of the words that I don't understand.
3) When I guess the meaning of a word in a class lecture or discussion, I think back to everything else that I have heard, to see if my guess makes sense.

<u>Planning-Evaluation</u>

4) Before I start to listen in class, I have a plan in my head for how I am going to listen.
5) After listening in class, I think back to how I listened, and about what I might do differently next time in class.
6) As I listen in class, I periodically ask myself if I am satisfied with my level of comprehension.
7) I have a goal in mind as I listen in class.

<u>Directed Attention</u>

8) I consciously make meaning in my head as I listen to class lectures and discussions.
9) When my mind wanders in class, I recover my concentration right away.
10) I try to get back on track when I lose concentration while listening in class.
11) As I listen in class, I quickly adjust my interpretation if I realize that it is not correct.

Note: Labels should be removed and items randomized prior to administration. Items 1, 2, 3, 8, 10, and 11 comprise the MLSI-R.

Profile 41

Microanalysis of Face-to-Face Dialogue (MFD)

Profiled by: Janet Bavelas, Jennifer Gerwing, Sara Healing, and Christine Tomori[1]

Construct

Microanalysis of face-to-face dialogue (MFD) is defined as the systematic examination of specific observable behaviors in face-to-face dialogue, focusing on their immediate communicative functions. Its applications are broader than a single construct or set of constructs.

Instrument

Behavioral observation

Description

MFD was developed to study the unique features of communication in face-to-face dialogue, which differs from other forms of communication (e.g., writing and public speaking) in several respects: (a) The speaker is talking to a particular person, the addressee (e.g., Bavelas & Gerwing, 2011); (b) the speaker and addressee create their dialogue collaboratively through highly interdependent actions (e.g., Bavelas, 2011); and (c) the speaker and addressee use multimodal, integrated messages consisting of words, prosody, hand gestures, facial gestures, and gaze (e.g., Bavelas & Chovil, 2006).

Specific applications of MFD are as varied as the behavioral phenomena that researchers might find interesting in dialogue. In general, the purpose of the method is to understand

1 The authors appreciate the contribution of Graham D. Bodie to this profile.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

what those behaviors reveal about how dialogue works (rather than what they might reveal about mental processes or abstract variables). In short, MFD focuses on the details of social interaction—details that often occur at the level of seconds. The method is applicable to any setting where dialogue is important, including but not limited to face-to-face dialogues in lab experiments, psychotherapy sessions, doctor–patient consultations, and mediation.

Administration

It is not necessary to use any previous MFD system; instead, researchers can follow the inductive approach that is common in other areas of science (Bavelas, 1987). Using this kind of microanalysis requires digitized video-recorded conversations between two or more individuals who are continuously visible and audible at all times. We recommend analysis with ELAN software (http://tla.mpi.nl/tools/tla-tools/elan/; Wittenburg, Brugman, Russel, Klassman, & Sloetjes, 2006). ELAN is a free download from the Max Planck Institute for Psycholinguistics that permits viewing at any speed, including frame by frame. The annotation feature lets analysts make any number of selections that identify specific occurrences of a behavior of interest (e.g., questions, addressee responses, or hand or facial gestures), then save those selections as annotations, and add their own text to the annotation (e.g., a transcription, a description, or a function). The result is a direct and stable link between the annotation and the relevant behavior on the video. The ability to layer multiple, overlapping annotations in any number of tiers can make these direct links superior to a separate transcript. The annotations can be exported to other formats for sorting or to generate summaries or comparisons. Even with appropriate software, the rapid and precise nature of face-to-face dialogue makes MFD a labor-intensive method, often involving high ratios of analysis time to real time on the video.

Scoring

The term *scoring* implies the goal of parametric data. A useful feature of MFD is that it spans the qualitative versus quantitative divide. That is, most behaviors examined with MFD are nominal. For example, a psychotherapist's formulation may be a combination of words that the client said, paraphrases of these words, and/or the therapist's own technical terms (Korman, Bavelas, & De Jong, 2013). Similarly, mutual gaze either occurs at a particular moment or not (Bavelas, Coates, & Johnson, 2002). If, however, the analyst wishes to have parametric data for statistical tests, the nominal data can readily be quantified as frequencies, rates, proportions, or the like without destroying its qualitative nature (for numerous examples, see Bavelas, Gerwing, Sutton, & Prevost, 2008).

Development

The Natural History of an Interview (NHI) project at the Center for Advanced Study in the Behavioral Sciences was "the first major study to use microanalysis as the primary method of analyzing social interaction" (Leeds-Hurwitz, 1987, p. 2). The NHI project focused on therapeutic interviews, which were beginning to be of interest to other research groups, especially the Palo Alto group, one of the first to focus on interpersonal communication more generally (Watzlawick, Beavin Bavelas, & Jackson, 1967). Beginning in the 1980s, Bavelas and her colleagues at the University of Victoria began using microanalysis to study addressees' responses to a communicative conflict (Bavelas, Black, Chovil, & Mullet, 1990; Bavelas & Smith, 1982) and motor mimicry (Bavelas, Black, Lemery, MacInnis, & Mullett, 1986; Bavelas, Black, Lemery, & Mullett, 1986; Chovil, 1989, 1991). As it became clear that these elusive phenomena were orderly when viewed closely and in detail, MFD slowly emerged as a unique, systematic approach.

By the 1990s, MFD was also confirming that, when interacting face to face, the participants have visible as well as audible means of communication available at all times. The research group began to study hand and facial gestures (see review in Bavelas, Gerwing, & Healing, 2014) and gaze (Bavelas *et al.*, 2002), which are tightly synchronized in both timing and meaning to the ongoing speech. For example, Chovil (1989, 1991/1992) conducted the first-ever study of how interlocutors' faces function in dialogue, which differed from previous studies of the emotions that tended to focus on faces of individuals who were either alone (not interacting with anyone) or in still photos. In Chovil's study, 12 dyads discussed a variety of conversational topics (e.g., close calls and planning a meal) while recorded in split screen with a close-up of each face. Chovil viewed the video recordings intensively, often in slow motion, because the interlocutors' facial actions in a dialogue are both rapid and transient, with a precise temporal and semantic relationship to the accompanying speech (rather than displaying an emotional state). As is typical for MFD, the analysis was both inductive and functional, asking, "What is this display doing at this point in the conversation? How is it conveying meaning in the context in which it occurs?" (Bavelas & Chovil, 1997, p. 341). In almost 900 meaningful displays, Chovil found about 40 different conversational functions, which were then confirmed by an independent analyst. The person speaking at the moment made most of these displays, but the addressees' faces were also active, primarily providing information that overlapped but did not interrupt the speaker. The most common function of addressees' displays was analogous to verbal back channels (e.g., raising the eyebrows or squinting slightly). Two other common functions were personal reactions (e.g., a "disgust" face at the mention of a disliked food) and motor mimicry (e.g., wincing as the speaker described being injured).

With other scholars, the Victoria Group proposed that dialogue is composed of functionally integrated speech and co-speech acts, rather than physically separable ("verbal" and "nonverbal") actions. These integrated, multimodal utterances serve the other unique feature of dialogue, which is the constant moment-by-moment collaboration between participants that produces their conversation. For example, the addressee's facial gestures, as described, both influence and are influenced by the speaker.

In this century, MFD has moved outside the lab and examined psychotherapy sessions (e.g., Korman *et al.*, 2013; Smock Jordan, Froerer, & Bavelas, 2013; Tomori & Bavelas, 2007), medical interactions (e.g., Gerwing & Dalby, 2014; Healing, 2013; Gerwing & Indseth, 2015; Gerwing, Indseth, & Gulbrandsen, 2016), parent–infant interaction (Gerwing, 2008), and computer-mediated versus face-to-face interactions (Phillips, 2007).

Close analysis of dialogue is not unique to MFD. Indeed, the method has much in common with discourse and conversation analytic methods as developed in linguistics, anthropology, communication studies, psychology, and other allied disciplines. A key difference is in the explicit, experimentally derived framework of theoretical assumptions that guides the analysis: Dialogue consists of the interlocutors' collaborative, coordinated actions rather than the isolated actions of individuals; verbal and nonverbal behaviors are integrated rather than separable; and behavioral sequences are understandable and interesting in themselves without making mental inferences or attributions about the interlocutors. Within this framework, MFD analysts often prefer to begin with an inductive approach, deriving their analysis from the particulars of their data and developing a comprehensive and systematic set of definitions and procedures for analysis, tailored to that particular project and the phenomenon they have uncovered. Using the new system, it is then possible to go on to the more familiar deductive, hypothesis-testing phase.

Reliability

The reliability of MFD applications is most often computed as interanalyst agreement. Our group aims for 90% or better agreement among independent analysts. The value of establishing a relatively high level of interanalyst agreement for microanalysis was emphasized by Bavelas, Kenwood, and Phillips (2002):

> Achieving high agreement on complex interpretations of discourse requires careful and explicit description of the interpretive and reasoning process. This requirement is as valuable as the goal of demonstrating agreement itself – we always understand the phenomenon more clearly and deeply after we have done the hard and iterative work of describing (and debating!) it sufficiently to achieve agreement. (p. 114)

Two levels of agreement are important for MFD. First, analysts have to agree on what constitutes the unit of analysis. Because dialogue occurs as a stream of interconnected behaviors, the analysts must divide the stream into units and agree on their number and placement (e.g., locating all addressee responses to a close-call story or all utterances by patients in oncology consultations). After assessing their agreement and resolving disagreements, the analysts must then agree on the function of each unit. For example, was the addressee's response generically or specifically related to what the speaker said (Bavelas, Coates, & Johnson, 2000)? Did the patient's utterance provide biomedical information or information about this individual patient's perspective (Healing, 2013)? MFD does not create categories or taxonomies using formal properties of behaviors, because these inevitably abstract the act from its function in the immediate dialogue. Instead, MFD aims to sort the behaviors of interest according to their function at that moment in the dialogue. (Chapter 6 of this volume provides more details on these elements of interrater reliability.)

Validity

Because MFD does not assess traits or states of the individuals or characteristics of their relationship, many of the traditional validity criteria (e.g., temporal or cross-situational consistency) are not relevant. Instead, internal and construct validity are required.

Internal validity addresses possible biases that could affect interpretation of results. For example, it is often appropriate to use naïve as well as expert analysts to show that the analysis is not idiosyncratic or that being aware of the hypothesis does not affect the analyst's decision. For example, to ensure internal validity in a study of generic versus specific listener responses (Bavelas *et al.*, 2000), two analysts knew nothing of the hypothesis or rationale for the distinction they were making. The same study tested the construct validity of the distinction between generic and specific listener responses by testing (and confirming) hypotheses about differences in where they would occur within a narrative as well as the effect of the absence of specific responses on the quality of the narrative.

Most experimentalists will admit that internal and external validity are usually antithetical—the tighter the controls for internal validity, the less the situation may resemble any other contexts. Simply moving outside the lab, however, does not guarantee external validity either, because there is not just one "nonlab" context out there in the world, nor is random sampling of a given context usually an option. Sometimes, it is helpful to narrow the question. When asking how psychotherapists with different theoretical models shape their therapeutic dialogues (Korman *et al.*, 2013; Smock Jordan *et al.*, 2014; Tomori & Bavelas, 2007), random sampling of all possible psychotherapy sessions was obviously not an option. Instead, we chose to focus on sessions that had been conducted by experts for the explicit purpose of demonstrating their approach and training others in it. Our conclusions were that experts in at least two models differed significantly in how they talked with clients, so the models as presented by experts did differ. In any case, caution about generalization has to be a guideline for all researchers. That is, rather than concluding that our results show that people behave a certain way, it is more accurate to conclude that, in this context, with these characteristics, these people tended to behave a certain way.

Availability

It is probably clear by now that there is no single coding system available for MFD. It is a metamethod, with dozens of examples of how different researchers have used the theoretical framework and procedures to pursue their particular interest. Published articles as well as information on seminars, workshops, and training in MFD can be found at http://www.microanalysis.ca/.

Sample Studies

Because of the focus of the *Sourcebook*, this profile has used primarily examples that involved addressees (e.g., their facial gestures, formulations, generic and specific responses, and motor mimicry, as well as infant responses). Bavelas, Gerwing, Healing, and Tomori (2016) included an appendix that summarizes 24 MFD studies, of which at least 15 focused either on the addressees' actions or on their moment-by-moment collaboration with the speaker. Several studies have focused specifically on the influence of addressees who were psychotherapists (Korman *et al.*, 2013), emergency telephone operators (Gerwing & Indseth, 2015), or just listening to a close-call story (e.g., Bavelas *et al.*, 2000; Chovil, 1992).

Critique

MFD was developed exclusively for the systematic and close analysis of interaction in face-to-face dialogue. The method has helped to demonstrate that dialogues are a process of co-construction in which addressees play a crucial role. Although MFD has made important contributions to theory and practice in its area, absent from this research is a focus on outcomes deemed important in the wider personal, relational, or professional lives of the interlocutors. That is, although MFD has shown with great precision how dialogue functions in a variety of settings, there is rarely any assessment of the impact of these specific functions outside the dialogue, on variables or outcomes that Duncan (1969) called external to the dialogue. There are two recent exceptions: Gerwing (2008) studied infant triplets' social responsiveness, which accurately predicted the onset of autism in one infant. Healing (2013) identified all instances of patient-centered information in oncology consultations and related them to the patients' subsequent reports on the degree to which their goals were met, their understanding of the information the oncologist provided, and the decisions that were made in the consultation. Note, however, that the outcomes in both cases were very closely related to the focus of the MFD analysis. We would not, for example, predict where the infant would ultimately be on the autism spectrum or whether patients who contributed more patient-centered information would be more likely to comply with treatment. We respect the importance of far too many other factors in the panorama of these individuals' lives. Instead, the value of MFD arises from the importance and ubiquity of face-to-face dialogue in human social life.

References

Bavelas, J. B. (1987). Permitting creativity in science. In D. N. Jackson & J. P. Rushton (Eds.), *Scientific excellence: Origins and assessment* (pp. 307–327). Beverly Hills, CA: Sage.

Bavelas, J. B. (2011). Connecting the lab to the therapy room. Microanalysis, co-construction, and solution-focused therapy. In C. Franklin, T. Trepper, W. J. Gingerich, & E. McCollum (Eds.), *Solution-focused brief therapy: From practice to evidence-informed practice* (pp. 144–162). Oxford: Oxford University Press.

Bavelas, J. B., Black, A., Chovil, N., & Mullett, J. (1990). *Equivocal communication*. Newbury Park, CA: Sage.

Bavelas, J. B., Black, A., Lemery, C. R., MacInnis, S., & Mullett, J. (1986). Experimental methods for studying "elementary motor mimicry." *Journal of Nonverbal Behavior*, *10*, 102–119. doi:10.1007/BF01000007

Bavelas, J. B., Black, A., Lemery, C. R., & Mullett, J. (1986). "I show how you feel." Motor mimicry as a communicative act. *Journal of Personality and Social Psychology*, *50*, 322–329. doi:http://dx.doi.org/10.1037/0022-3514.50.2.322

Bavelas, J. B., & Chovil, N. (1997). Faces in dialogue. In J. A. Russell & J. M. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 334–346). Cambridge: Cambridge University Press.

Bavelas, J. B., & Chovil, N. (2006). Hand gestures and facial displays as part of language use in face-to-face dialogue. In V. Manusov & M. Patterson (Eds.), *Handbook of nonverbal communication* (pp. 97–115). Thousand Oaks, CA: Sage.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, *79*, 941–952. doi:http://dx.doi.org/10.1037/0022-3514.79.6.941

Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. *Journal of Communication*, *52*, 566–580. doi:10.1111/j.1460-2466.2002.tb02562.x

Bavelas, J. B., & Gerwing, J. (2011) The listener as addressee in face-to-face dialogue. *International Journal of Listening*, *25*, 178–198. doi:10.1080/10904018.2010.508675

Bavelas, J. B., Gerwing, J., & Healing, S. (2014). Hand gestures and facial displays in conversational interaction. In T. Holtgraves (Ed.), *Oxford handbook of language and social psychology* (pp. 111–130). New York: Oxford University Press.

Bavelas, J. B., Gerwing, J., Healing, S., & Tomori, C. (2016). Microanalysis of face-to-face dialogue: An inductive approach. In C. A. Van Lear & D. J. Canary (Eds.), *Researching communication interaction behavior: A sourcebook of methods and measures* (pp. 129–156). Thousand Oaks, CA: Sage.

Bavelas, J. B., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, *58*, 495–520. doi:10.1016/j.jml.2007.02.004

Bavelas, J. B., Kenwood, C., & Phillips, B. (2002). Discourse analysis. In M. Knapp & J. Daly (Eds.), *Handbook of Interpersonal Communication* (3rd ed., pp. 102–129). Thousand Oaks, CA: Sage.

Bavelas, J. B., & Smith, B. J. (1982). A method for scaling verbal disqualification. *Human Communication Research*, *8*, 214–227. doi:10.1111/j.1468-2958.1982.tb00665.x

Chovil, N. (1989). *Communicative functions of facial displays in conversation*. Unpublished PhD dissertation, University of Victoria, Victoria, BC, Canada.

Chovil, N. (1991a). Discourse-oriented facial displays in conversation. *Research on Language and Social Interaction*, *25*, 163–194. doi:10.1080/08351819109389361

Chovil, N. (1991b). Social determinants of facial displays. *Journal of Nonverbal Behavior*, *15*, 141–154. doi:10.1007/BF01672216

Duncan, S., Jr. (1969). Nonverbal communication. *Psychological Bulletin*, *72*, 118–137. doi:10.1037/h0027795

Gerwing, J. (2008). *Quantifying infant social responsiveness: Microanalysis of home videos of a set of triplets for early indications of autism*. Unpublished doctoral dissertation. University of Victoria, Victoria, BC, Canada. Retrieved from https://dspace.library.uvic.ca/handle/1828/1226?show=full.

Gerwing, J., & Dalby, A. M. (2014). Gestures convey content: An exploration of the semantic functions of physicians' gestures. *Patient Education and Counseling*, *96*, 308–314. doi:10.1016/j.pec.2014.07.014

Gerwing, J., & Indseth, T. (2015) Enhancing caller comprehensibility in emergency calls with a language barrier: Operators in the role of addressee. *International Journal of Listening*, *30*, 163–179. doi:10.1080/10904018.2015.1056878

Gerwing, J., Indseth, T., & Gulbrandsen, P. P. (2016). A microanalysis of the clarity of information in physicians' and patients' discussions of treatment plans with and without language barriers. *Patient Education and Counseling*, *99*, 522–529. doi:10.1016/j.pec.2015.10.012

Healing, S. (2013). *Development of a method of analysis for identifying an individual patient's perspective in video-recorded oncology consultations*. Unpublished master's thesis. University of Victoria, Victoria, BC, Canada. Retrieved from http://hdl.handle.net/1828/4835

Korman, H., Bavelas, J. B., & De Jong, P. (2013). Microanalysis of formulations in solution-focused brief therapy, cognitive behavioral therapy, and motivational interviewing. *Journal of Systemic Therapies*, *32*, 32–45. doi:0.1521/jsyt.2013.32.3.31

Leeds-Hurwitz, W. (1987). The social history of The Natural History of an Interview: A multidisciplinary investigation of social communication. *Research on Language and Social Interaction*, *20*, 1–51. doi:10.1080/08351818709389274

Phillips, B. (2007). *A comparison of autonomous and collaborative models in computer-mediated communication*. Unpublished doctoral dissertation, Department of Psychology, University of Victoria, Victoria, BC, Canada. Retrieved from http://hdl.handle.net/1828/226

Smock Jordan, S., Froerer, A., & Bavelas, J. B. (2013). Microanalysis of positive and negative content in solution-focused brief therapy and cognitive behavioral therapy expert sessions. *Journal of Systemic Therapies*, *32*, 46–59. doi:10.1521/jsyt.2013.32.3.46

Tomori, C., & Bavelas, J. B. (2007). Using microanalysis of communication to compare solution-focused and client-centered therapies. *Journal of Family Psychotherapy*, *18*, 25–43. doi:10.1300/J085v18n03_03

Watzlawick, P., Beavin Bavelas, J., & Jackson, D. D. (1967). *Pragmatics of human communication*. New York: Norton.

Wittenburg, P., Brugman, H., Russel, A., Klassman, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation* (pp. 1556–1559). Genoa, Italy: European Language Resources Association. Retrieved from http://www.lrec-conf.org/proceedings/lrec2006/pdf/153_pdf.pdf

Profile 42

Multidimensional Evaluation of Enacted Social Support (MEESS)

(Goldsmith, McDermott, & Alexander, 2000)

Profiled by: Daena J. Goldsmith and Abbey Griscom

Lewis & Clark College

Construct

The Multidimensional Evaluation of Enacted Social Support (MEESS) was designed to measure individuals' multifaceted evaluations of social support they provide, receive, or observe.

Instrument Type

Self-Report; Behavioral Observation

Description

The MEESS differentiates among three criteria by which a message, conversation, or other enactment of support might be evaluated. The *problem-solving utility* dimension refers to the informational and instrumental benefits of an interaction, including whether the social support is helpful, knowledgeable, useful, and generous. The *relational assurance* dimension taps loyalty and standing by someone as expressed in being supportive, comforting, reassuring, and encouraging. The *emotional awareness* dimension reflects being attuned to feelings by showing sensitivity, understanding, compassion, and consideration.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

The 12 semantic differential items that make up the scale can be self-administered to tap an individual's perception of her own or other's behavior in a recalled or hypothetical interaction. The items could also be used by outside observers to rate an interaction. Reporters or raters respond to adjective pairs that reflect opposite ends of a continuum (e.g., *helpful–unhelpful*) and select a number between 1 and 7 to indicate their evaluation of the interaction for each criterion.

Scoring

There are four items for each of the three dimensions. The items associated with each dimension and instructions for scoring are provided at the end of this profile. Although developed as a multidimensional scale, a number of studies have adapted the MEESS, using some or all of the original items to create a unidimensional, global assessment of message quality (e.g., Bodie, 2013; Bodie, Burleson, & Jones, 2012; Ledbetter, 2008; Matsunaga, 2010; Murray, Derrick, Leder, & Holmes, 2008).

Development

The MEESS was developed in three stages, detailed in Goldsmith, McDermott, and Alexander (2000). They began by asking a sample of adults from the local community to respond to open-ended prompts about what the terms *helpful*, *sensitive*, and *supportive* meant to them. The authors used computerized cluster analysis supplemented by interpretive thematic analysis to develop a pool of 30 semantic-differential items. In a second study, a sample of students used these items to evaluate a recalled conversation about a problem, stress, or hassle. The authors used confirmatory factor analysis to select 12 final items and to test the superiority of a three-factor oblique model compared to a single-factor model. Finally, the authors asked another sample of students to use the 12-item scale to evaluate hypothetical advice messages that varied in form and content. Through confirmatory factor analysis, the authors provided evidence for the construct validity of the scale by fitting a three-factor model to the data: problem-solving utility, relational assurance, and emotional awareness.

Separating the interrelated aspects of support is useful for revealing different pathways through which support might have positive or negative effects. Many studies of support presume that the form of social support implies its function: Informational support confers benefits by improving problem-solving coping; and emotional support benefits us by soothing negative feelings. Explicitly measuring reactions to support and using a multidimensional measure to do so might reveal surprising pathways through which benefits (or harms) occur. Even advice that is useless for problem solving might be valued as an expression of support, for instance. Emotional support not only addresses feelings but also enables individuals to think more clearly about solutions

A multidimensional measure also recognizes that support attempts may receive mixed evaluations, beneficial in some respects and less so in others. Unidimensional measures designed to reflect a global assessment fail to differentiate between support that is mediocre and support that is exemplary in some respects but problematic in others (e.g., the blunt advice that hurts our feelings even as we grudgingly admit that it is exactly what we should do, or the inept attempt at comforting that we nonetheless appreciate for the effort and caring it expresses). It is precisely the *mixed* evaluation of these messages that could be important for coping and for relational outcomes.

Reliability

Alphas for the original three dimensions have been reported to range from .80 to .90 for problem-solving utility, .84 to .92 for relational assurance, and .78 to .94 for emotional awareness (Caughlin *et al*., 2008; Goldsmith *et al*., 2000; Thompson & O'Hair, 2008).

Validity

The original validation studies utilized confirmatory factor analysis to support the construct validity of the scales (Goldsmith *et al*., 2000). Subsequent studies provide additional evidence of predictive validity. For example, some studies have examined variation in message quality (e.g., facework: Fowler, Fisher, & Pitts, 2014; quality advice: Thompson & O'Hair, 2008; and message design logic: Caughlin *et al*., 2008) and have found that the message types that this theory would predict would be rated more positively did receive higher ratings on the MEESS. The MEESS also has successfully predicted desired outcomes of a conversation, such as elders' willingness to discuss future care needs with their children (Fowler *et al*., 2014), Japanese and American students' willingness to discuss bullying incidents with a family member (Matsunaga, 2010), and relational partners' willingness to seek or provide support (Murray *et al*., 2008).

Availability

The semantic differential items are provided at the end of this profile and are reprinted under fair use provisions. When using the scale to evaluate a recalled conversation, Goldsmith and colleagues (2000) asked several open-ended questions prior to the scale presentation in order to facilitate memory and promote focus on a specific conversation (e.g., "Briefly describe the problem you talked about," "What is your relationship to the person with whom you talked?", "Where did the conversation take place?", and "How long ago did the conversation occur?"). The scale is free to use for research purposes.

Sample Studies

The scales have been used to evaluate a range of types of conversations in a variety of contexts and with several different populations. For example, Caughlin and colleagues (2008) measured the evaluation of responses to a sibling's disclosure of an HIV diagnosis and found that messages exhibiting more sophisticated message design logics received more favorable MEESS ratings. Fowler and colleagues (2004) used the relational assurance and emotional sensitivity subscales to examine how adults older than 60 evaluated a hypothetical conversation in which an adult child initiated a conversation about the elder's future care needs. Conversations that included facework were evaluated more favorably on the MEESS scales than conversations that did not; in turn, MEESS ratings predicted willingness to discuss eldercare with one's own adult children. Thompson and O'Hair (2008) examined advice reported by 184 cancer survivors and found that MEESS ratings were more favorable when the recipient of advice was open to receiving advice and when the advice was optimistic and did not include limitations.

Critique

A primary concern with the MEESS derives from the intercorrelations among the three dimensions, which can be substantial. In the original validation study, interfactor correlations ranged from .62 to .77, but some subsequent studies report correlations of .90 or greater (e.g., for the relational assurance and emotional awareness scales in Fowler *et al.* [2014] and for all three scales in Matsunaga [2010]). Although the scales are highly correlated, Goldsmith and colleagues (2000) found that in two different student samples, a three-factor model provided a better fit to the data than a single-factor model.

It seems likely that the nature of the messages evaluated might influence the degree to which the factors emerge as distinct. In the validation studies, the authors intentionally included messages that varied in quality. In Study 2, half of the sample was instructed to recall a conversation when they "ended up being really happy about talking to that person," and the other half recalled a conversation when they "ended up being really sorry they talked with that person." In Study 3, participants rated advice messages that varied in content of advice and in regard for face. This produced messages that were rated positively in some regards and less so in others (e.g., the advice that has problem-solving utility but shows little emotional awareness or relational assurance). There is a conceptual basis for differentiating the ratings, even if some populations or situations tend to produce messages that are perceived as more uniformly good or bad.

References

Bodie, G. D. (2013). The role of thinking in the comforting process: An empirical test of a dual-process framework. *Communication Research*, *40*, 533–558. doi:10.1177/0093650211427030

Bodie, G. D., Burleson, B. R., & Jones, S. M. (2012). Explaining the relationships among supportive message quality, evaluations, and outcomes: A dual-process approach. *Communication Monographs*, *79*, 1–22. doi:10.1080/03637751.2011.646491

Caughlin, J. P., Brashers, D. E., Ramey, M. E., Kosenko, K. A., Donovan-Kicken, E., & Bute, J. J. (2008). The message design logics of responses to HIV disclosures. *Human Communication Research*, *34*, 655–684. doi:10.1111/j.1468-2958.2008.00336.x

Fowler, C., Fisher, C. L., & Pitts, M. J. (2014). Older adults' evaluations of middle-aged children's attempts to initiate discussion of care needs. *Health Communication*, *29*, 717–727. doi:10.1080/10410236.2013.786278

Goldsmith, D. J., McDermott, V. M., & Alexander, S. C. (2000). Helpful, supportive and sensitive: Measuring the evaluation of enacted social support in personal relationships. *Journal of Social and Personal Relationships*, *17*, 369–391. doi:10.1177/0265407500173004

Ledbetter, A. M. (2008). Chronemic cues and sex differences in relational e-mail perceiving immediacy and supportive message quality. *Social Science Computer Review*, *26*, 466–482. doi:10.1177/0894439308314812
Matsunaga, M. (2010). Testing a mediational model of bullied victims' evaluation of received support and post-bullying adaptation: A Japan-US cross-cultural comparison. *Communication Monographs*, *77*, 312–340. doi:10.1080/03637751003758235
Murray, S. L., Derrick, J. L., Leder, S., & Holmes, J. G. (2008). Balancing connectedness and self-protection goals in close relationships: A levels-of-processing perspective on risk regulation. *Journal of Personality and Social Psychology*, *94*, 429–459. doi:10.1037/0022-3514.94.3.429
Thompson, S., & O'Hair, H. D. (2008). Advice-giving and the management of uncertainty for cancer survivors. *Health Communication*, *23*, 340–348. doi:10.1080/10410230802229712

Scale

The Multidimensional Evaluation of Enacted Social Support (MEESS) (Goldsmith *et al.*, 2000)

Please circle the number that best represents how you feel about this interaction.

1)	sensitive	1--------2--------3--------4--------5--------6--------7	insensitive
2)	upsetting	1--------2--------3--------4--------5--------6--------7	reassuring
3)	useless	1--------2--------3--------4--------5--------6--------7	useful
4)	comforting	1--------2--------3--------4--------5--------6--------7	distressing
5)	encouraging	1--------2--------3--------4--------5--------6--------7	discouraging
6)	heartless	1--------2--------3--------4--------5--------6--------7	compassionate
7)	supportive	1--------2--------3--------4--------5--------6--------7	unsupportive
8)	helpful	1--------2--------3--------4--------5--------6--------7	hurtful
9)	ignorant	1--------2--------3--------4--------5--------6--------7	knowledgeable
10)	selfish	1--------2--------3--------4--------5--------6--------7	generous
11)	considerate	1--------2--------3--------4--------5--------6--------7	inconsiderate
12)	misunder-standing	1--------2--------3--------4--------5--------6--------7	understanding

Scoring instructions: The scale should be scored so that higher values correspond to more positive evaluations; consequently, items 1, 4, 5, 7, 8, and 11 are reverse scored. Items 1, 6, 11, and 12 are averaged to provide a measure of *sensitive* (emotional awareness). Items 2, 4, 5, and 7 are averaged to provide a measure of *supportive* (relational assurance). Items 3, 8, 9, and 10 are averaged to provide a measure of *helpful* (problem-solving utility).

Profile 43

Multitasking While Listening

Profiled by: Jonathon Frost, MA and Brock Bybee, MA

Louisiana State University and Agricultural & Mechanical College

Construct

Multitasking can be understood as attempting to simultaneously perform multiple routine tasks or multiple tasks that require higher level cognitive processing (Bratfisch & Hagman, 2003).

Instrument Type

Experimental Manipulation

Description

When people multitask, they are managing "multiple task goals in the same general time period by engaging in frequent switches between individual tasks" (Delbridge, 2000, p. 1). Task switches can occur very rapidly, as quickly as one tenth of a second, giving people the illusion that the tasks occur simultaneously (Meyer & Keiras, 1997). Research has shown, however, that humans are actually incapable of simultaneously managing two tasks at the same time (Meyer & Keiras, 1997).

Listening tasks are often used in multitasking studies in conjunction with written or visual tasks to access multitasking proficiency. Generally speaking, multitasking decreases proficiency at individual tasks (Bowman, Levine, Waite, & Gendron, 2010; Kahneman, 1973; Kushniryk & Levine, 2012; Meyer & Keiras, 1997; Oswald, Hambrick, & Jones, 2007). Due to differences in cognitive processing between aural and visual information (Paivio, 1986), however, presenting information in two different channels

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

can in fact increase memory retention (Kushniryk & Levine, 2012; Mantei, 2000; Weinraub, 1998), but only conditionally.

Administration

Multitasking studies most often involve the experimental manipulation of multiple tasks, asking participants to complete them simultaneously. In many studies, the delivery method of the task varies. Listening tasks are common, and may include listening to and being able to recall messages delivered by speakers, music, television programing, and movie clips (see, e.g., der Horst, Klehe & Leeuwen, 2012; Jeong & Hwang, 2012).

Other studies ask participants to report on their typical or daily tendencies to multitask or task switch. These studies either administer surveys at one point in time, or ask participants to keep daily logs of their activities (Czerwinski, Horvitz, & Wilhite, 2004; Sanbonmatsu, Strayer, Medeiros-Ward, & Watson, 2013). Such self-report studies tap tendencies to engage in multitasking or perceptions of multitasking ability rather than true ability to do so or the actual outcomes associated with such activity.

A third strain of research involves creating a cognitive map of brain activity during task switches using functional magnetic resonance imaging (fMRI) imagery (Dux, Ivanoff, Asplund, & Marois, 2006; Yin, Wang, Pan, Liu, & Chen, 2015; see Profile 21). These studies typically focus on the physiological processes and implications of task switching by monitoring brain activity while participants engage in activities that require task switching.

Scoring

Individuals are not technically scored on multitasking ability, but are more often given scores on constructs thought to be outcomes of multitasking. A common outcome is the average score of retention questions given after engaging in tasks (der Horst *et al.*, 2012), whereas others score participants separately for the written and listening components (Kushniryk & Levine, 2012). Close-ended and open-ended items also have been used, with scoring of responses including word counts and coding the accuracy of open-ended responses (Jeong & Hwang, 2012; Kushniryk & Levine, 2012). Finally, self-report questionnaires that employ multiple scaled items have been used (e.g., asking participants to rate their perceived comprehension from *strongly agree* to *strongly disagree*; Jeong & Hwang, 2012). Although it is possible for experimental manipulations of multitasking to be scored for their instantiation of the construct, most studies do not include manipulation checks.

Development

The term *multitasking* was developed in the computer engineering industry (Abate, 2008). Research about multitasking, however, can be traced back to William James (1890), who described differences between active attention processes and passive attention processes. To James, active attention was the default of the human brain, and things that divide attention were seen as distractions and obstructions, preventing what should be an effortless focus. James's line of reasoning was used in the development of the single-channel hypothesis, which was additionally influenced by Broadbent's (1958) filter model of attention. The single-channel hypothesis posits that people have a limited amount of cognitive processing ability and, as such, can only process a single stimulus at a time (Allport, Antonis, & Reynolds, 1972). The single-channel hypothesis was not ultimately supported, as it did not account for the parallel processing of two signals, and was thus replaced with the limited-capacity model of information processing (Kahneman, 1973; see also the Audio Message Complexity profile, Profile 8), which, although maintaining that humans have a finite ability to process information, allows for parallel processing.

Researchers eventually determined that multitasking or simultaneous processing is actually impossible, although the brain can switch tasks incredibly rapidly, in as quickly as one tenth of a second (Meyer & Keiras, 1997). More recent research, using fMRI testing to monitor brain activity while multitasking, revealed a neural "bottleneck" that delays the brain's ability to process nearly simultaneous tasks (Dux *et al.*, 2006; see fMRI profile, Profile 21). The bottleneck delay increases with task complexity, unfamiliarity, and the frequency of interruptions, indicating that the brain does in fact hinder our ability to multitask. Ongoing research still seeks to fully understand the cognitive processes involved with task switching (Yin *et al.*, 2015).

Listening became a primary emphasis of multitasking research with the introduction of Paivio's (1986) dual coding theory, which posits that the cognitive processes for interpreting visual and auditory information are independent yet interconnected. As such, the brain is capable of processing information from the two channels (aural and visual) easier than it would be able to process two pieces of information from the same channel (e.g., watching a video while reading). Since then, there have been numerous studies that have manipulated multitasking by designing measures to test for memory retention (Kushniryk & Levine, 2012; Mantei, 2000; Weinraub, 1998) and task performance (Bowman *et al.*, 2010; Kushniryk & Levine, 2012; Meyer & Keiras, 1997; Oswald *et al.*, 2007) after participants have engaged in simultaneous tasks.

Reliability

Multitasking involves doing two or more things, seemingly simultaneously. The only indicator of reliability thus relevant for multitasking experiments is whether a particular set of tasks used in a particular set of studies consistently produces variability in attentional switching (e.g., vis-à-vis brain activity). Because of the inherent difficulty of establishing this reliability criterion for a single study, it is no surprise that such a measure is never reported, at least in studies reviewed for this profile.

Take, for example, the study reported by der Horst *et al.* (2012) in which applicants for a call center position were asked "to look up information, listen to [a] voice message, and read and type all at the same time" (p. 436). Do these tasks consistently produce brain activity in line with how multitasking is conceptually defined? That question remains unanswered in that study. Instead, like other studies, the authors focused more on the outcomes of multitasking—in this case, the retention of information as assessed with postactivity questions.

Validity

Similar to reliability, manipulations of multitasking can be more or less valid indicators of the construct. Validity concerns for experimental multitasking studies include whether the tasks used involve (a) higher order cognitive capacity and (b) a need for simultaneous completion. Asking people to complete multiple tasks might also interfere with anxiety, for instance, which means any results of multitasking studies can equally be attributed to other constructs. In general, concerns about validity center on internal validity when the study is experimental (e.g., manipulating the presence of multitasking). If the study is interested in reported tendencies to multitask, then construct validity concerns are important as are psychometric properties of any scales used.

When experimental methods are employed, it is typical for participants to be asked to perform several tasks simultaneously. This decision seems valid insofar as the definition of multitasking is the attempted performance of simultaneous tasks. Of course, what tasks should be used is at least partially contextual. As noted, der Horst *et al.* (2012) were interested in the demands posed on potential call center agents and thus decided to simulate multitasking by having applicants listen to voice messages while looking up information on the computer and filling out forms. They argued that their study design simulated the qualitative and quantitative demands of a call center employee as identified by researchers. Their measure of how much information the applicants retained after the tasks, however, seems more akin to a measure of listening comprehension than multitasking, per se. So, although their manipulation of multitasking seems valid, the measure they employed may not actually tap multitasking ability but an outcome of that ability.

Overall, it seems more honest to say that measures of multitasking should be restricted to brain activity or other physiological indicators such as eye tracking that can actually provide information about the attention being allocated to tasks. Manipulations of multitasking like the performance of simultaneous tasks important for a job can be judged as valid only if they cause physiological activity to vary in ways characteristic of multitasking. Any other measures that are given after a multitasking simulation are then more accurately labeled as measures of some other construct—listening comprehension if the test is geared toward retention of orally presented information, or perhaps task difficulty if items read something like "I had a hard time concentrating on my tasks."

Availability

There are numerous manipulations available in the literature, as multitasking studies are quite diverse. However, there is no universal way to manipulate multitasking while listening. The interest could be with whether listening can be adequately accomplished while simultaneously completing other tasks like writing (Coens, Degryse, Senecaut, Cottyn, & Clarebout, 2011; der Horst *et al.*, 2012; Kushniryk & Levine 2012), or the study could attempt to explore what happens when people attempt to listen to multiple sources of information (Brungart, Simpson, Ericson, & Scott, 2001; Freyman, Balakrishnan, & Helfer, 2004). Readers are encouraged to explore the literature on multitasking to find an appropriate manipulation and/or measure given their unique research questions.

Sample Studies

The study by der Horst *et al.* (2012) illustrates a common format for multitasking studies. In their study, participants were asked to listen to an audio recording while simultaneously browsing the Internet for additional information to answer a set of questions. der Horst *et al.*'s (2012) study was designed to determine whether or not multitasking proficiency has an effect on job performance. The test was designed to mirror the multiple tasks required of a call center employee. They found that applicants who answered more multiple-choice questions after engaging in a multitasking exercise scored higher on various metrics of job performance related to customer contact; their scores also were correlated with performance-based job termination.

Kushniryk and Levine (2012) tested for both written task performance and listening memory retention. They instructed participants to answer open-ended questions while listening to a short 15-minute lecture. The lecture was presented in three different formats: video lecture, live lecturer, and live lecturer with a PowerPoint aid. The participants were instructed to write throughout the entire duration of the lecture, while simultaneously remembering the content from the lecture. Afterwards, participants were given a short quiz to test for information retention. A control group performed both tasks sequentially, rather than simultaneously. They found that participants' performance and retention both dropped when multitasking; retention increased, however, when information was presented verbally and visually (i.e., lecturer with PowerPoint), even if the participants were multitasking.

Jeong and Hwang (2012) sought to determine the effect of multitasking on persuasive messages. They divided participants into three groups and presented each with a persuasive message in written format. One group was presented only with the persuasive message, one group was presented with the written message and a video (a 4-minute movie clip) and asked to focus primarily on the written message, and a final group was given the written message and the video and asked to focus primarily on the video. Afterwards, participants were given a self-report and assessment to determine levels of comprehension and counterarguing. Results showed that multitasking decreased both comprehension and participant inclination to counterargue, which could have mixed implications for persuasive attempts. On one hand, lower comprehension could mitigate the effect of persuasion. On the other hand, decreased motivation to counterargue could increase the effectiveness of persuasion.

Recent research has highlighted the need to investigate the relations between multitasking ability and noncognitive processes (Sanderson, Bruk-Lee, Viswesvaran, Gutierrez, & Kantrowitz, 2016). Sanderson *et al.* (2016) employed a multitasking test used by actual businesses to screen employees, which involved answering questions on one screen, while receiving frequent email interruptions on a second screen that demanded attention. The researchers measured multitasking ability against emotional stability, conscientiousness, openness to experience, and cognitive ability. They failed to find a relation between multitasking and emotional stability or openness to experience, but they found a negative correlation between conscientiousness and multitasking, as well as a strong positive correlation between multitasking and cognitive ability.

Critique

Studies involving multitasking are limited by the fact that there is no precise way to manipulate the construct. Moreover, the various manipulations found in the literature were never checked against physiological standards that seem to define the construct operationally (e.g., variability in attention, and task switching as shown with fMRI scans). In one sense, the inconsistency of manipulation is logical—when studies are conducted to understand the impact of multitasking in particular contexts, the manipulation of multitasking should be applicable to those contexts. Because the manipulations differ, however, it is difficult to claim that the studies accurately reflect the same construct, making comparisons among studies challenging.

Despite differences among studies, there appears to be some consistency in the finding that multitasking decreases task performance on multiple levels, but there are exceptions (i.e., presenting information visually and aurally can increase memory retention).

References

Abate, C. J. (2008). You say multitasking like it's a good thing. *Thought & Action*, *24*, 7–13.

Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology*, *24*, 225–235. doi:10.1080/00335557243000102

Bowman, L. L., Levine, L. E., Waite, B. M., & Gendron, M. (2010). Can students really multitask? An experimental study of instant messaging while reading. *Computers & Education*, *54*, 927–931. doi:10.1016/j.compedu.2009.09.024

Bratfisch, O., & Hagman, E. (2003). *Simultankapazität/Multi-Tasking (SIMKAP) Version 24.00: Handanweisung* (Simultaneous capacity/multi-tasking (SIMKAP) release 24.00: Manual). Mödling, Austria: Schuhfried.

Broadbent, D. (1958). *Perception and communication*. London: Pergamon.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, *110*(5), 2527–2538.

Coens, J., Degryse, E., Senecaut, M., Cottyn, J., & Clarebout, G. (2011). Listening to an educational podcast while walking or jogging: Can students really multitask? *International Journal of Mobile and Blended Learning*, *3*, 23–33. doi:10.4018/jmbl.2011070102

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 175–182). New York: ACM.

Delbridge, K. A. (2000). *Individual differences in multi-tasking ability: Exploring a nomological network*. Unpublished doctoral dissertation. Michigan State University, East Lansing.

der Horst, A. C., Klehe, U., & Leeuwen, L. (2012). Doing it all at once: Multitasking as a predictor of call center agents' performance and performance-based dismissal. *International Journal of Selection & Assessment*, *20*, 434–441. doi:10.1111/ijsa.12006

Dux, P. E., Ivanoff, J. G., Asplund, C. L., & Marois, R. (2006). Isolation of a central bottleneck of information processing with time-resolved fMRI. *Neuron, 52,* 1109–1120. doi:10.1016/j.neuron.2006.11.009

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America, 115*(5), 2246–2256

James, W. (1890). *The principles of psychology*. New York: Henry Holt.

Jeong, S., & Hwang, Y. (2012). Does multitasking increase or decrease persuasion? Effects of multitasking on comprehension and counterarguing. *Journal of Communication, 62,* 571–587. doi:10.1111/j.1460-2466.2012.01659.x

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kushniryk, A., & Levine, K. J. (2012). Impact of multitasking on listening effectiveness in the learning environment. *The Canadian Journal for the Scholarship of Teaching and Learning, 3*(2), 1–11. doi:10.5206/cjsotl-rcacea.2012.2.7

Mantei, E. J. (2000). Using Internet class notes and PowerPoint in physical geology lecture: Comparing the success of the computer technology with traditional teaching techniques. *Journal of College Science Teaching, 29,* 301–305.

Meyer, D. E., & Keiras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review, 104,* 3–65. doi:10.1037/0033-295x.104.1.3

Oswald, F. L., Hambrick, D. Z., & Jones, L. A. (2007). Keeping all the plates spinning: Understanding and predicting multi-tasking performance. In D. H. Jonassen (Ed.), *Scientific problem solving: The state of the art* (pp. 77–97). Mahwah, NJ: Erlbaum.

Paivio, A. (1986). *Mental representations: a dual coding approach*. New York: Oxford University Press.

Sanbonmatsu, D. M., Strayer, D. L., Medeiros-Ward, N., & Watson, J. M. (2013). Who multi-tasks and why? multi-tasking ability, perceived multi-tasking ability, impulsivity, and sensation seeking. *Plos ONE, 8*(1), 1–8. doi:10.1371/journal.pone.0054402

Sanderson, K. R., Bruk-Lee, V., Viswesvaran, C., Gutierrez, S., & Kantrowitz, T. (2016). Investigating the nomological network of multitasking ability in a field sample. *Personality & Individual Differences, 9,* 152–157. doi:10.1016/j.paid.2015.11.013

Weinraub, H. J. (1998). Using multimedia authoring software: The effects on student learning perception and performance. *Financial Practice and Education, 8,* 88–92.

Yin, S., Wang, T., Pan, W., Liu, Y., & Chen, A. (2015). Task-switching cost and intrinsic functional connectivity in the human brain: Toward understanding individual differences in cognitive flexibility. *Plos ONE, 10*(12), 1–15. doi:10.1371/journal.pone.0145826

Profile 44

Narrative Believability Scale (NBS-12)

(Yale, 2013)

Profiled by: Graham D. Bodie, PhD

Meek School of Journalism and New Media, The University of Mississippi

Construct

Narrative believability is the extent to which a narrative seems "veridical, and thus acceptable for informing decisions" (Yale, 2013, p. 579).

Instrument Type

Self-Report

Description

The NBS-12 was developed to investigate the extent to which believable narratives are more influential on decision making. Overall, "a believable narrative is one that avoids leaving loose ends, is internally consistent and consistent with the perceiver's prior knowledge, and contains the expected elements and structure of a story" (Yale, 2013, p. 580). Drawing from the story model in the context of juror decision making, Yale (2013) asserted that two "certainty principles" determine story acceptability. *Coverage* is "the extent to which the story accounts for evidence presented at trial" (Pennington & Hastie, 1992, pp. 527–528). *Coherence* is the degree to which a story is consistent, plausible, and complete. These four constructs—coverage, consistency, plausibility, and completeness—are proposed as first-order latent constructs, each measured by three items.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

The NBS-12 can be administered online or using pencil and paper, and it is likely to take fewer than 5 minutes to complete. The scale is administered after participants are exposed to one or more narratives and thus acts as a measure of the degree to which participants (as an aggregate or separated into groups based on some other variable) perceive the narrative(s) to be believable. Each item is scaled along seven points bounded by *strongly disagree* (1) and *strongly agree* (7).

Scoring

The NBS-12 can generate five scores per participant: an overall believability score and a score for each of the four subscales. Items within subscales can be averaged, or the entire scale can be used (average across all 12 items) to generate a total believability score.

Development

Yale developed the NBS-12 over three studies. Study 1 involved generating an initial pool of items in the following manner. First, Yale wrote 8 items for each of the four constructs for a total of 32 items. Second, four trial attorneys and two litigation consultants familiar with the story model evaluated all items for face validity by first reviewing definitions for each construct, then categorizing each item into one of the four constructs. A final pool of 28 items was constructed after (a) removing items with low agreement and (b) adding items suggested by the experts; Yale also (c) wrote four additional items "by providing a short conceptual definition of [each] construct and asking for a rating of the narrative on that dimension" (p. 581). Third, these 28 items were administered to a group of 474 US undergraduate students who were exposed to one of five narratives that varied in manipulated believability. Fourth, each 7-item subscale was analyzed with ALPHAMAX, an item reduction technique that generates short forms of scales for use in public opinion research (Hayes, 2005). Fifth, the final 12 retained items were submitted to confirmatory factor analyses that compared a two- and four-factor model. The four-factor model was deemed the best fitting model, S–B $\chi 2 = (48,\ N = 474) = 104.14$, $p < .001$, RMSEA = .05 (90% CI: .037–.063), NNFI = .99, CFI = .99. These 12 items were then submitted to tests of validity in two subsequent studies (reviewed in this profile).

Reliability

Reported internal consistency estimates of the NBS-12 (Cronbach's alpha) were all within acceptable ranges: .88–.91 for the full scale, .81–.87 for Plausibility, .81–.87 for Completeness, .81–.82 for Consistency, and .72–.78 for Coverage.

Validity

Using responses from the same undergraduate sample that helped produce the 12-item scale, Yale (2013) also gathered evidence for criterion-related validity by "testing the ability of the subscale scores to differentiate between the master trial narrative and the four trial narratives that were manipulated to be low in plausibility, completeness, consistency, and coverage" (p. 585). Results suggested all subscales were able to discriminate among manipulated narratives. In addition, evidence for convergent validity was offered by showing that three of the NBS-12 subscales (not Completeness) were positively related to the Pinocchio circling task, a measure of narrative acceptance.

In a third study, Yale (2013) asked 269 undergraduates who were eligible to serve as US jurors to view web-based videos of plaintiff and defense statements, then complete the NBS-12. Participants also "rendered verdicts in the case and completed credibility measures for each of the attorneys they viewed" (Yale, 2013, p. 589). Evidence of predictive validity was weak. In several predictive models, Plausibility was the only consistent predictor of verdict decisions and verdict confidence. Although the subscales were associated with the dependent variables at the bivariate level, when added to an overall model their predictive power was subverted.

Availability

The NBS-12 was presented originally in Yale (2013) and is reproduced here, with permission. It is free to use for research purposes.

Sample Studies

Other than the studies published by Yale, the NBS-12 has not been used to measure narrative believability in any other published work.

Critique

The primary critique of the NBS-12 is that it has only been employed in three studies by one author. Although the measure seems to exhibit strong psychometric properties, it must continue to stand the test of time. Moreover, evidence for convergent, predictive, and criterion-related validity was weak; no evidence for discriminant validity was offered. Listening scholars are encouraged to think through how people process narratives that are more or less believable as well as what makes narratives seem veridical even if they contain false information.

References

Hayes, A. F. (2005). *A computational tool for survey shortening applicable to composite attitude, opinion, and personality measurement scales*. Paper presented at the Midwestern Association for Public Opinion Research, Chicago, IL.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology*, *62*, 189–206. doi:10.1037/0022-3514.62.2.189

Yale, R. N. (2013). Measuring narrative believability: Development and validation of the Narrative Believability Scale (NBS-12). *Journal of Communication*, *63*, 578–599. doi:10.1111/jcom.12035

Scale

The Narrative Believability Scale (Yale, 2013)

Source: Yale (2013). Reproduced with permission of John Wiley & Sons.

Please answer the following items with reference to the story you just heard.

Plausibility

1) I believe this story could be true.
2) This story was plausible.
3) This story seems to be true.

Completeness

4) It was easy to follow the story from beginning to end.
5) It was hard to follow this story.*
6) If I were writing this story, I would have organized it differently.*

Consistency

7) The information presented in this story was consistent.
8) All of the facts in this story agreed with each other.
9) The "consistency" of a story refers to the extent to which a story does not contradict itself or contradict other things you know to be true or false. How would you rate this story in terms of "consistency"?*

Coverage

10) There was important information missing from this story.*
11) There were lots of "holes" in this story.*
12) The "coverage" of a story refers to the extent to which the story accounts for all of the information presented in the story. How would you rate this story in terms of "coverage"?

Note: Labels should be removed and items randomized prior to administration. All items are scaled along seven points, but end points differ. Items 9 and 12 are scaled using Very Low (1) to Very High (7); all other items are scaled using Strongly Disagree (1) to Strongly Agree (7). Items marked with an asterisk should be reverse-scored prior to creating subscales or an overall score.

Profile 45

Narrative Engagement Measure (NEM)

(Busselle & Bilandzic, 2009)

Profiled by: Jenny L. Crowley, PhD[1] and Jennifer A. Jackl, PhD[2]

[1] *University of Tennessee*
[2] *Roanoke College*

Construct

The Narrative Engagement Measure (NEM) was developed to assist researchers in the quantification of how much, and in what ways, an individual engages with and processes narratives.

Instrument Type

Self-Report

Description

The goal of measuring narrative engagement is to better understand the ways in which audience members make sense of narratives and the ways narratives influence audience members' perceptions of the world. The 12-item NEM was created to measure four dimensions of narrative engagement. There are three items for each of the four subscales: narrative understanding, attentional focus, narrative presence, and emotional engagement (Busselle & Bilandzic, 2009).

Narrative understanding is conceptualized as the audience member's ability to take the perspective of characters in the narrative and how the audience member makes sense of the narrative. *Attentional focus* measures the extent to which the participant is focused on or distracted from the narrative. *Narrative presence* captures the extent to which the audience member is transported into the narrative world. *Emotional engagement* assesses the level of empathy and/or sympathy that viewers or readers

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

have for characters in the narrative. Researchers can use the full NEM or use the subscales independently, depending on the goals of the study.

Administration

After participants have watched/read/listened to a narrative, they are asked to fill out the 12-item NEM or, if the researcher wishes, particular subscales of the NEM. Responses may be recorded via an online survey or analog with pencil and paper. Administration time will vary with the length of narrative chosen and whether all or part of the scale is utilized.

Scoring

There are three items for each of the four dimensions of narrative engagement. Each of the 12 items is scaled along 7 points (1 = *completely disagree*; 7 = *completely agree*). Participants receive subscores for each of the four areas (i.e., narrative understanding, attentional focus, narrative presence, and emotional engagement) as well as an overall narrative engagement score. Items for the narrative understanding and attentional focus subscales are reverse-coded. The score for each subscale is computed by averaging the three items associated with the subscale, and an overall narrative engagement score is computed by calculating the mean for all 12 items.

Development

The NEM was created by Busselle and Bilandzic (2009) in order to achieve a more comprehensive way to quantify the factors that lead individuals to fully engage with a narrative. A key, underlying assumption of the measure is an individual's ability to construct mental models. A mental model is the individual's understanding of the world based on previous experiences, combined with information gained from engaging with a narrative. When audience members consume a narrative, they either become highly engaged or disengaged depending upon how the narrative matches up, or differs from, their current mental models of meaning.

Busselle and Bilandzic (2009) compiled 40 previous scale items involving empathy, sympathy, cognitive perspective taking, loss of time, loss of self-awareness, narrative presence, narrative involvement, distraction, ease of cognitive access, narrative realism, and narrative transportation. Next, Busselle and Bilandzic (2009) engaged in multiple rounds of exploratory and confirmatory factor analyses that revealed that the 40 items could be condensed into 12 items. These 12 items loaded onto the four main narrative engagement dimensions: narrative understanding, attentional focus, narrative presence, and emotional engagement.

Reliability

Reliability estimates for the entire scale have been good, with Cronbach's alpha typically exceeding .80 (e.g., Busselle & Bilandzic, 2009; Sukalla, Bilandzic, *et al.*, 2015). Reliability estimates of the subscales generally exceed .70; however, internal consistencies for the subscales have varied widely. For example, reliability estimates from the narrative understanding subscale have ranged from α = .58 (Busselle & Bilandzic, 2009) to α = .89 (Sukalla, Bilandzic, *et al.*, 2015), and the reliability estimates from the narrative presence subscale have ranged from α = .58 (Appel & Mara, 2013) to α = .91 (Sukalla, Shoenberger, *et al.*, 2015).

Validity

The NEM was developed to provide clarity between existing constructs and measures related to experiencing a narrative, such as narrative transportation (Green & Brock, 2000), identification (Cohen, 2001), and presence (Kim & Biocca, 1997). Over the course of three studies, each of which used a different sample, Busselle and Bilandzic (2009) used exploratory and confirmatory analyses to identify different dimensions of narrative engagement. In Study 1, exploratory factor analyses were conducted on 40 items from existing measurements and constructs, using both varimax and promax rotation (extraction method was not specified). The results of the exploratory factor analyses suggested a 20-variable, 4-factor solution. Study 2 employed confirmatory factor analysis, and goodness-of-fit tests suggested a modification of the measure, resulting in a 12-item scale. Finally, Study 3 confirmed the 12-item measure with data from an independent sample. The final NEM includes four subscales, each of which is predicted by a second-order narrative engagement factor, χ^2 (50) = 83.60; χ^2/df = 1.67; CFI = .963; RMSEA = .061 (90% CI = .037 - .084). Items loaded on factors that were unique from their original measures; thus, the NEM represents a distinct measure of narrative processing.

Previous studies suggest narrative engagement is associated with enjoyment (Busselle & Bilandzic, 2009), story-consistent attitudes (Busselle & Bilandzic, 2009), and media migration (Shade, Kornfield, & Oliver, 2015). In addition, media clips with higher negative emotional content are associated with higher levels of emotional engagement, attentional focus, and narrative presence; less cohesive stories are associated with lower narrative understanding, emotional engagement, and narrative presence (Sukalla, Bilandzic, *et al.*, 2015). In addition, self-reported narrative engagement is associated with psychophysiological measures of narrative processing, such as heart rate and corrugator muscle activity, which suggest increased mental effort and internal information processing (Sukalla, Bilandzic, *et al.*, 2015).

Availability

The English version of the NEM is provided at the end of this profile. Prior research has utilized a German version as well (Appel & Mara, 2013; Busselle & Bilandzic, 2009). Other versions of the NEM can be created by altering the wording of the scale to account for different narrative formats (e.g., "During reading …"). All items are displayed at the end of this profile with permission and are free to use for research purposes.

Sample Studies

Within the relatively short duration of the NEM's existence, scholars interested in narrative processing have applied the NEM to a variety of narrative formats, such as text-based stories (Appel & Mara, 2013), television programs (Sukalla, Shoenberger, *et al.*, 2015), and even video games (Sangalang, Quintero Johnson, & Ciancio, 2013). Narrative engagement has been associated with the likelihood of following entertainment across different media (e.g., watching a television program and then visiting that program's website). In addition, Sukalla, Shoenberger, and Bolls (2015) discovered that narrative engagement moderates the relation between a surprise event in a narrative (e.g., plot twist) and the allocation of mental resources for narrative processing.

Researchers also have discovered several associations between the individual dimensions of narrative engagement and the experiential and persuasive outcomes related to narratives. For example, narrative understanding significantly and positively predicted feeling transported by interactive games (Sangalang *et al.*, 2013). This finding suggests that the more an audience member engages with the perspective of the characters in a narrative, the more the audience member will feel immersed in the story environment and disconnected from their physical environment.

Appel and Mara (2013) discovered that narrative presence can have a significant effect on the persuasive outcomes of reading a story. Individuals low in narrative presence had stronger intentions to change behaviors when receiving information from a trustworthy source (rather than an untrustworthy source), whereas individuals high in narrative presence were equally persuaded to change behaviors irrespective of the trustworthiness of the source. This suggests that when listeners are able to fully enter the story world, they are less critical of the messages within the narrative. Finally, Bal and Veltkamp (2013) found that greater emotional engagement with a fictional story increased reader empathy, and longitudinal data suggest that these increases in empathy last for one week or longer. Thus, narrative engagement might contribute to an individual's development of emotional responses, such as empathy.

Critique

The NEM has consistently reached acceptable levels of reliability. Reliability estimates for the subscales have varied, although estimates typically reach acceptable levels as well. The NEM possesses several strengths. Research provides validity evidence for the scale, and the measure more clearly defines constructs that were previously confounded in the literature. Furthermore, we believe the recommendation by Busselle and Bilandzic (2009) to use subscales independently will productively enhance a more refined understanding of the four distinct constructs of narrative engagement. To date, however, use of the scale is limited, especially in listening contexts. Future work should continue to investigate the scale's applicability to listening contexts, as well as its utility for understanding narrative engagement with different narrative types (e.g., fantasy).

References

Appel, M., & Mara, M. (2013). The persuasive influence of a fictional character's trustworthiness. *Journal of Communication*, *63*, 912–932. doi:10.1111/jcom.12053

Bal, P. M., & Veltkamp, M. (2013). How does fiction reading influence empathy? An experimental investigation on the role of emotional transportation. *PLoS ONE*, *8*, e55341. doi:10.1371/journal.pone.0055341

Busselle, R., & Bilandzic, H. (2009). Measuring narrative engagement. *Media Psychology*, *12*, 321–347. doi:10.1080/15213260903287259
Cohen, J. (2001). Defining identification: A theoretical look at the identification of audiences with media characters. *Mass Communication & Society*, *4*, 245–264. doi:10.1207/S15327825MCS0403_01
Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*, 701–721. doi:10.1037/0022-3514.79.5.701
Kim, T., & Biocca, F. (1997). Telepresence via television: Two dimensions of telepresence may have different connections to memory and persuasion. *Journal of Computer-Mediated Communication*, *3*. doi:10.1111/j.1083-6101.1997.tb00073.x
Sangalang, A., Quintero Johnson, J. M., & Ciancio, K. E. (2013). Exploring audience involvement with an interactive narrative: Implications for incorporating transmedia storytelling into entertainment-education campaigns. *Critical Arts*, *27*, 127–146. doi:10.1080/02560046.2013.766977
Shade, D. D., Kornfield, S., & Oliver, M. B. (2015). The uses and gratifications of media migration: Investigating the activities, motivations, and predictors of migration behaviors originating in entertainment television. *Journal of Broadcasting & Electronic Media*, *59*, 318–341. doi:10.1080/08838151.2015.1029121
Sukalla, F., Bilandzic, H., Bolls, P. D., & Busselle, R. W. (2015). Embodiment of narrative engagement: Connecting self-reported narrative engagement to psychophysiological measures. *Journal of Media Psychology*, *27*, 1–12. doi:10.1027/1864-1105/a000153
Sukalla, F., Shoenberger, H., & Bolls, P. D. (2016). Surprise! An investigation of orienting responses to test assumptions of narrative processing. *Communication Research*, *43*, 844–862. doi:10.1177/0093650215596363

Scale

Narrative Engagement Measure (Busselle & Bilandzic, 2009)

Source: Busselle and Bilandzic (2009). Reproduced with permission of Taylor & Francis.

Instructions: Now that you have watched/listened to/interacted with the program, please carefully read and respond to the following 12 statements. Keep in mind, 1 indicates that you completely disagree, while 7 indicates you completely agree.

<u>Narrative Understanding</u>

1) At points, I had a hard time making sense of what was going on in the program.*
2) My understanding of the characters is unclear.*
3) I had a hard time recognizing the thread of the story.*

<u>Attentional Focus</u>

4) I found my mind wandering while the program was on.*
5) While the program was on I found myself thinking about other things.*
6) I had a hard time keeping my mind on the program.*

Narrative Presence

7) During the program, my body was in the room, but my mind was inside the world created by the story.
8) The program created a new world, and then that world suddenly disappeared when the program ended.
9) At times during the program, the story world was closer to me than the real world.

Emotional Engagement

10) The story affected me emotionally.
11) During the program, when a main character succeeded, I felt happy, and when they suffered in some way, I felt sad.
12) I felt sorry for some of the characters in the program.

Note: The instructions were not included with original scale but were crafted by the profile authors as a general guideline; they should be modified to fit the design of the research project. Labels should be removed and items randomized prior to administration. Items marked with an asterisk (*) should be reverse coded prior to scoring.

Profile 46

Weinstein Noise Sensitivity Scale (WNSS)

(Weinstein, 1978)

Profiled by: Debra L. Worthington, PhD

Auburn University

Construct

The Weinstein Noise Sensitivity Scale (WNSS) was designed to measure individual sensitivity to perceived noise.

Instrument Type

Self-Report

Description

Previous research has demonstrated that individuals respond differently to noise (see Hill, 2012). Noise sensitivity, as a personality trait, is considered an antecedent of noise annoyance. Weinstein (1978) developed the Noise Sensitivity Scale (WNSS) as a unidimensional, self-report measure of sensitivity to noise. The scale is composed of 21 items addressing affective reactions and attitudes to both general noise and daily environmental sounds. Individual sensitivity is seen as lying on a continuum ranging from high to low. Highly sensitive individuals are more sensitive to, and react more negatively to, perceived noise than their less sensitive counterparts (see also Highly Sensitive Person Scale, Profile 22).

Administration

The scale can be administered on paper or online and takes approximately 5 minutes to complete.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Scoring

Responses to each of the 21 items are scaled along 6 points with 0 indicating *strong disagreement* and 5 indicating *strong agreement*. After reverse coding relevant items, the unweighted sum of scores from each of the items is tallied. Stronger agreement with the items results in a higher score (ranging from 0 to 105), indicating greater individual noise sensitivity.

Development

The 21-item Weinstein Noise Sensitivity Scale (WNSS-21) was introduced in a longitudinal study of college student reactions to noise in a dormitory context (Weinstein, 1978). This investigation, a germinal study of individual noise sensitivity, is notable for the development of the first comprehensive self-report measure of noise sensitivity, the WNSS-21, as well as moving the study of noise sensitivity beyond transportation contexts. Weinstein noted that the WNSS-21 was designed to assess affective reactions to noise, while avoiding the role of noise as an environmental issue.

Weinstein (1978) explored individual differences to noise in a college dormitory context, examining students' initial reactions to noise along with their ability to adapt over time. He found that highly sensitive individuals had a greater need for privacy, were more critical of noise around them, and were more likely to express their dissatisfaction when annoyed. Less sensitive participants scored higher in a variety of personality traits, including extraversion, social desirability, sociability, and social presence. Although some items of the WNSS-21 appear specific to the college environment (e.g., studying), Weinstein contended that the items can be applied beyond this context.

Shorter versions of the WNSS-21 scale exist (e.g., the WNSS-10, NSS-10, NSS-6B, and WNS-SF). These scales typically exclude or modify context-specific items (Benfield *et al*., 2014; Kishikawa *et al*., 2006; Weinstein, 1980).

Reflecting its popularity, the WNSS-21 has been translated into several languages, including German (Zimmer & Ellermeier, 1999), Italian (Senese *et al*., 2012), Swedish (Ekehammar & Dornic, 1990), Persian (Alimohammadi, Nassiri, Azkhosh, Sabet, & Hosseini, 2006), and Serbian (Belojevic and Jakovljevic, 2001). Other noise sensitivity measures exist (e.g., NoiSeq: Schutte, Marks, Wenning, & Griefahn, 2007; and LEF: Zimmer & Ellermeier, 1998). These scales are longer (35 and 52 items, respectively) and broader in scope, measuring global noise sensitivity as well as sensitivity in a variety of daily contexts (e.g., leisure, work, sleep, etc.). A similar measure for sensitivity is presented in this volume as well, the Highly Sensitive Person Scale (HSPS; Profile 22).

Reliability

English and other language versions of the measure report similar reliability estimates. Weinstein (1978) reported acceptable reliability estimates using the Kuder Richardson reliability formula ($r = .83$). Other studies have reported reliability estimates ranging in the .70s (Stansfeld, 1992) but more frequently in the .80s (Ekehammar & Dornic, 1990; Zimmer & Ellermeier, 1999). Test–retest reliability has been reported as .87 over a 4-week interval (Zimmer & Ellermeier, 1997), .75 at 9 weeks (Weinstein, 1978), .69 at a 4-month interval (Stansfeld, 1992), and .63 over an 8-month interval (Weinstein, 1978).

Shorter versions of the scale also report acceptable reliability estimates (e.g., WNSS-10, Cronbach's alpha = .78; Stansfeld, Sharp, Gallacher, & Babisch, 1993; and WNS-SF, test–retest at 35 days = .80) (for an exception, see Worthington *et al.*, 2015).

Validity

Most studies support the unidimensionality of the scale using confirmatory procedures. Although some have found a highly correlated two-factor model, others have suggested dropping some items to improve the psychometric indices (Benfield *et al.*, 2014; Ekehammar & Dornic, 1990; Senese, 2012).

Job (1988) reported associations between noise sensitivity and subjective reactions to noise of .30 (i.e., annoyance), explaining 9% of the variance in reactions to noise. Following noise exposure, noise sensitivity is the second largest predictor of individual noise annoyance, with associations ranging from .25 to .45. It is also predictive of individual attitudes and behaviors toward specific sounds, such as personal electronic devices (Benfield *et al.*, 2014; Worthington *et al.*, 2015).

Senese *et al.* (2012) found support for the invariance of the construct across participant age, sex, and context (quiet vs. noisy), whereas Ekehammar and Dornic (1990) found it to be inversely correlated with self-rated stress tolerance (SST): SST-noise ($r = -.33$), SST-emotional ($r = -.18$), and SST-information overload ($r = -.20$).

Noise sensitivity has been associated with general trait-negative affect (Ohrstom *et al.*, 1989), trait anxiety (Zimmer & Ellermeier, 1998, 1999), and neuroticism (Stansfeld, 1992), as well as trait anxiety (Stansfeld, Sharp, Gallacher, & Babisch, 1993), depression, stress, and trait anger (Zimmer & Ellermeier, 1999). Higher sensitivity has been associated with a greater need for privacy (Weinstein, 1978; Worthington *et al.*, 2015) and a desire to experience quiet (Benfield *et al.*, 2014).

Availability

The English version of the WNSS-21 is readily available in several publications (Senese, Ruotolo, Ruggiero & Iachini, 2012; Weinstein, 1978), as are the various short forms of the measures (e.g., NSS-10, Kishikawa *et al.*, 2006; and WNS-SF, Benfield, 2014). The full version is presented at the end of this profile, with permission, and is free for use in research studies with appropriate citation.

Sample Studies

The WNSS-21 is frequently used to examine subjective reactions to transportation-related noise (e.g., aircraft, traffic, and railroad) (Marks & Griefahn, 2007; Ouis, 2002) and sleep disturbance (Aasvang, Moum, & Engdahl, 2008). Some studies have suggested that sensitivity can negatively affect performance in the workplace (Leathers *et al.*, 2003) and in educational settings (Arezes, Barbosa and Miguel, 2010; Weinstein, 1978), and add to the stress of hospitalization (Oreyzi, Fakhri, & Bahadorian, 2012). It also mediates customer responses to the acoustical soundscape of a restaurant (Novak, La Lopa, & Novak, 2010).

Relatively little research has addressed its potential contributions to listening and listening-related attitudes and behaviors. One exception is a study by Worthington *et al.* (2015), who reported that highly noise-sensitive individuals differ in their use of mobile phones and in their phone-related privacy-seeking behaviors. Based on self-report data, their findings suggest that highly noise-sensitive individuals were more likely to seek privacy when speaking with others on their mobile phone and are more likely to use alternative means of communication (e.g., texting, returning a call later, and removing themselves from a group) when utilizing a mobile phone in a social setting. In addition, using the WNSS-21, Kliuchko *et al.* (2015) investigated the relation between long-term musical training and noise sensitivity utilizing both Italian and Finnish subjects. They found that those who were more sensitive were less likely to play background music (i.e., passive listening to music) and rated music as less important in their daily life. Notably, there were no differences in responses between Italian and Finnish participants nor between responses of men and women.

Critique

The WNSS-21 has been identified as the most widely used measure of noise sensitivity (Miedema & Vos, 2003; Zimmer & Ellermeier, 1999). It was originally developed, however, for use with a student population, and the scale features items relevant to this population that may not be relevant for other populations. Although this limitation should be considered when moving to use the scale beyond academic settings, other studies (e.g., Senese *et al.*, 2012) have reported reliable scores and results in line with theoretical expectations when used with an adult sample and adapted to other contexts.

Questionnaire measures of noise sensitivity have been criticized for their lack to sensitivity to contextual factors (Smith, 2003). Schutte, Marks, Wenning, and Griefahn (2007) argued that noise sensitivity measures should be able to measure both global sensitivity as well as sensitivity in specific contexts (e.g., leisure, work, habitation, communication, and sleep). They subsequently developed their own scale.

Job (1999) suggested two possible intervening factors between annoyance and noise sensitivity—general negativity and noise itself. He also suggested that findings related to transportation noise might be confounded by self-selection. In other words, individuals who are highly sensitive may avoid moving into such areas when provided a choice.

Given its relation to anxiety and stress, noise sensitivity may be related to or contribute to communication-related anxiety (i.e., communication apprehension, receiver apprehension, and communicator style). Worthington *et al.* (2012) suggested noise sensitivity and interaction involvement (see Profiles 46 and 25, respectively) might be related.

References

Aasvang, G. M., Moum, T., &. Engdahl B. (2008). Self-reported sleep disturbances due to railway noise: Exposure-response relationships for nighttime equivalent and maximum noise levels. *Journal of the Acoustical Society of America*, *124*, 257–268. doi:10.1121/1.2932074

Alimohammadi, I., Nassiri, P., Azkhosh, M., Sabet, M., & Hosseini, M. (2006). Reliability and validity of the Persian translation of the Weinstein Noise Sensitivity Scale. *Psychological Research*, *9*, 74–87.

Arezes, P. M, Barbosa, M. S., & Miguel, A. S. (2010, August). Noise as a cognitive impairment factor: A case study amongst teachers. In *Proceedings of the International Congress on Acoustics* (pp. 2022–2026), Sydney, Australia.

Benfield, J. A., Nurse, G. A., Jakubowski, R., Gibson, A. W., Taff, B. D., Newman, P., & Bell. P.A. (2014). Testing noise in the field: A brief measure of individual noise sensitivity. *Environment and Behavior*, *46*, 353–372. doi:10.1177/0013916512454430

Ekehammar, B., & Dornic, S. (1990). Weinstein's noise sensitivity scale: Reliability and construct validity. *Perceptual and Motor Skills*, *70*, 129–130. doi:0.2466/pms.1990.70.1.129

Hill, E. M. (2012). *Noise sensitivity and diminished health: The role of stress-related factors*. Unpublished doctoral thesis. Auckland University of Technology, Auckland, New Zealand. Retrieved from http://hdl.handle.net/10292/4755

Job, R. F. S. (1988). Community response to noise: A review of factors influencing the relationship between noise exposure and reaction. *Journal of the Acoustical Society of America*, *83*, 991–1001. doi:10.1121/1.396524

Job, R. F. S. (1999). Noise sensitivity as a factor influencing human reaction to noise. *Noise & Health*, *1*, 57–68. PMID: 12689500

Kishikawa, H., Matsui, T., Uchiyama, I., Miyakawa, M., Hiramatsu, K., & Stansfeld, S. (2006). The development of Weinstein's noise sensitivity scale. *Noise and Health*, *8*, 154–160. doi:10.4103/1463-1741.50696

Kliuchko M., Heinonen-Guzejev, M., Monacis, L., Gold, B. P., Heikkilä, K. V., Spinosa V., Tervaniemi, M., & Brattico, E. (2015). The association of noise sensitivity with music listening, training, and aptitude. *Noise & Health*, *7*, 350–357. doi:10.4103/1463-1741.165065

Marks, A., & Griefahn, B. (2007). Associations between noise sensitivity and sleep, subjectively evaluated sleep quality, annoyance, and performance after exposure to nocturnal traffic noise. *Noise & Health*, *9*, 1–7. doi:10.4103/1463-1741.34698

Miedema, H. M. E., & Vos, H. (2003). Noise sensitivity and reactions to noise and other environmental conditions. *Journal of the Acoustical Society of America*, *113*, 1492–1504. doi:10.1121/1.1547437

Oreyzi, H. R., Fakhri, M., & Bahadorian, P. (2012). Noise psychological effect in Isfahan hospitals environment *Journal of Mazandaran University of Medical Sciences*, *21*, 190–191.

Ouis, D. (2002). Annoyance caused by exposure to road traffic noise: An update. *Noise & Health*, *4*, 69–79. PMID:12678952

Schutte, M., Marks, A., Wenning E., & Griefahn, B. (2007). The development of the noise sensitivity questionnaire. *Noise & Health*, *9*, 15–24. doi:10.4103/1463-1741.34700

Senese, V. P., Ruotolo, F., Ruggiero, G., & Iachini, T (2012). The Italian version of the Weinstein Noise Sensitivity scale: Measurement invariance across age, sex, and context. *European Journal of Psychological Assessment*, *28*, 118–124. doi:10.1027/1015-5759/a000099

Shepherd, D., Welch, D., Dirks, K. N., & Mathews, R. (2010). Exploring the relationship between noise sensitivity, annoyance and health-related quality of life in a sample of adults exposed to environmental noise. *International Journal of Environmental Research and Public Health*, *7*, 3579–3594. doi:10.3390/ijerph7103580

Smith, A. (2003). The concept of noise sensitivity: Implications for noise control. *Noise & Health*, *5*, 57–59. doi:10.1177/0013916586182004

Stansfeld, S. (1992). Noise, noise sensitivity and psychiatric disorder: Epidemiological and psychophysiological studies. *Psychological Medicine* (Monograph Suppl. 22), 1–44. doi:10.1017/S0264180100001119

Stansfeld, S. A., Sharp, D. S., Gallacher, J., & Babisch, W. (1993). Road traffic noise, noise sensitivity and psychological disorder. *Psychological Medicine*, *23*, 977–985. PMID: 8134521

Weinstein, N. D. (1978). Individual differences in reactions to noise: A longitudinal study in a college dormitory. *Journal of Applied Psychology*, *63*, 458–466. doi:10.1037/0021-9010.63.4.458

Weinstein, N. D. (1980). Individual differences in critical tendencies and noise annoyance. *Journal of Sound and Vibration*, *68*, 241–248. doi:10.1016/0022-460X(80)90468-X

Worthington, D. L., Fitch-Hauser, M., Välikoski, T.-R., Imhof, M., & Kim, S.-H. (2012). Listening and privacy management in cell phone conversations among young adults: A cross-cultural comparison of Finnish, German, Korean, and U.S. American students. *Empedocles: European Journal for the Philosophy of Communication*, *3*, 43–60. doi:10.1386/ejpc.3.1.43_1

Worthington, D. L., Keaton, S., Imhof, M., & Valikoski, T.-R. (2015). Impact of noise sensitivity on mobile phone attitudes and behaviors. *Mobile Media and Communication*, *4*, 3–18. doi:10.1177/2050157915581435

Zimmer, K., & Ellermeier, W. (1998). Construction and validation of the Noise Sensitivity Questionnaire. *Diagnostica*, *44*, 11–20.

Zimmer, K., & Ellermeier, W. (1999). Psychometric properties of four measures of noise sensitivity: A comparison. *Journal of Environmental Psychology*, *19*, 295–302. doi:10.1006/jevp.1999.0133

Scale

Weinstein's Noise Sensitivity Scale (WNSS-21)

Source: Weinstein (1978). Reproduced with permission of American Psychological Association.

<u>*Instructions*</u>: Below are a number of statements addressing individual reactions to noise. After reading each statement, please circle the number that best represents your level of agreement with the statement. For each item please use the following scale:

Strongly Disagree 1 2 3 4 5 Strongly Agree

1) I wouldn't mind living on a noisy street if the apartment I had was nice.
2) I am more aware of noise than I used to be.*
3) No one should mind much if someone turns up his or her stereo full blast once in a while.
4) At movies, whispering and crinkling candy wrappers disturb me.*
5) I am easily awakened by noise.*
6) If it's noisy where I'm studying, I try to close the door or window or move someplace else.*
7) I get annoyed when my neighbors are noisy.*
8) I get used to most noises without much difficulty.

9) It would matter to me if an apartment I was interested in renting were located across from a fire station.
10) Sometimes noises get on my nerves and get me irritated.*
11) Even music I normally like will bother me if I'm trying to concentrate.*
12) It wouldn't bother me to hear the sounds of everyday living from neighbors (footsteps, running water, etc.).
13) When I want to be alone, it disturbs me to hear outside noises.*
14) I'm good at concentrating no matter what is going on around me.
15) In a library, I don't mind if people carry on a conversation if they do it quietly.
16) There are often times when I want complete silence.*
17) Motorcycles ought to be required to have bigger mufflers.*
18) I find it hard to relax in a place that's noisy.*
19) I get mad at people who make noise that keeps me from falling asleep or getting work done.
20) I wouldn't mind living in an apartment with thin walls.
21) I am sensitive to noise.*

Note: Items should be randomized prior to administration. Item 3 was revised to be gender neutral. Item 9 was reworded to allow use of Likert response categories. Original wording: How much would it matter to you if an apartment you were interested in renting was located across from a fire station? Items marked with an asterisk (*) should be reverse coded prior to scoring. The asterisk should also be removed prior to administration.

Profile 47

Nonverbal Immediacy Measures

(Andersen, Andersen & Jensen, 1979; Richmond, McCroskey & Johnson, 2003)

Profiled by: Jane B. Teel, PhD

Auburn University

Construct

Immediacy is based on the belief that "people are drawn toward persons and things they like, evaluate highly, and prefer; and they avoid or move away from things they dislike, evaluate negatively, or do not prefer" (Mehrabian, 1971, p.1.) Nonverbal immediacy (NVI) is demonstrated through facial expressions, oculesics, kinesics, vocalic communication, proxemics, haptics, and chronemics.

A number of measures of NVI have been developed. This profile highlights several of the more common measures used in instructional settings.

Instrument Type

Self- and Other-Report; Behavioral Assessment

Description

Behavioral Indicants of Immediacy Scale (BIIS)

The BIIS was designed to assess an individual's evaluation of another's immediacy (Andersen & Andersen, 2005). Two versions of the BIIS were constructed to aid the measurement of the following nonverbal behaviors: distance, body orientation, forward lean, increases in touch, increases in eye contact, positive facial expressions, positive head nods, increases in gestures, bodily relaxation, use of purposeful body movements, spending time with other interactants, informal dress, orientation of head toward the other interactant, and vocal variations (Andersen *et al.*, 1979; Andersen & Andersen,

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

2005). The *instructional* version is a 28-item instrument based on the 15 NVI behaviors listed above (Andersen & Andersen, 2005). Participants are asked to evaluate their perception of their instructor as a teacher (Andersen *et al.*, 1979). The *interpersonal* version is a 22-item scale that measures the decoder's perception of their relationship with another person (Andersen *et al.*, 1979).

The Generalized Immediacy Scale (GIS)

The two versions of the GIS, the instructional and the interpersonal, were "developed to measure immediacy as a conceptual gestalt" (Andersen *et al.*, 1979, p. 157). Both versions contain nine semantic differential items. The instructional version measures teacher immediacy, and the interpersonal version measures the level of immediacy in a conversational setting.

Raters' Perception of Immediacy Scale (RIS)

The RIS is a third immediacy scale developed by Andersen *et al.* (1979). The 11-item scale was created for use by trained observers to rate the NVI of instructors in an instructional setting.

Nonverbal Immediacy Scale (NIS-S; NIS-O)

The NIS was developed by Richmond, McCroskey, and Johnson, (2003) to measure general tendencies toward enacting NVI behaviors. The NIS is a modified version of the BIIS and based on the limitations reported by Andersen *et al.* (1979). The two versions of the NIS are a self-report (NIS-S) and an observer-report (NIS-O) instrument. The NIS-S consists of statements related to NVI that are answered according to the level the respondent believes most accurately reflects his or her nonverbal communication behaviors. The NIS-O is used by an observer to evaluate a designated subject's general tendency to enact immediacy behaviors. Both instruments use 13 positively worded items and 13 negatively worded items; all items are scaled along five points with the end points *never* (1) and *very often* (5).

Administration

Although all of the measures were developed and originally delivered via paper and pencil, they could be adapted to online delivery (see, e.g., Andersen *et al.*, 1979).

For the BIIS (Instructional Context), participants respond to 28 (instructional) or 22 (interpersonal) statements using 7-point Likert scaling. The BIIS takes about 10 minutes to complete for either version.

For the GIS, participants respond to nine statements using semantic differential responses (e.g., *cold/warm* and *close/distant*). The GI takes fewer than 5 minutes to complete by an observer.

The RIS (16 items) was intended as a rating instrument and so involves intensive training of judges. Guidelines for training are found in Chapter 6 of this volume, and readers are additionally referred to Anderson's work and work by Susanne Jones (Jones, 2004; Jones & Guerrero, 2008) for examples of how to train and assess intercoder reliability.

For both the NIS-S and NIS-O, participants respond to 26 statements along five points: 1 = *never*, 2 = *rarely*, 3 = *occasionally*, 4 = *often*, and 5 = *very often*. Either version of the NIS will take approximately 10 minutes to complete.

Scoring

For the BIIS Instructional measure, the scores of designated items are entered into the following formula: X − Y + 56 = immediacy score; and, for the Interpersonal version: X − Y + 80 = immediacy score (Andersen *et al.*, 1979). The authors did not include a range for a high or low level of immediacy; however, they noted that higher scores are associated with higher levels of immediacy. Scores on the BIIS-Instructional ranged from a high of 98 to the lowest possible of 7. For the BIIS-Interpersonal, the highest possible score is 140, and the lowest possible is 20. The formula for both GI measures is X – Y + 40; higher scores reflect greater levels of immediacy (Andersen *et al.*, 1979).

Richmond *et al.* (2003) used a three-step scoring process. First, responses to the positively worded items are added together. Second, response to the negatively worded items are added together. Then, the following formula is applied:

78 + Sum of Positively Worded Items – Sum of Negatively Worded Items

According to the authors, high immediacy for females is greater than 112, and low immediacy for females is less than 92 (based on $M = 102$, $SD = 10.9$). High immediacy for males is greater than 104, and low immediacy for males is less than 83 (based on $M = 93.8$, $SD = 10.8$).

Development

Early communication research into immediacy behaviors and measures focused on the instructional context. Andersen (1978, 1979) simultaneously developed the BIIS, GIS, and RIS. Each measure has a different focus, but all reflect Andersen's interest in the relation between teacher immediacy and student learning (Andersen *et al.*, 1979; Richmond *et al.*, 2003). The three measures were used to rate nine male and four female instructors in 13 sections of an introductory communication course (Andersen *et al.*, 1979). Students enrolled in each course completed the BIIS and GIS, and three trained observers evaluated the same instructor's immediacy using the RIS (Andersen *et al.*, 1979).

The BIIS was designed to measure 15 NVI behaviors based on facial expressions, oculesics, proxemics, haptics, kinesics, vocalics, and chronemics in instructional and interpersonal contexts (Andersen & Andersen, 2005). The measure was developed with 28 statements for the instructional context and 22 statements for the interpersonal. The measure for instructional context was completed in two sessions by a total of 548 participants. The scale was reduced to 15 items following a principal component analysis (PCA) with varimax rotation (Andersen *et al.*, 1979).

When using the BIIS, students are instructed to compare the teacher being evaluated against their other teachers (Richmond *et al.*, 2003). Like the BIIS, the GIS also measures general immediacy in an instructional context and an interpersonal context. The GIS includes an opening paragraph explaining immediacy, then asks respondents to consider nine bipolar adjective pairs (e.g., *warm/cold* and *close/distant*) (Andersen, 1978).

According to Richmond and her colleagues, the GIS was based on McCroskey's unpublished Generalized Belief Scale. Results of a PCA of responses ($N = 548$) to the 9-item GIS yielded a single-component solution. The first component had an eigenvalue of 6.78, with remaining values below 1.00.

In contrast, the RIS was developed for use by trained observers and was created, in part, to provide validity evidence for the BIIS (Andersen *et al.*, 1979). The 16-item measure was also submitted to a PCA using data from 136 observations from three trained judges. Andersen *et al.* (1979) acknowledged that the initial analysis needed a greater number of observations to produce a stable factor structure. However, their findings suggested that the RIS was a unidimensional measure, with loadings from the unrotated solution ranging from .47 to .87 (Andersen *et al.*, 1979).

The NIS-S and NIS-O were based on previously used measures developed by Andersen (1978, 1979); Richmond, Gorham, and McCroskey (1987); McCroskey, Richmond, Sallinen, Fayer, and Barraclough (1995); and Richmond and McCroskey (2000). According to Richmond *et al.* (2003), earlier instruments often produced scores with low internal consistency. The NIS-S and NIS-O were developed to address this problem. The goal of both NIS measures was to balance the statements that indicated high immediacy with statements that indicated low immediacy. Richmond *et al.* (2003) argued that the content validity of the NIS is strong because of the 13 positively worded and 13 negatively worded components. An initial PCA with Promax rotation produced a single component for both the self-report and other-report instruments (Richmond *et al.*, 2003).

Reliability

Overall, reliability estimates of the NVI measures profiled here have been good. Estimates for the interpersonal version of the BIIS and the 9-item instructional version of the GIS have been reported in the .90s (Andersen, 1979; Andersen & Andersen, 2005; Andersen *et al.*, 1979; Andersen, Norton, & Nussbaum, 1981; Carrell & Menzel, 2001). Reported intercoder reliability of the RIS used by trained observers in an instructional setting has ranged from .79 to .97 (Andersen, 1979; Andersen *et al.*, 1979). Andersen *et al.* (1979) reported split-half reliability coefficients for the interpersonal version of the BIIS from .70 to .78 and a split-half internal reliability of .82 for the revised 11-item RIS.

According to Richmond *et al.* (2003), reliability estimates for the NIS-S and NIS-O were assessed through the completion of the instruments by 1241 participants; both were above .90. In other studies, internal consistency estimates of the NIS have ranged from .88 to .94 (Allen, Long, O'Mara, & Judd, 2008; Mottet, Beebe, Raffeld, & Paulsel, 2004; Santilli, Miller, and Katt, 2011). Alpha values for the NIS modified for use outside of the United States have been lower: Brazil ($\alpha = .89$), Japan ($\alpha = .79$), and Turkey ($\alpha = .81$) (Özmen, 2011; Santilli *et al.*, 2011). Özmen (2011) suggested the NIS is "culturally and cognitive[ly] more accessible" for US students (p. 871).

Validity

According to Andersen *et al.* (1979), content validity was provided for the BIIS by making sure each desired immediacy behavior was represented in an item. Andersen (1979) determined that predictive validity was evidenced because the scores on the BIIS are significant predictors of student affect toward the course and the instructor. Concurrent validity was demonstrated by examining the correlation between scores on the BIIS and RIS when students and raters evaluated the same instructors ($r = .92$) (Andersen *et al.*, 1979). However, other studies suggest there is little relationship between student reports of instructor NVI and other reports. Andersen (1978) found that reported NVI is not highly associated with student-rated NVI. Hess and Smythe (2001) found a slight correlation between perceived teacher NVI and cognitive learning ($r = .07$). Basically, research has focused on student perception of teacher's NVI rather than on actual teacher immediacy.

Andersen (1979) found GIS scores were more predictive of student behavioral commitment to an instructor than BIIS scores because the impression of nonverbal behaviors is probably more informative than actually analyzing individual behaviors (Andersen, 1979). In several studies, the GIS has predicted student affective learning, behavioral commitment, and positive attitude toward the instructor and the course (Andersen, 1979; Andersen *et al.*, 1979, Carrell & Menzel, 2001). Similarly, using an early version of the NIS, McCroskey, Sallinen, Fayer, Richmond, and Barraclough (1996) found increased instructor immediacy positively associated with perceived learning and negatively associated with learning loss. Notably, their cross-cultural study reported that this finding held true in samples from Australia, Finland, Puerto Rico, and the United States.

Based on the 13 different nonverbal components included in the NIS measures, the content validity of the NIS-S and NIS-O appears quite solid. In addition, the NIS has been predictive of students' affective learning, learning style, rate of course completion, and commitment to course (Mottet, Parker-Raley, Cunningham, & Beebe, 2005; Singletary, 2013; Teel, 2011).

Availability

All of the scales discussed in this measurement profile are presented here, with permission. The BIIS, GIS, and RIS appear in the *Communication Research* article by Andersen *et al.* (1979). The NIS-S and NIS-O are currently available at no charge at http://www.jamescmccroskey.com/measures/.

Sample Studies

Studies utilizing measures of NVI have covered a broad range of contexts, two of which are reviewed in this section. The most extensive use of the NVI scales is within instructional communication, with most work exploring the relations between student perceptions of instructor NVI and learning. Summarizing this work in meta-analytic form, Witt, Wheeless, and Allen (2004) ($k = 81$, $N = 24{,}474$) found moderately sized average associations between student perceptions of teacher NVI and student reports of perceived learning ($r = .51$) and affective learning ($r = .49$). The association between teachers' NVI and students' performance on cognitive learning measures was, however, much lower ($r = .17$).

Much of the work on teacher NVI asks students to report their impressions, begging the question of whether certain individual differences might be related to these reports. For instance, Allen *et al.* (2008) administered the NIS-S and NIS-O to determine if students' communication avoidance and sociocommunicative orientation were related to their perceptions of instructors' immediacy. They found that students who were high in communication avoidance viewed instructors as less nonverbally immediate, less assertive, and less responsive. They also reported that students who view themselves as higher in NVI view teachers as higher in NVI, assertiveness, and responsiveness.

Other work has explored the relation between perceptions of teacher immediacy and other relevant perceptions. For instance, Mottet *et al.* (2005) investigated student perceptions of teacher NVI using the NIS-O. The purpose of the study was to determine if the level of teacher NVI influenced student expectations for class workload and teacher accessibility. Results revealed that as student perception of teacher NVI increased, so did students' willingness to complete higher course workload demands. Given the increasing number of online offerings at US colleges and universities, it is worth noting that Carrell and Menzel (2001) found that perceived instructor immediacy was higher in a live setting compared to a video classroom or audio classroom with a PowerPoint display. Although short-term learning was highest in the PowerPoint classroom, no significant differences were reported in cognitive learning style based on the classroom setting. Teel (2011) examined the relation between undergraduate students' NVI behaviors as measured by the NIS-S and cognitive learning style preferences as measured by the Gregorc Style Delineator. She reported a slight relation between the NIS-S scores and the abstract sequential (AS) learning style ($r = .25$), which is in line with the finding of Witt *et al.* (2004).

The second area of work on NVI is within the supportive communication context. Researchers have long speculated that supportive listening can be readily operationalized with immediacy cues (Andersen & Andersen, 2005; Fassaert, van Dulmen, Schellevis, & Bensing, 2007), although only a few studies have investigated this claim. One study reported by Bodie and Jones (2012) asked participants to watch one of several videotaped supportive conversations. One person in each conversation was trained to enact low, moderate, or high levels of NVI. After watching their randomly assigned conversation, participants judged how active listeners were on the Active-Empathic Listening Scale (AELS) and the Active Listening Observation Scale (ALOS) (see Profiles 2 and 4, respectively). Although results showed that ratings of active listening were a linear function of NVI, that result was dependent on the scale (results were significant only for the ALOS). Moreover, the effect size for the NVI–ALOS relation was small (.01). Bodie, Vickery, Cannava, and Jones (2014) replicated this result in a study that asked participants to disclose a personal problem to either a trained or untrained active listener. Coded NVI behaviors, again, influenced perceptions of listeners as competent, but the effect size was small in magnitude (.05).

Critique

There appears to be some disagreement among researchers as to which measures of NVI are most likely to capture the construct. Andersen and Andersen (2005) rated the BIIS, GIS, and RIS as "the most reliable and valid measures of perceived nonverbal immediacy available" (p. 119). Several others have identified the BIIS as one of the most reliable of nonverbal measures (Finn & Schrodt, 2012; Houser & Frymier, 2009; Witt *et al.*, 2004). Richmond *et al.* (2003), however, disagree, arguing that the elimination of 13 of the 28 original items of the BIIS may have increased reliability estimates, but doing so reduced its validity. Similarly, McCroskey *et al.* (1996) suggested that having students compare their teacher's behavior to that of another teacher is problematic. Ultimately, whichever measure is chosen, researchers should first test and verify the dimensionality and reliability of the chosen instrument.

The biggest issue within the instructional context is the reliance on student perceptions of teacher immediacy rather than on ratings of actual teacher–student interactions. The meta-analysis presented by Witt *et al.* (2004) showed that perceptions of teacher immediacy are much more highly related to other self-report measures than to objective tests of learning. Such a result questions not only the convergent validity of self-report measures of immediacy but also the practical guidance our research can provide to instructors. Based on the research to date, all we can really say to teachers is "Get your students to think you are immediate," to which most teachers would ask, "How?" Although scale items provide some insight into how teachers might influence student perceptions of their immediacy, the studies that actually code behavior do not always show strong predictive power for individual nonverbal behaviors. The same is true in the supportive communication context, where coded NVI is not strongly related to perceptions of individuals as active listeners.

Finally, immediacy is clearly multidimensional and this profile focuses on nonverbal immediacy. Measures of verbal immediacy are receiving increasing attention (see, e.g., Wilson & Locker, 2007–2008).

References

Allen, J. L., Long, K. M., O'Mara, J., & Judd, B. B. (2008). Students' predispositions and orientation toward communication and perceptions of instructor reciprocity and learning. *Communication Education*, *57*, 20–40. doi:10.1080/03634520701670908

Andersen, J. F. (1978). *The relationship between teacher and teaching effectiveness*. Unpublished doctoral dissertation. West Virginia University, Morgantown, WV.

Andersen, J. F. (1979). Teacher immediacy as a predictor of teaching effectiveness. *Communication Yearbook*, *3*, 543–559.

Andersen, J. F., Andersen, P. A., & Jensen, A. D. (1979). The measurement of nonverbal immediacy. *Journal of Applied Communication Research*, *7*, 153–180. doi:10.1080/00909887909365204

Andersen, J. F., Norton, R. W., & Nussbaum, J. F. (1981). Three investigations exploring relationships between perceived teacher communication behaviors and student learning. *Communication Education*, *30*, 377–393. doi:10.1080/03634528109378493

Andersen, P. A., & Andersen J. F. (2005). Measurements of perceived nonverbal immediacy. In V. Manusov (Ed.), *The sourcebook of nonverbal measures: Going beyond words* (pp. 113–126). London: Routledge.

Bodie, G. D., & Jones, S. M. (2012). The nature of supportive listening II: The role of verbal person centeredness and nonverbal immediacy. *Western Journal of Communication*, *76*, 250–269. doi:10.1080/10570314.2011.651255

Bodie, G. D., Vickery, A. J., Cannava, K., & Jones, S. M. (2014). The role of "Active Listening" in informal helping conversations: Impact on perceptions of listener helpfulness, sensitivity, and supportiveness and discloser emotional improvement. *Western Journal of Communication*, *79*, 151–173. doi:10.1080/10570314.2014.943429

Carrell, L. J., & Menzel, K. E. (2001). Variations in learning, motivation, and perceived immediacy between live and distance education classrooms. *Communication Education*, *50*, 230–240. doi:10.1080/03634520109379250

Fassaert, T., van Dulmen, S., Schellevis, F., & Bensing, J. (2007). Active listening in medical consultations: Development of the active listening observation scale (ALOS-global). *Patient Education and Counseling*, *68*, 258–264. doi:10.1016/j.pec.2007.06.011

Finn, A. N., & Schrodt, P. (2012). Students' perceived understanding mediates the effects of teacher clarity and nonverbal immediacy on learner empowerment. *Communication Education*, *61*, 111–130. doi:10.1080/03634523.2012.656669.

Houser, M. L., & Frymier, A. B. (2009). The role of student characteristics and teacher behaviors in students' learner empowerment. *Communication Education*, *58*, 35–53. doi:10.1080/03634520802237383

Jones, S. (2004). Putting the person into person-centered and immediate emotional support: Emotional change and perceived helper competence as outcomes of comforting in helping situations. *Communication Research*, *31*, 338–360. Sdoi:10.1177/0093650204263436

Jones, S., & Guerrero, L. K. (2008). The effects of nonverbal immediacy and verbal person centeredness in the emotional support process. *Human Communication Research*, *27*, 567–596. doi:10.1111/j.1468-2958.2001.tb00793.x

McCroskey, J. C., Richmond, V. P., Sallinen, A., Fayer, J. M., & Barraclough, R. A. (1995). A cross-cultural and multi-behavioral analysis of the relationship between nonverbal immediacy and teacher evaluation. *Communication Education*, *44*, 281–291. doi:10.1080/03634529509379019

McCroskey, J. C., Sallinen, A., Fayer, J. M., Richmond, V. P., & Barraclough, R. A. (1996). Nonverbal immediacy and cognitive learning: A cross-cultural investigation. *Communication Education*, *45*, 200–211. doi:10.1080/03634529609379049

Mehrabian, A. (1971). *Silent messages*. Belmont, CA: Wadsworth.

Mottet, T. P., Beebe, S. A., Raffeld, P. C., & Paulsel, M. L. (2004). The effects of student verbal and nonverbal responsiveness on teachers' liking of students and willingness to comply with student requests. *Communication Quarterly*, *52*, 27–38. doi:10.1080/01463370409370176

Mottet, T. P., Parker-Raley, J., Cunningham, C., & Beebe, S. A. (2005). The relationships between teacher nonverbal immediacy and student course workload and teacher availability expectations. *Communication Research Reports*, *22*, 275–282. doi:10.1080/000368105000317482

Özmen, K. S. (2011). Perception of nonverbal and effective teaching among student teachers: A study across cultural extremes. *International Online Journal of Educational Sciences*, *3*, 865–881.

Richmond, V. P., Gorham, J. S., & McCroskey, J. C. (1987). The relationship between selected immediacy behaviors and cognitive learning. *Communication Yearbook*, *10*, 574–590.

Richmond, V. P., & McCroskey, J. C. (2000). The impact of supervisor and subordinate immediacy on relational and organizational outcomes. *Communication Monographs*, *67*, 85–95. doi:10.1080/03637750009376496

Richmond, V. P., McCroskey, J. C., & Johnson, A. D. (2003). Development of the nonverbal Immediacy Scale (NIS): Measures of self- and other-perceived nonverbal immediacy. *Communication Quarterly*, *51*, 504–517. doi:10.1080/01463370309370170

Santilli, V., Miller, A. N., & Katt, J. (2011). A comparison of the relationship between instructor nonverbal immediacy and teacher credibility in Brazilian and U. S. classrooms. *Communication Research Reports*, *28*, 266–274. doi:10.1080/08824096.2011.588583

Singletary, J. R. (2013.) *Teacher nonverbal immediacy: A study of its effect on student academic progress and end of course test performance in a rural alternative high school*. PhD dissertation, Liberty University. Retrieved from http://digitalcommons.liberty.edu/doctoral/719/

Teel, J. B. (2011). *An examination of the relationship between cognitive learning style preference and nonverbal immediacy behaviors in undergraduate students*. Unpublished doctoral dissertation, Auburn University, Auburn, AL. Retrieved from http://hdl.handle.net/10415/2671

Wilson, J. H., & Locker, L. (2007–2008). Immediacy scale represents four factors: Nonverbal and verbal components predict student outcomes. *Journal of Classroom Interaction*, *42*, 4–10. Retrieved from http://www.jstor.org/stable/23869788

Witt, P. L., Wheeless, L. R., & Allen, M. (2004). A meta-analytical review of the relationship between teacher immediacy and student learning. *Communication Monographs*, *71*, 184–207. doi:10.1080/036452042000228054

Scales

Nonverbal Immediacy Scale-Self Report (NIS-S)* (Richmond, McCroskey, & Johnson, 2003)

Source: Richmond *et al.* (2003). Reproduced with permission of Taylor & Francis.

DIRECTIONS: The following statements describe the ways some people behave while talking with or to others. Please indicate the degree to which you believe the statement applies **TO YOU**.

Please use the following 5-point scale:
1 = Never; 2 = Rarely; 3 = Occasionally; 4 = Often; 5 = Very Often

1) I use my hands and arms to gesture while talking to people.
2) I touch others on the shoulder or arm while talking to them.
3) I use a monotone or dull voice while talking to people.
4) I look over or away from others while talking to them.
5) I move away from others when they touch me while we are talking.
6) I have a relaxed body position when I talk to people.
7) I frown while talking to people.
8) I avoid eye contact while talking to people.
9) I have a tense body position while talking to people.
10) I sit close or stand close to people while talking with them.
11) My voice is monotonous or dull when I talk to people.
12) I use a variety of vocal expressions when I talk to people.
13) I gesture when I talk to people.
14) I am animated when I talk to people.
15) I have a bland facial expression when I talk to people.

16) I move closer to people when I talk to them.
17) I look directly at people while talking to them.
18) I am stiff when I talk to people.
19) I have a lot of vocal variety when I talk to people.
20) I avoid gesturing while I am talking to people.
21) I lean toward people when I talk to them.
22) I maintain eye contact with people when I talk to them.
23) I try not to sit or stand close to people when I talk with them.
24) I lean away from people when I talk to them.
25) I smile when I talk to people.
26) I avoid touching people when I talk to them.

Scoring:

Step 1. Add the scores from the following items: 1, 2, 6, 10, 12, 13, 14, 16, 17, 19, 21, 22, & 25.

Step 2. Add the scores from the following items: 3, 4, 5, 7, 8, 9, 11, 15, 18, 20, 23, 24, & 26.

Total Score = 78 + (Step 1) – (Step 2)

Norms:

Females	Mean = 102.0	S.D. = 10.9	High = >112	Low = <92
Males	Mean = 93.8	S.D. = 10.8	High = >104	Low <83

Note: Items should be randomized prior to administration. To create the **Nonverbal Immediacy Scale-Observer Report (NIS-O)** "I" is replaced with either she or he. The NIS-O is available online at: http://www.jamescmccroskey.com/measures/.

Norms for Other-Report version:

Females	Mean = 96.7	S.D. = 16.1	High = >112	Low = <81
Males	Mean = 91.6	S.D. = 15.0	High = >106	Low = <77
Combined	Mean = 94.2	S.D. = 15.6	High = >109	Low = <79

Behavioral Indicants of Immediacy (BII) Scale: Instructional Context (Andersen *et al.*, 1979)

DIRECTIONS: Please mark these scales to indicate how you perceive your instructor in the teaching role. Please mark the following statements to indicate whether you: (7) strongly agree; (6) agree; (5) moderately agree; (4) are undecided; (3) moderately disagree; (2) disagree; (1) strongly disagree. There is no correct answer. Simply record your perceptions. Some of the questions may seem similar, but this is necessary.

1) This instructor engages in more eye contact with me when teaching most other instructors.*
2) Students discuss less in this class than in most other classes.
3) This instructor has a more tense body position while teaching than most other instructors.*

4) This instructor gestures more while teaching than most other instructors.*
5) This instructor engages in less movement while teaching than most other instructors.*
6) This instructor sits in a student desk less than most other instructor when teaching.
7) This instructor touches students less than most other instructors when teaching.
8) This instructor has a more relaxed body position while teaching than most other instructors.*
9) This instructor directs his/her body position more toward students while teaching than most other instructors.*
10) This instructor stands in front of the classroom less than most other instructors while teaching.
11) This instructor smiles more during class than most other instructors.*
12) This instructor dresses less informally than most other instructors when teaching.
13) This instructor engages in less eye contact with me when teaching than most.*
14) This instructor spends less time with students before and after class than most instructors.
15) This instructor touches students more than most other instructors when teaching.
16) Students discuss more in this class than in most other classes.
17) This instructor is more vocally expressive while teaching than most other instructors.*
18) This instructor is more distant from students while teaching than most other instructors.*
19) This instructor directs his/her body position less toward students while teaching them than most other instructors.*
20) This instructor gestures less while teaching than most other instructors.*
21) This instructor engages in more movement while teaching than most other instructors.*
22) This instructor sits in a student desk more often than most other instructors while teaching.
23) This instructor dresses more informally than most other instructors when teaching.
24) This instructor stands in front of the classroom more than most other instructors while teaching.
25) This instructor is less vocally expressive while teaching than most other instructors.*
26) This instructor smiles less during class than most other instructors.*
27) This instructor is less distant from students than most other instructors while teaching.
28) This instructor spends more time with students before and after class than most other instructors.

Scoring Instructions:

To obtain an immediacy score, use this formula:

1) Total the subject's response for the following scale items: 1, 4, 8, 9, 11, 17, 21. Call this X.
2) Total the subject's response for the following scale items: 3, 5, 13, 18, 19, 20, 25, 26. Call this Y.
3) Immediacy score = X – Y + 56.

Note: Items should be randomized prior to administration. Items marked with an asterisk (*) constitute the 15-item behavioral indicants of immediacy scale.

Generalized Immediacy (GI) Scale: Instructional Context (Andersen *et al.*, 1979)

Immediate behaviors are those communication behaviors that reduce distance between people. Immediate behaviors may actually decrease the physical distance, or they may decrease the psychological distance. The more immediate a person is, the more likely he/she is to communicate at close distances, smile, engage in eye contact, use direct body orientations, use overall body movement and gestures, touch others, relax and be vocally expressive. In other words, we might say that an immediate person is perceived as overtly friendly and warm.

DIRECTIONS: Please place an "X" in *each* of the following scales to indicate your agreement with the following statement:

In your opinion, the teaching style of your instructor is very immediate.

Agree ____:____:____:____:____:____:____ Disagree
False ____:____:____:____:____:____:____ True
Incorrect ____:____:____:____:____:____:____ Correct
Wrong ____:____:____:____:____:____:____ Right
Yes ____:____:____:____:____:____:____ No

Please place an "X" in *each* of the following scales to indicate the word that best describes the teaching style of your instructor:

Immediate ____:____:____:____:____:____:____ Not Immediate
Cold ____:____:____:____:____:____:____ Warm
Unfriendly ____:____:____:____:____:____:____ Friendly
Close ____:____:____:____:____:____:____ Distant

Scoring Instructions:

1) Number each subject's response by numbering each scale from left to right (1-7).
2) Total the subject's response for the following scales: false/true, wrong/right, cold/warm, and unfriendly/friendly. Call this X.
3) Total the subject's response for the other five scales. Call this Y.
4) Generalized immediacy score = X – Y + 40.

Behavioral Indicants of Immediacy (BII) Scale: Interpersonal Context (Andersen *et al.*, 1979)

DIRECTIONS: Please complete the following scales to indicate how you see the relationship between you and the other person. Please mark the following statements to indicate whether you:

(7) strongly agree
(6) agree
(5) moderately agree
(4) are undecided
(3) moderately disagree
(2) disagree
(1) strongly disagree.

There is no correct answer. Simply record your perceptions. Some of the questions may seem similar, but it is necessary.

1) This person engages in more eye contact with me than most other people.
2) This person's body is more tense than most other people.
3) This person gesture more than most other people.
4) This person engages in less movement than most other people.
5) This person teaches me less than most other people usually do.
6) This person has a more relaxed body position than most other people.
7) This person directs his/her body position more toward me than most other people usually do.
8) This person smiles more than most other people do.
9) This person dress more formally than most other people do.*
10) This person engages in less eye contact with me than most other people.
11) This person seems eager to spend time talking with me.
12) This person touches me more than most other people.
13) This person is more vocally expressive than most other people.
14) This person seems more distant from me than most other people.
15) This person directs his/her body position less toward me than most.
16) This person gestures less than most other people.
17) This person engages in more movement than most other people.
18) This person dresses more informally than most other people.*
19) This person is less vocally expressive than most other people.
20) This person smiles less than most other people.
21) This person seemed less distant from me than most other people.
22) This person seemed reluctant to spend time talking to me.

Scoring Instructions:

1) Total the subject's response for the following scale items: 1, 3, 6, 7, 8, 11, 12, 13, 17, 21. Call this X.
2) Total the subject's response for the following scale items: 2, 4, 5, 10, 14, 15, 16, 19, 20, 22. Call this Y.
3) Immediacy Score = X – Y + 80.

Note: Items should be randomized prior to administration. Items marked with an asterisk (*) were dropped from the scale because of failure to load above 40.

Generalized Immediacy (GI) Scale: Interpersonal Context (Andersen *et al.*, 1979)

Immediate behaviors are those communication behaviors that reduce distance between people. Immediate behaviors may actually decrease the physical distance, or they decrease the psychological distance. The more immediate a person is, the more likely they are to communicate at close distance, smile, engage in eye contact, use direct body orientations, use overall body movement and gestures, touch others, relax, and be vocally expressive. In other words, we might say that an immediate person is perceived as overtly friendly and warm.

Is, in your opinion, the conversational style of the other person very immediate?

DIRECTIONS: Please place an "X" in *each* of the following scales to indicate your agreement with the above statement.

Agree ____:____:____:____:____:____:____ Disagree
False ____:____:____:____:____:____:____ True
Incorrect ____:____:____:____:____:____:____ Correct
Wrong ____:____:____:____:____:____:____ Right
Yes ____:____:____:____:____:____:____ No

Please place an "X" in *each* of the following scales to indicate the word that best describes the conversational style of the other person:

Immediate ____:____:____:____:____:____:____ Not immediate
Cold ____:____:____:____:____:____:____ Warm
Unfriendly ____:____:____:____:____:____:____ Friendly
Close ____:____:____:____:____:____:____ Distant

Scoring Instructions:

1) Number each subject's response by numbering each scale from left to right (1-7).
2) Total the subject's response for the following scales: false/true, wrong/right, cold/warm, and unfriendly/friendly. Call this X.
3) Total the subject's response for the other five scales. Call this Y.
4) Generalized immediacy score = X – Y + 40.

Profile 48

Normative Message Processing Scale (NMPS)

(Aune & Reynolds, 1994)

Profiled by: R. Kelly Aune, PhD[1] and Rodney A. Reynolds, PhD[2]

[1] *University of Hawai'i at Mānoa*
[2] *California Lutheran University*

Construct

The Normative Message Processing Scale was designed to assess the extent to which a person prefers to process messages in a more analytical or intuitive fashion.

Instrument Type

Self-Report

Description

The Normative Message Processing Scale (NMPS) was developed to test hypotheses concerning the role of effort in message processing. Extant scales such as the Need for Cognition Scale (NCS) are effort-based approaches to assess how much a person enjoys or is inclined to think effortfully (Cacioppo, Petty, & Morris, 1983). Effort, it is argued, is a proxy for a host of systematic differences in information and message processing. The degree of cognitive effort applied to an information- or message-processing task may be indicative of qualitatively different processing strategies (e.g., "dual" processes; see Bodie & Eldredge, in press) as well as quantitative differences in expended effort.

Administration

The NMPS consists of 26 items scaled along 7 points ranging from 1 = *strongly disagree* to 7 = *strongly agree*. Fifteen items assess the preference for processing messages in an

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Analytical (A) manner, and nine items assess the preference for processing messages in a more Intuitive (I) fashion. The items should be randomized prior to administration either using pencil-and-paper surveys or online software.

Scoring

Several items for each subscale are reverse-coded. Items for each scale are then averaged to create an Analytical scale and an Intuitive scale.

Development

The NMPS was developed by Aune and Reynolds (1994) partially in response to a perceived need to elaborate on the role that cognitive effort plays in message processing. Self-reports of greater expenditure of cognitive effort had been reliably associated with more systematic processing (e.g., Cacioppo & Petty, 1982, 1984). The implicit assumption that lower expenditures of cognitive effort suggest diminished processing, however, are not necessarily valid. Respondents who claimed they expended little cognitive effort when processing information might have been lacking focus or attention to the task, but they also might have been employing a processing style that does not require greater expenditures of cognitive effort. The NMPS was developed to test hypotheses that self-reports of cognitive effort could predict qualitative as well as quantitative differences in message processing.

Aune and Reynolds (1994) developed an initial set of items that were created to reflect

> the tendency to rely on hunches and intuition (to tap the experience of a meaningful gestalt), to employ deliberate analysis and evaluation (as is characteristic of a High Need for Cognition individual), preference for a particular message form and structure (to assess the preference for messages that are amenable to analytical decoding or processing meaningful gestalts), indication of memory processes and retrieval processes (since it was assumed that message decoding and assessment would be related to subsequent ability to recall message content), the tendency to easily process multiple message codes simultaneously (characteristic of Andersen's 1991 concept of intuitive communication). (p. 141)

This initial version of the NMPS consisted of 40 items.

More than 400 students completed the initial version of the NMPS, and these scores were then subjected to a principal axis factor analysis with varimax rotation (Aune & Reynolds, 1994). The factor analysis produced a 2-factor solution with items loading in a manner consistent with the conceptual framework of the scale. Items that loaded on the 2 factors were retained for a second round of testing, with additional items written to reinforce the conceptual content of the 2 factors.

The second study tested the 30-item revised version of the NMPS. Seventeen items assessed more systematic processing, and 13 items were intended to assess more intuitive processing (Aune & Reynolds, 1994). More than 200 respondents completed the NMPS in the second study that employed confirmatory factor analysis (CFA) and simultaneously measured theoretically relevant constructs for convergent and discriminant validity testing.[1] Results of the CFA reinforced the 2-factor solution, with the final version of the NMPS retaining 15 Analysis items and nine Intuitive items.

Reliability

Alpha reliabilities of the final version of the NMPS were reported as .87 for the Analytical Scale and .79 for the Intuitive Scale (Aune & Reynolds, 1994). During the validity testing of the scales, Aune and Reynolds found comparable reliability estimates (Analysis $\alpha = .86$, Intuition $\alpha = .78$). Reynolds (1997) employed an abbreviated version of the NMPS, using four items from each subscale that had the highest factor loadings. He reported alpha reliabilities for the Analysis subscale as .74 and for the Intuition subscale .74. Bodie, Worthington, and Gearhart (2013) used revised versions of the NMPS subscales, using only eight of the Analysis items and four of the Intuition items (after a CFA suggested model fit was poor because of low loadings), and found reliabilities of .84 and .66, respectively.

Validity

Aune and Reynolds (1994) provided evidence of construct validity through CFA. The NMPS measurement model was not, however, supported by data collected by Bodie *et al.* (2013). In their study, several items did not load highly on their respective latent factors and had high standardized residual values. A 12-item version of the scale (8 items for Analytical and 4 items for Intuitive) was supported by the following fit statistics: $\chi2$ (53) =148.58, $p < .001$, CFI = .90, SRMR = .07, RMSEA = .05 (CI 90% = .06–.10).

Aune and Reynolds (1994) also engaged in two stages of validity testing for the two subscales of the NMPS. Evidence for convergent validity was gathered by comparing respondents' scores on the subscales of the NMPS to their scores on the Human Information Processing Survey (HIP; Taggart & Torrance, 1984), the Need for Cognition Scale (NCS; Cacioppo & Petty, 1982), the Dogmatism scale (Troldahl & Powell, 1965), and subscales of Gundersen and Perrill's (1989) adaptation of Zahn and Hopper's (1985) Speech Evaluation Instrument. Evidence for divergent validity was gathered by comparing the subscales of the NMPS to Crowne and Marlowe's (1964) Social Desirability Scale and Burgoon's (1976) Unwillingness to Communicate Scale.

As seen in Table P48.1, correlations were in line with expectations. The Analytical subscale was negatively correlated with the Intuition subscale and positively correlated with the NCS and the Left Hemisphere subscale of the HIP. The Intuition subscale was negatively correlated with the NCS, negatively correlated with the Left Hemisphere subscale of the HIP, and positively correlated with the Right Hemisphere subscale of the HIP. There were also unexpected correlations between Social Desirability and the NMPS (both scales), the NCS, and the Left Hemisphere subscale of the HIP. The authors suggested that these findings could be related to the multiethnic, multicultural sample

1 It should be noted that this analysis followed guidelines of Hunter and Gerbing (1982), which didn't employ currently used goodness-of-fit tests but examined correlations matrices for evidence of deviations from unidimensionality, specifically tests for homogeneity, internal consistency, parallelism, and reliability.

Table P48.1 Correlations providing evidence of validity

	Analytical	Intuition	NCS	Left Hem.	Right Hem.
Analytical		−.35***	.30***	.50***	.01
Intuition			−.17*	−.24***	.43***
NCS				.12	−.02
Left Hem.					−.01
Right Hem.	.01	.43***	−.02	−.01	
UC Fearful	−.03	−.04	−.21**	.04	−.15*
UC reward	−.07	−.09	−.05	.07	−.09
Dogmatism	.32***	−.15*	−.08	.43***	.07
L. Compet	.16*	.01	.21**	.14*	.18**
L. Status	.07	−.07	.21**	.14*	.14*
L. Dynamism	.14	.04	.15*	.05	.16*
L. Attract	.08	.02	−.03	.19**	.20**

Note: All correlations after partialing out effects of social desirability. $N = 237$; * $p < .05$; ** $p < .01$; *** $p < .001$.
Source: Aune and Reynolds (1994). Reproduced with permission of National Communication Association.

used in the study, a significantly different sample than had been used in earlier tests of the NCS. A follow-up review of the study results found that Caucasians responded in a manner consistent with respondents in Cacioppo and Petty's (1982) earlier tests of the NCS. Significant relations between social desirability and the scales in question were, however, discovered for certain Asian and Pacific Islander respondents. Subsequent analyses were run controlling for social desirability scores, and all correlations in Table P48.1 are partial correlations controlling for scores on the SD scale.

The final validity tests consisted of predictive validity assessments. In the third study (Aune & Reynolds, 1994), almost 200 respondents completed the NMPS (Analytical $\alpha = .84$; Intuition $\alpha = .83$). Within 1–3 weeks, these respondents were asked to read several paragraphs advocating a position on tuition increases. After reading the paragraphs, respondents' self-reports of expended cognitive efforts were recorded, and respondents were asked to "please list as many of the message arguments as you can remember." Results were mostly consistent with expectations. Small but significant correlations were found between reported expended effort and the Analytical scale, $r = .18$, but not the Intuition scale. Recall of arguments was correlated in the predicted manner with both scales. Analytical scores were positively associated with recall, $r = .25$; Intuition scores were negatively associated with recall, $r = -.27$.

Additional predictive validity evidence was gathered by showing that the two NMPS subscales could predict respondents' ability to recognize emotional expressions (Aune & Reynolds, 1994). Respondents completed the NMPS, and within 3 weeks of completing the NMPS they also completed the Facial Meaning Sensitivity Test (FMST; Leathers, 1992). The latter test employs 30 pictures of a woman's face in each of which she is expressing one of 10 different affective states (disgust, happiness, interest, sadness, bewilderment, contempt, surprise, anger, determination, and fear). Respondents were asked to select the 3 pictures that depicted each of the 10 affective states. They were given an abbreviated time in which to do so. Results were consistent with expectations.

The Intuition subscale was positively correlated with the number of correct scores on the FMST, $r = .28$, and the Analytical scale was negatively correlated with the number of correct scores on the FMS, $r = -.35$. All correlations can be found in Table P48.1.

Availability

The full Analytical and Intuitive subscales of the NMPS as well as the complete account of the development of the NMPS can be found in Aune and Reynolds (1994). All items are presented below, with permission, and are free to use for research purposes with appropriate citation.

Sample Studies

The NMPS was developed to test hypotheses about how individuals may be processing messages when they report extending little cognitive effort. It is not a surprise, then, that the scale and its conceptual basis are referenced and often employed in validation testing of other instruments. These include scales developed to examine how individuals elaborate on messages (Reynolds, 1997), process media messages (Schemer, Matthes, & Wirth, 2008), support various listening goals (Bodie *et al.*, 2013), and consider themselves competent listeners (Fontana, Cohen, & Wolvin, 2015).

Reynolds (1997) used an abbreviated version of the NMPS in the validation process of his Message Elaboration Instrument and found that respondents' scores on the instrument were positively correlated with the Analysis subscale but not correlated with the Intuition subscale. Using a subset of the items based on a CFA, Bodie *et al.* (2013) showed the Analysis subscale correlating positively with Relational Listening, Analytical Listening, and Critical Listening. The Intuition subscale also correlated positively with Relational Listening and produced a small, negative correlation with Analytical Listening.

Critique

The most useful discussion of the conceptual basis and properties of the NMPS can be found in Radler's (2000) examination of the limitations of the NCS. Radler attempted to further Aune and Reynolds's (1994) argument that the NCS, being an effort-focused scale, contributes little toward understanding how exactly low-NCS persons process messages. Radler accurately pointed out that Aune and Reynolds's (1994) more intuitive processors may conflate different forms of low effort processing: rule-based heuristic systems and associative systems. These forms of processing are substantially different, yet the phenomenological experience can be similar (i.e., respondents may describe the experience of either form of processing as low effort and intuition- or hunch-based). Radler replicated and extended a study by Hayes and Broadbent (1988) and produced results that suggest high-NCS persons engage in cognitive activity that is rule-based and systematic, whereas low-NCS persons seem to be relying on a more associative, intuitive system. Although this does not invalidate or challenge the value of the NMPS, it may offer some conceptual clarity regarding the Intuitive subscale. It suggests the

Intuition subscale may be assessing the inclination to process messages by relying on associative capacities as opposed to heuristic understandings.

This does point to the general concern for the genre of information- and message-processing scales such as the NMPS (see also the Rational-Experiential Inventory, Profile 53). These scales almost uniformly attempt to assess cognitive phenomena that are variably accessible to consciousness, and attempt to do so with self-report items that get at those cognitive phenomena only indirectly. Given that a wealth of information processing occurs below levels of conscious awareness, it stands to reason that attempts to assess (in valid and reliable ways) those forms of processing most inaccessible to awareness will always be unsatisfyingly indirect and produce attenuated effects at best.

References

Andersen, P. A. (1991). When one cannot not communicate: A challenge to Motley's traditional communication postulates. *Communication Theory*, *3*, 73–82. doi:10.1080/10510979109368346

Aune, R. K., & Reynolds, R. A. (1994). The empirical development of the normative message processing scale. *Communication Monographs*, *61*, 135–160. doi:10.1080/03637759409376329

Bodie, G. D., & Eldredge, J. (2016). Dual processing perspectives. In C. Berger & M. Roloff (Eds.), *International encyclopedia of interpersonal communication*. Hoboken, NJ: John Wiley & Sons. [Online First]. doi:10.1002/9781118540190.wbeic0044

Bodie, G. D., Worthington, D. L., & Gearhart, C. C. (2013). The Listening Styles Profile-Revised (LSP-R): A scale revision and evidence for validity. *Communication Quarterly*, *61*, 72–90. doi:10.1080/01463373.2012.720343

Burgoon, J. K. (1976). The unwillingness to communicate scale: Development and validation. *Communication Monographs*, *43*, 60–69. doi:10.1080/03637757609375916

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–131. doi:10.1037/0022-3514.42.1.116

Cacioppo, J. T., & Petty, R. E. (1984). The need for cognition: Relationship to attitudinal processes. In R. P. McGlynn, J. E. Maddux, C. D. Stoltenberg, & J. H. Harvey (Eds.), *Social perception in clinical and counseling psychology* (Vol. *2*, pp. 113–140). Lubbock: Texas Tech University Press.

Cacioppo, J. T., Petty, R. E., & Morris, K. J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology*, *45*, 805–818. doi:10.1037/0022-3514.45.4.805

Crowne, D., & Marlowe, D. (1964). *The approval motive*. New York: Wiley.

Fontana, P. C., Cohen, S. D., & Wolvin, A. D. (2015). Understanding listening competency: A systematic review of research scales. *The International Journal of Listening*, *29*, 148–176. doi:10.1080/10904018.2015.1015226

Gundersen, D. F., & Perrill, N. K. (1989). Extending the "Speech Evaluation Instrument" to public speaking settings. *Journal of Language and Social Psychology*, *8*, 59–61. doi:10.1177/0261927X8900800105

Hayes, N. A., & Broadbent, D. E. (1988). Two modes of learning for interactive tasks. *Cognition*, *28*, 249–276. doi:10.1016/0010-0277(88)90015-7

Radler, B. (2000). *Elucidating need for cognition using a procedural learning paradigm*. Unpublished doctoral dissertation. University of Wisconsin, Madison.

Reynolds, R. A. (1997). A validation test of a message elaboration measure. *Communication Research Reports*, *14*, 269–278. doi:10.1080/08824099709388670

Schemer, C., Matthes, J., & Wirth, W. (2008). Toward improving the validity and reliability of media information processing measures in surveys. *Communication Methods and Measures*, *2*, 193–225. doi:10.1080/19312450802310474

Taggart, W., & Torrance, E. P. (1984). *Human information processing survey: Administrator's manual*. Bensenville, IL: Scholastic Testing Service.

Troldahl, V. C., & Powell, F. A. (1965). A short-form dogmatism scale for use in field studies. *Social Forces*, *44*, 211–214. doi:10.1093/sf/44.2.211

Zahn, C. J., & Hopper, R. (1985). Measuring language attitudes: The speech evaluation instrument. *Journal of Language and Social Psychology*, *4*, 113–123. doi:10.1177/0261927X8500400203

Scale

Normative Message Processing Scale (NMPS) (Aune & Reynolds, 1994)

<u>Analysis Subscale</u>

1) Objectivity and analysis are not my primary tools for assessing persuasive messages.*
2) After making a decision about someone's argument, I usually know the thought processes that led to my decision.
3) The best way for me to assess a person's argument is through careful analysis.
4) I analyze each point of a message one at a time and very carefully.
5) When developing a message, I don't think very much about the order of the specific points of the message.*
6) I don't need to completely understand a message to know if it makes sense.*
7) When I read or listen to a message I pay close attention to each point that is made and decide whether it is a good point or not.
8) When I'm listening to an explanation about something, I stop everything else so that I can pay close attention to what is being said.
9) My best decisions about a message come from careful analysis and reflection about the content of the message.
10) It takes me a while to understand an argument because I carefully think about each point presented.
11) When assessing the validity of an argument, I rank each point in order of importance and then consider whether it makes sense.
12) When assessing a persuasive argument I try to remain objective and analyze the content of the message.
13) When I listen to a speaker I concentrate on the content of the message and don't let myself get distracted by anything else.
14) I'm not very careful or deliberate when I'm listening to a message.*
15) I assess a person's argument by evaluating each point, one at a time.

<u>Intuition Subscale</u>

16) I know when a message makes sense because it just seems to feel right.
17) My intuition plays only a weak role in my analysis of a person's message.*
18) Hunches and intuitions are not my primary tools for assessing persuasive messages.*
19) I don't like to rely on my hunches about the validity of people's arguments.*
20) When assessing the validity of a person's argument I rely a lot on my feelings and intuitions.
21) I don't usually have hunches or intuitions about messages.*
22) I don't usually go with my first impressions when making an important decision; I prefer to take my time.*
23) Having a good hunch is often as useful as developing a good understanding.
24) I don't always know what leads me to believe or reject an argument; it just happens.

Note: Labels should be removed and items randomized prior to administration. Items marked with an asterisk (*) should be reverse coded prior to scoring.

Profile 49

Ordinary Conversation Scale (OCS)

(Lakey, Vander Molen, Fles, & Andrews, 2016)

Profiled by: Brian Lakey, PhD and Travis Sain, MA

Grand Valley State University
Southern Illinois University

Construct

The Ordinary Conversation Scale (OCS) was designed to assess the perceived quality of conversations about ordinary topics.

Instrument Type

Self-Report

Description

The OCS is an 8-item, self-report questionnaire. Each item asks the respondent to rate aspects of the quality of conversation with a specific relationship partner (e.g., friend, spouse, or parent).

Administration

The OCS is self-administered and can be completed in fewer than 2 minutes. It can be easily modified for use with any type of relationship. All items are rated using 5-point Likert scaling.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Scoring

Scoring consists of taking the average across the eight items. There are no subscales.

Development

The OCS was developed to assess the perceived quality of ordinary conversations, a key construct in Relational Regulation Theory (RRT; Lakey & Orehek, 2011). People with high perceived support have better mental health on a wide range of outcomes, including low levels of normal distress, lower rates of diagnosed mental disorder, and greater happiness (Cohen & Wills, 1985; Lakey & Orehek, 2011). Most social support theory is intended to explain stress-buffering effects whereby social support protects people from the adverse effects of stress. In contrast, main effects occur when people with high perceived support have better emotional well-being regardless of the presence of stress—even in its absence. RRT is designed to explain main effects between perceived support and emotional well-being. According to RRT, the main effect reflects the regulation of affect, action, and thought through ordinary conversation and shared activities (e.g., TV, movies, and sport). Thus, RRT does not rely upon stress and coping mechanisms.

Eight items were selected from the original 20 by identifying those that loaded most highly on the first principal component in each of three samples, using varimax rotation (Lakey, Vander Molen, Fles, & Andrews, 2016). Two of the samples were composed of college students, and one was composed of US Marines.

Reliability

The median internal consistency reliability for the scale in five samples was .89, with a range of .80 to .93 (Lakey *et al.*, 2016; Woods, Lakey, & Sain, 2016).

Validity

If ordinary conversation can explain the main effect between perceived support and emotional well-being, then the OCS should show findings very similar to those observed with perceived support measures. As expected, OCS scores are highly correlated with perceived support scores (median $r = .69$, range $= .54 - .79$). As is perceived support, OCS scores are strongly correlated with high positive affect (median $r = .61$, range $= .54 - .77$), as well as significantly linked to low negative affect (median $r = -.35$, range $= -.01$ to $-.60$), few automatic negative thoughts ($r = -.47$), and provider similarity (median $r = .62$, range $= .58 - .65$). That is, when a support provider is rated by the recipient as evoking unusually good ordinary conversation, the provider (a) is seen as unusually supportive and similar to the recipient and (b) evokes high positive affect, low negative affect, and few automatic negative thoughts (Lakey *et al.*, 2016; Woods *et al.*, 2016). Evidence for the discriminant validity of the scale is provided by nonsignificant correlations with conflict, $r = -.10$ (Lakey *et al.*, 2016), and incremental validity beyond received support in predicting high positive ($\beta = .45$, $p < .05$) and low negative affect ($\beta = -.39$, $p < .05$; Woods *et al.*, 2016).

Availability

The OCS is provided at the end of this profile with permission and is free to use for research purposes with appropriate citation.

Sample Studies

RRT defines relational as in the social relations model (Kenny, 1994). Relationship effects occur when a support recipient sees a provider as more supportive, for example, than (a) how the recipient typically sees providers (actor or recipient effects) and (b) how the provider is typically seen by others (partner or provider effects). Thus, relationship effects reflect the extent to which a judgment is idiosyncratic to a specific recipient, in the same way that tastes in art are often idiosyncratic to perceivers. One way to isolate relationship effects is to use a round-robin design in which groups of people each rate each other (Kenny, 1994).

Lakey *et al.* (2016, Study 1) used a round-robin design to isolate relationship effects. Groups of four Marines who trained and worked as a team rated each other before deploying to Afghanistan. Both perceived support and ordinary conversation were strongly relational, accounting for 54% and 67% of the variance, respectively. Moreover, relational ordinary conversation was significantly linked to high positive affect, $r = .54$, and low negative affect, $r = -.35$. That is, when a provider evoked unusually good ordinary conversation in a recipient, the provider also evoked unusually high positive and low negative affect. According to RRT, ordinary conversation can explain most of the main effect between perceived support and emotional well-being. If so, then when ordinary conversation is controlled, most of the link between perceived support and positive and low negative affect should disappear. In fact, when ordinary conversation was controlled in multiple regression analyses, perceived support's correlation with positive affect dropped from accounting for 29% of the variance to accounting for 6%. For low negative affect, perceived support's link dropped from explaining 19% of the variance to explaining 7% when ordinary conversation was controlled.

One disadvantage of round-robin designs is that because groups of people rate each other, one is often limited to small groups who know each other well. This often leaves out important relationships. For example, in the study of Marines just described, Marines did not rate important relationships such as those with parents, romantic partners, or closest friends, unless these people were members of the four-person teams. An alternative is a one-with-many design in which each participant (the one) rates many relationship partners (the many). Although this design permits the study of each person's most important relationships, the design combines relationship effects with provider effects into a single social influences effect. This is not a major problem in social support research as provider effects are typically very small (Lakey & Orehek, 2011), and thus nearly all social influences reflect relationship effects.

An example of a one-with-many design using the Ordinary Conversation Scale is provided by Woods *et al.* (2016, Study 1). Each student rated her mother, father, and closest peer, permitting the isolation of social influences. Both perceived support and ordinary conversation were strongly socially influenced, accounting for 74% and 69% of the variance, respectively. As predicted, socially influenced conversation quality was significantly correlated with perceived support ($r = .79$), positive affect ($r = .77$), and low negative affect ($r = -.60$). That is, when a provider evoked high-quality ordinary conversation, the provider was seen by the recipient as supportive and as evoking high positive and low negative affect. In addition, when ordinary conversation was controlled, perceived support's link to positive affect was reduced from accounting for 50% of the variance to accounting for 2%. Perceived support's link to low negative affect was reduced from accounting for 31% to accounting for 1%.

Critique

The OCS is relatively new, and thus its construct validity is not fully developed. Additional studies are needed to replicate initial findings with a wider range of constructs, samples, and investigators. Moreover, RRT describes mechanisms by which ordinary conversation regulates affect, action, and thought, but the OCS was not designed to assess those specific mechanisms. Although there is some evidence that independent observers agree with participants when unusually good conversations occur (Lakey *et al*., 2016, Study 3), the magnitude of the agreement is not strong. It will be important to draw clearer links between listening and perceptions of ordinary conversation. For example, research has identified a range of verbal and nonverbal behaviors that participants believe indicate effective listening (Bodie, St. Cyr, Pence, Rold, & Honeycutt, 2012). It will be important to determine the role of these types of behaviors in the quality of ordinary conversation.

References

Bodie, G. D., St. Cyr, K., Pence, M., Rold, M., & Honeycutt, J. M. (2012). Listening competence in initial interactions I: Distinguishing between what listening is and what listeners do. *International Journal of Listening*, *26*, 1–28. doi:10.1080/10904018.2012.639645

Cohen, S., & Wills, T. A. (1985). Stress, social support and the buffering hypothesis. *Psychological Bulletin*, *98*, 310–357. doi:10.1037/0033-2909.98.2.310

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York, NY: Guilford Press.

Lakey, B. & Orehek, E. (2011). Relational Regulation Theory: A new approach to explain the link between perceived support and mental health. *Psychological Review*, *118*, 482–495. doi:10.1037/a0023477

Lakey, B., Vander Molen, R., Fles, E., & Andrews, J. (2016). Ordinary social interaction and the main effect between perceived support and affect. *Journal of Personality*, *84*, 671–684.

Woods, W., Lakey, B., & Sain, T. (2016). The role of ordinary conversation and shared activity in the main effect between perceived support and affect. *European Journal of Social Psychology*, *46*, 356–368.

Scale

Ordinary Conversations Scale (OCS) (Lakey, Vander Molen, Fles, & Andrews, 2016)

Source: Lakey. Reproduced with permission of Brian Lakey.

Please read each statement carefully and indicate how much you agree or disagree with each. Please mark your answers according to the following format:

1. Strongly disagree
2. Mildly disagree
3. Neutral
4. Mildly agree
5. Strongly agree

1) I enjoy talking with her because we have interesting conversations that last a long time.
2) It is difficult to find something she and I both want to talk about.*
3) It is hard to have a conversation with her because she repeatedly says things that have no relevance to what I am talking about.*
4) When we have a conversation, we can go back and forth for as long as we want.
5) My conversations with her usually end quickly.*
6) I hardly ever change the subject when talking to her because she always has something interesting to talk about.
7) It is hard to talk with her because she never has anything new to say.*
8) I normally forget our conversations soon after they are done.*

Note: Instructions should be modified to fit the relationship under study. Items should be randomized prior to administration. Items marked with an asterisk (*) should be reverse coded prior to scoring.

Profile 50

Organizational Listening Survey (OLS)

(Cooper & Buchanan, 1999, 2003)

Profiled by: Laura A. Janusik, PhD, MBA

Rockhurst University

Construct

The Organizational Listening Survey (OLS) was designed to measure an individual's listening competency within a business context (Cooper & Husband, 1993).

Instrument Type

Self- or Other-Report

Description

The OLS is a 30-item measure of listening competency in an organizational context. Both self- and other-report versions are available. Cooper and Husband (1993) defined listening competency as the "knowledge and ability to effectively use behaviors which show an accurate understanding of the message as well as demonstrate support for the relationship between the communication participants, within the appropriate boundaries of the organizational situation" (pp. 13–14). These two factors (accuracy and support) are thought to underlie self-reports of listening competence, although some research has reported evidence for other factor structures. The other-report version of the scale has generated data consistent with a five-factor model: accuracy, support, openness, verbal cues, and nonverbal cues (Cooper & Buchanan, 2003).

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

The OLS is administered as a paper-and-pencil assessment or can be used in online surveys. The assessment takes less than 5 minutes to complete. Each of the 30 items assesses how the individual perceives herself as a listener in her organization. Versions can be created for self-report by mostly adding the pronoun *I*, and the prompts can be modified for other-report by changing the pronoun to *she* or *he* to denote the target of interest. For example, "I say something to show understanding" is reworded as "She says something to show understanding." Care should be taken when administered in an organizational context, as some respondents who are calculating their own scores may find reverse scoring confusing.

Scoring

Items for both the self- and other-report versions are rated from 1 (*never*) to 7 (*always*). Twelve items are reverse-coded to reduce response bias. Scoring is done by either summing or averaging items within a subscale. Because of variability in factor structure (discussed further in this profile), it is advisable to conduct a factor analysis prior to generating scales if used for research purposes. Normative data have not yet been reported for the OLS.

As a self-report measure, final scores indicate what respondents perceive themselves to do, not necessarily what they actually do. This is a useful feature when the OLS is used as a 360-degree feedback instrument, as the focus becomes one's perception of oneself compared to others' perceptions.

Development

Communication competency has its conceptual roots in classical rhetoric and involves the impression or judgment of the appropriateness and effectiveness of communication behavior (Rubin, 1990; Spitzberg & Cupach, 1984). Listening competency assumes that competency is based on perceived appropriateness and effectiveness; that is, individuals understand the content of the encounter, have not violated any norms or rules excessively, and are able to achieve their interaction goals. Competency is contextual; that is, situational realities and constraints determine what is appropriate and effective communication behavior (Chomsky, 1964). Competency has both cognitive and behavioral aspects, both of which contribute to an "impression" of an interaction and the individuals involved in it (Diez, 1984). Competency is functional because particular behaviors are used to achieve intentional or unintentional goals. The literature suggests that listening will be perceived in clusters of related behavior that are seen as competent (Brownell, 1987; Husband, Cooper, & Monsour, 1988; Husband, Schenck, & Cooper, 1988; Lewis & Reinsch, 1988).

The original OLS, known as the Managerial Listening Survey (MLS), was designed to assess supervisor perceptions of their own listening competency in an organizational context. Husband, Cooper, and Monosour (1988) created 60 items to assess "seven basic components … most often identified and discussed as facets or dimensions of listening: 1) receiving stimuli or hearing, 2) attending to, 3) understanding, 4) remembering or retaining, 5) interpreting or assessing meaning, 6) evaluation, and 7) responding" (pp. 99–100). These items were reviewed by a group of communication scholars and professional managers who suggested retaining 40 items. These 40 items were administered to 122 utility company supervisors who assessed how often they displayed each of the 40 behaviors using the previously described 7-point scale (see the Scoring section). A principal component analysis (PCA) with varimax rotation resulted in seven components accounting for 81.3% of item variance: attending, clarifying, responding, discriminating and evaluating, recalling, affiliating, and accommodating. The correlation matrix for these seven factors was submitted to "a higher order factor analysis" (extraction and rotation not specified), which further endorsed the previously described two dimensions: support (attending, clarifying, affiliating, and accommodating) and accuracy (discriminating and recalling).

Additional studies reported revisions to the MLS, including removing five items that were more dispositional than behavioral in nature and adding an other-report version of the scale (Husband, Schenck, & Cooper, 1988). A final revision of the scale was completed by Cooper and Husband (1993), who reduced the number of items to 30 and renamed the scale the Organizational Listening Survey. Most recently, with slight wording modifications, these 30 items were administered in the education context in order to expand the contexts within which listening competency can be measured by the scale (Cooper & Buchanan, 2010).

Reliability

The OLS has generated appropriate internal consistency estimates, as well as interrater agreement (Cooper & Buchanan, 1999, 2003, 2010; Cooper, Seibold, & Suchner, 1997). Throughout the years, the MLS and OLS have achieved estimates of Cronbach's alpha ranging from .82 (Husband *et al.*, 1988; Stine, Thompson, & Cusella, 1995) to .93 (Cooper & Husband, 1993). Husband *et al.* (1988) reported alpha values for the seven subscales found in the original PCA between .65 and .77.

Validity

The validity portfolio of the OLS includes tests of model fit as well as convergent validity. The original MLS was generated using PCA (Husband *et al.*, 1988), and the most recent test of the OLS in the educational context used this same method (Cooper & Buchanan, 2010). PCA is a data reduction technique and inappropriate for making claims about the underlying factor structure of data. To date, there are two reported confirmatory factor analyses in the published literature, neither of which generated fit statistics that adhere to conventional standards. Cooper and Husband (1993), using 22 of the 30 OLS items, tested a two-factor model with items loading on either accuracy (11 items) or support (8 items) or double loading on both latent factors (3 items). This model returned the following: GFI = .813 and RMSEA = .245 for the other-report version; and GFI = .629 and RMSEA = .401 for the self-report version. Similarly poor fit statistics were reported by Cooper, Seibold, and Suchner (1997), who used 19 of the OLS items and tested single-factor models for both the self- and other-report versions.

Three additional factors were suggested by the Cooper *et al.* (1997) data: openness to others' ideas, use of verbal cues, and use of nonverbal cues. In general, the data reported to date do not support the construct validity of the OLS as measuring two latent factors; the unidimensional model also is questionable. More work is needed to discover a consistent underlying factor structure that is invariant across versions.

In terms of convergent validity, data from the previously described studies suggest that perceptions of listening competency are positively correlated with satisfaction with the work relationship and satisfaction with the professor. The greatest threat to convergent validity within OLS research has been with the lack of correlation between self- and other-report versions. To date, there has been no multigroup factor-analytic work to test factorial invariance. It may be that results showing that perceptions do not correlate are either a function of different factor structures or a function of true differences. Most reports do, however, suggest a high degree of correspondence between raters (e.g., Cooper & Buchanan, 2010, $r_{wg} = .84$), providing evidence that judges can reliably assess a single target.

Evidence for divergent validity, in the form of assessing whether perceptions of listening competence correlate with actual behaviors, has not yet been reported. The degree to which what one thinks she does and what actually occurs in an organizational or educational setting is questionable for socially desirable behaviors such as listening (Ford, Wolvin, & Chung, 2000).

Availability

The prompts used for the OLS by Cooper and Buchanan (2010) with a student sample are provided at the end of this profile with permission. However, they should not be used without first contacting the primary author, Dr. Lynn Cooper, for specific instructions. The most current version of the instruments, both self-report (Form C) and other-report (Form D), may be obtained directly from Dr. Cooper via email (lynn.cooper@wheaton.edu).

Sample Studies

Most studies of the OLS have been dedicated to the development of the instrument and are covered elsewhere in this profile (Cooper, 1992, 1997; Cooper & Buchanan, 1999, 2003, 2010; Cooper & Husband, 1993; Cooper *et al.*, 1997; Husband *et al.*, 1988a, 1988b).

Stine *et al.* (1995), using a version of the MLS, assessed the relations between listening, supportiveness, and trustworthiness among 89 employees of a Midwestern tool manufacturing company. Some of the probes were slightly reworded, although specific details on the rewording were not provided. Results supported that perceived listening was positively correlated with perceived supportiveness ($rs > .70$) and trustworthiness ($rs > .50$).

Using data from their 1999 study, which included 444 employees of a Midwestern petroleum-refining company, Cooper and Buchanan (2003) found that interrater agreement between self- and other-perceptions of listening were strongest for those perceived to be the worst listeners, even though a significant relationship was reported for all three groups: Less Competent ($r = .93$), Moderately Competent ($r = .90$), and Most Competent ($r = .84$).

In the student data presented by Cooper and Buchanan (2010), student perceptions of instructor listening competence were associated with satisfaction with the class ($r = .47$), frequency of communication with the professor ($r = .44$), time spent communicating with the professor one-on-one ($r = .36$), relationship with the professor ($r = .45$), and overall evaluation of listening effectiveness ($r = .40$).

Critique

The OLS is drawn from a strong theoretical framework, and the initial study proposed the systematic development of an instrument that could measure one's own and others' perceptions of listening competency in an organization (Husband *et al.*, 1988). As suggested above, the factor structure of the 30 items is not stable across administrations. The original structure was proposed as a second-order latent model (support and accuracy) with seven first-order latent constructs (attending, clarifying, responding, discriminating, recalling, affiliating, and accommodating). Later versions tested only a two-factor and unidimensional model, even though these models have yet to be supported by data. Moreover, the self- and other-report versions have not met standards for factorial invariance. As such, comparisons of self- and other-report scores are questionable. When factor structures are not similar, any differences in ratings could be due to method effects and not to actual differences (Little, 1997).

As with most instruments, further data collection with independent samples could assist with further refining the instrument, making it even more useful. What is noteworthy is that if the OLS items do, in fact, cluster around a single factor, there is a great deal of redundancy in the instrument. To date, the primary method to generate the factor structure has been PCA, a data reduction technique that has different assumptions compared to factor analytic techniques. More work is needed to build a case for construct validity.

Because this is a business instrument used in an organizational setting, it has likely been used by many practitioners who are not researchers and do not publish results for a number of reasons. Even with profound validity issues that remain to be settled, its practical significance cannot go unstated, particularly when used as a 360-degree evaluation. How others view us and how we view ourselves are often different. When we can see how others view us, we can learn to view our listening differently (and perhaps improve). Having access to another's report is often the best way to begin struggling with one's own listening competency, as competency is a judgment made by others.

References

Brownell, J. (1987). *Perceptions of listening behavior: A management study*. Unpublished manuscript, Cornell University.

Chomsky, N. (1964). *Current issues in linguistic theory*. The Hague: Mouton.

Cooper, L. O. (1992). *A survey of research in organizational listening competency*. Paper presented at the International Listening Association convention, Seattle, WA.

Cooper, L. O. (1997). Listening competency in the workplace: A model for training. *Business Communication Quarterly*, *60*, 75–84. doi:10.1177/108056999706000405

Cooper, L. O., & Buchanan, T. (1999). Inter-rater agreement in judgments of listening competency: An item-based analysis of the Organizational Listening Survey. *Communication Research Reports*, *16*, 48–54. doi:10.1080/08824099909388700

Cooper, L. O., & Buchanan, T. (2003). Taking aim at good targets: Inter-rater agreement of listening competency. *International Journal of Listening, 17*, 88–114. doi:10.1080/10904018.2003.10499057

Cooper, L. O., & Buchanan, T. (2010). Listening competency on campus: A psychometric analysis of student listening. *International Journal of Listening, 24*, 141–163. doi:10.1080/10904018.2010.508681.

Cooper, L. O., & Husband, R. L. (1993). Developing a model of organizational listening competency. *International Journal of Listening, 7*, 6–34. doi:10.1080/10904018.1993.10499112

Cooper, L. O., Seibold, D. R., & Suchner, R. (1997). Listening in organizations: An analysis of error structures in models of listening competency. *Communication Research Reports, 14*, 312–320. doi:10.1080/08824099709388674.

Diez, M. E. (1984). Communication competence: An interactive approach. In R. Bostrom (Ed.), *Communication Yearbook 8* (pp. 56–79). Newbury Park, CA: Sage.

Ford, W. S. Z., Wolvin, A. D., & Chung, S. (2000). Students' self-perceived listening competencies in the basic speech communication course. *International Journal of Listening, 14*, 1–13. doi:10.1080/10904018.2000.10499032

Husband, R. L., Cooper, L. O., & Monsour, W. M. (1988a). Factors underlying supervisors' perceptions of their own listening behavior. *International Journal of Listening, 2*, 97–112. doi:10.1080/10904018.1988.10499100

Husband, R. L., Schenck, T., & Cooper, L. O. (1988b, March). *A further look at managerial listening.* Paper presented at the meeting of the International Listening Association Convention, Scottsdale, AZ.

Lewis, M. H., & Reinsch, N. L., Jr. (1988). Listening in organizational environments. *Business Communication, 25*, 49–67. doi:10.1177/002194368802500304

Little, T. D. (1997). Mean and covariance structures (MACS) analysis of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*, 53–76. doi:10.1207/s15327906mbr3201_3

Rubin, R. B. (1990). Communication competence. In G. Phillips & J. T. Wood (Eds.), *Speech communication: Essays to commemorate the 75th anniversary of the Speech Communication Association* (pp. 94–129). Carbondale, IL: Southern Illinois University Press.

Spitzberg, B. H., & Cupach, W. R. (1984). *Interpersonal communication competence.*Beverly Hills, CA: Sage.

Stine, M., Thompson, T., & Cusella, L. (1995). The impact of organizational structure and supervisory listening indicators on subordinate support, trust, intrinsic motivation, and performance. *International Journal of Listening, 9*, 84–105. doi:10.1080/10904018.1995.10499143)

Scale

Organizational Listening Survey (OLS; Cooper & Buchanan, 2010)

Source: Cooper and Buchanan (2010). Reproduced with permission of Taylor & Francis.

Instructions: For your instructor, please respond to each of the following items using the following scale.

Always	1	2	3	4	5	6	7	Never

<u>Accuracy</u>

1) Asks when words used in an unfamiliar way
2) Reacts to detail; sometimes misses the point*
3) Can't remember significant details*[a]
4) Gives straightforward information
5) Remembers relevant details
6) Doesn't discriminate fact from opinions*
7) Finds it difficult to express ideas*[a]
8) Is easily distracted*
9) Analyzes what is said
10) Fails to hear the consistency in facts and logic when others talk
11) Allows preconceived attitudes to interfere*

<u>Support</u>

12) Says something to show understanding
13) Watches for tones and gestures
14) Makes eye contact when listening
15) Listens only when action is required*
16) Tries to make one feel at ease
17) Makes negative statements*
18) Restates what was said
19) Uses positive nonverbal expressions

<u>Double Loading</u>

20) Takes time to listen
21) Asks simple questions to clarify
22) Asks clarifying questions

<u>Additional Items</u>

23) Tasks more important than listening*
24) Listens to every word to get main ideas
25) Listens without concentrating*
26) Tries to think of points to contribute
27) Affects behavior by listening
28) Disregards situational factors*
29) Evaluates a student's credibility
30) Listens to students in the same way*

Note: Instructions and items can be adapted to reflect either a self- or other-report for a designated individual. This version is written as an other-report for a designated instructor. The scale's author has requested that researchers contact her for additional instructions (lynn.cooper@wheaton.edu). Items listed under *Accuracy* and *Support* were specified to load on these factors by Cooper and Husband (1993). Items listed under *Double Loading* were specified to load on both accuracy and support factors in the Cooper and Husband study. Labels should be removed and items randomized prior to administration. Items marked with an asterisk (*) should be reverse coded prior to scoring.

Profile 51

Perceived Partner Responsiveness Scale (PPRS)

(Reis & Carmichael, 2006)

Profiled by: Harry T. Reis, PhD[1], Dev Crasta[1], Ronald D. Rogge[1], Michael R. Maniaci[2] and Cheryl L. Carmichael[3]

[1] *University of Rochester*
[2] *Florida Atlantic University*
[3] *Brooklyn College*

Construct

The Perceived Partner Responsiveness Scale (PPRS) was designed to measure the degree to which people feel that their relationship partners are responsive to them.

Instrument Type

Self-Report

Description

The PPRS is a measure of people's perceptions of their relationship partners' responsiveness to themselves. This 18-item measure incorporates two closely related constructs, based on the interpersonal process model of intimacy originally proposed by Reis and Shaver (1988): understanding (the degree to which another person seems to "get things right" about oneself) and validation (the degree to which another person is believed to appreciate and value oneself). The measure is intended to assess a specific target's responsiveness to the respondent and is most commonly used for romantic partners. It can easily be adapted to refer to other relationship types, such as friends, family, and acquaintances. The scale pointedly asks about global perceptions of the partner's responsiveness; such perceptions will not necessarily correspond to the partner's attitudes, intentions, or perceptions, or to objectively coded behavior.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

The PPRS is a self-administered instrument that takes approximately 3 to 4 minutes to complete. Items are written to be general, so that participants can complete the scale with respect to a particular relationship (e.g., a romantic partner, spouse, best friend, work supervisor, parent, or coach). The relationship of interest is inserted into the instructions, although the stem "My partner usually" can be modified to refer to a specific target as well (e.g., "My coach usually"). The stem also can be adjusted to apply to a more specific moment (e.g., "Today, my partner …") or interaction (e.g., "During this conversation, my partner …"), with corresponding modifications to each item's verb tense. After reading the item, participants are asked to indicate the degree to which that item applies to the individual being considered. On the original measure, each item is followed by a 9-point scale with the following anchors: 1 = *not at all true*, 3 = *somewhat true*, 5 = *moderately true*, 7 = *very true*, and 9 = *completely true*. Five-point and 7-point versions have also been used, with the same end-anchors.

Scoring

There are eight items each for the understanding and validation subscales, along with two general items. Computing a total responsiveness score is the most common usage, calculated by simple summation of ratings across all 18 items. If subscale scores are desired, they can be calculated by summing ratings for the appropriate items.

Development

The 18-item PPRS was first introduced by Harry Reis and Cheryl Carmichael (2006) in an unpublished study of married spouses' experiences of intimacy and support. A few items have undergone minor changes in wording since then. Subsequently, a 12-item version of the measure has been developed (Reis, Maniaci, Caprariello, Eastwick, & Finkel, 2011), which typically generates similar reliability and has demonstrated adequate validity compared to the longer version. A three-item version has also appeared, which is better suited to protocols that demand brevity, such as experience sampling and daily diary studies (Gable, Gosnell, Maisel, & Strachman, 2012).

Reliability

Internal consistencies for both the 12-item and 18-item PPRSs tend to be high, ranging from .91 to .98 in most published and unpublished samples (e.g., Birnbaum & Reis, 2006; Reis *et al.*, 2011; Reis, Maniaci, & Rogge, 2014).

Validity

Using data from an Internet-based sample of over 2000 individuals who were asked to complete the PPRS with regard to a close other, exploratory factor analyses using principal axis extraction with an oblimin rotation supported a unidimensional solution, with all items loading highly on a single factor. This solution has emerged consistently (with similar factor loadings) across different relationship types (e.g., a romantic partner, close friend, and family members). In the Internet-based sample, as well as several other samples, there was mildly suggestive evidence of a two-factor solution corresponding to understanding and validation, but even in this circumstance, the two factors have a substantial correlation, $r = .94$. This likely reflects the fact that understanding and validation tend to co-occur in close relationships.

Various studies have demonstrated convergent validity through correlations with other scales designed to measure responsiveness, including relationship satisfaction ($r = .82$), trust ($r = .67$), empathy ($r = .51$), and emotional support ($r = .49$). Other findings that contribute to the measure's validity portfolio include an experimental study in which PPRS scores increased as first-year college students became better acquainted with each other (Reis *et al.*, 2011). In another study, PPRS scores were significantly correlated with reports of a partner's daily compassionate behaviors ($r = .33$), suggesting that donors and recipients agree about behavioral manifestations of responsiveness (Reis *et al.*, 2014). Another unpublished study found significant agreement in a laboratory conversation between ratings of responsiveness provided by relationship partners and independent coders ($r = .33$) (Rusbult, Kumashiro, & Reis, 2011).

Availability

The Romantic Partner version of the PPRS appears at the end of this profile. Other versions can be created by changing the wording of the relationship of interest (e.g., best friend, parent, or work supervisor). The scale is freely available to researchers with appropriate citation.

Sample Studies

PPR has been found to be associated with numerous relationship qualities, particularly those that relate to intimacy and support. For example, perceived partner responsiveness is positively associated with relationship satisfaction, trust, intimacy, and most forms of support. The PPRS has been used to clarify why, in some circumstances, social support may not be helpful. Maisel and Gable (2009) found that support provided by others is effective only when it is perceived as responsive; in other words, support intended to be helpful that nonetheless is perceived as nonresponsive tends to undermine well-being. Other studies have shown that responsiveness about personal positive events—good things that have happened in one's life—may actually benefit relationships more than responsiveness about negative events and stressors (the traditional focus of social support research). For example, ratings of PPR following a laboratory conversation about positive but not negative events predicted changes in relationship well-being over 2 months (Gable, Gonzaga, & Strachman, 2006). Another laboratory study found that socially anxious people view themselves—and are viewed by their partners—as less responsive in conversations about the partner's good news (Kashdan, Ferssizdis, Farmer, Adams, & McKnight, 2013).

PPR also has been related to sexual desire. Among women in established relationships, PPR predicts viewing sex as exciting and as a way of strengthening a relationship; it also is negatively related to feeling distracted, distant, and ashamed during sex (Birnbaum & Reis, 2006). In a daily diary study of newlyweds, PPR mediated the relationship between sexual satisfaction and relationship satisfaction (Gadassi *et al.*, in press). Birnbaum and Reis (2012) found, however, that PPR piqued sexual interest in a new acquaintance only among individuals high in attachment security.

Although mostly used in established relationships, studies of new acquaintances also have used the PPRS successfully. For example, previously unacquainted individuals who had consumed a moderate dose of alcohol, compared to a placebo, were rated as more responsive following conversations about a significant person in their lives (Kirkpatrick & de Wit, 2013). Forest and Wood (2011) found that responsiveness displayed by a new acquaintance increased expressivity among individuals with low self-esteem but not among individuals with high self-esteem. In two experiments, Reis *et al.* (2011) found that randomly paired college students increased their ratings of PPR the more they chatted with each other, and that these increases mediated increases in liking. In a similar design, perceived responsiveness (and liking) was rated lower among students engaging in computer-mediated text-only conversations compared to face-to-face interactions or computer-mediated conversations that included audio or video channels (Sprecher, 2014).

These diverse studies are consistent with the broad idea that PPR is central to the development and maintenance of intimate relationships (for a review, see Reis and Clark, 2013).

Critique

PPR is best considered as an outcome of good listening skills; that is, when a listener has been effective, in the large majority of circumstances, speakers will feel responded to. As such, the PPRS can contribute to a research program on listening by providing an index of the recipient's perceptions. Although the PPRS has demonstrated excellent reliability and convergent validity, it has been used primarily in relatively intimate relationships among white, educated, middle-class Westerners. Research is needed to demonstrate the measure's usefulness in other types of relationships and more diverse samples. PPRS scores also tend to be substantially correlated with other measures of relationship quality. It would be desirable to develop a measure that better distinguishes these characteristics.

References

Birnbaum, G. E., & Reis, H. T. (2006). Women's sexual working models: An evolutionary-attachment perspective. *The Journal of Sex Research*, *43*, 328–342. doi:10.1080/00224490609552332

Birnbaum, G. E., & Reis, H. T. (2012). When does responsiveness pique sexual interest? Attachment and sexual desire in initial acquaintanceships. *Personality and Social Psychology Bulletin*, *38*, 946–958. doi:10.1177/0146167212441028

Forest, A. L., & Wood, J. V. (2011). When partner caring leads to sharing: Partner responsiveness increases expressivity, but only for individuals with low self-esteem. *Journal of Experimental Social Psychology*, *47*, 843–848. doi:10.1016/j.jesp.2011.03.005

Gable, S. L., Gosnell, C. L., Maisel, N. C., & Strachman, A. (2012). Safely testing the alarm: Close others' responses to personal positive events. *Journal of Personality and Social Psychology*, *103*, 963–981. doi:10.1037/a0029488

Gable, S. L., Gonzaga, G. C., & Strachman, A. (2006). Will you be there for me when things go right? Supportive responses to positive event disclosures. *Journal of Personality and Social Psychology*, *91*, 904–917. doi:10.1037/0022-3514.91.5.904

Gadassi, R., Bar-Nahum, L. E., Newhouse, S., Anderson, R., Heiman, J. R., Rafaeli, E., & Janssen, E. (2016). Perceived partner responsiveness mediates the association between sexual and marital satisfaction: A daily diary study in newlywed couples. *Archives of Sexual Behavior*, *45*, 109–120. doi:10.1007/s10508-014-0448-2

Kashdan, T. B., Ferssizidis, P., Farmer, A. S., Adams, L. M., & McKnight, P. E. (2013). Failure to capitalize on sharing good news with romantic partners: Exploring positivity deficits of socially anxious people with self-reports, partner-reports, and behavioral observations. *Behaviour Research and Therapy*, *51*, 656–668.

Kirkpatrick, M. G., & de Wit, H. (2013). In the company of others: Social factors alter acute alcohol effects. *Psychopharmacology*, *230*, 215–226. doi:10.1007/s00213-013-3147-0

Maisel, N. C., & Gable, S. L. (2009). The paradox of received social support: The importance of responsiveness. *Psychological Science*, *20*, 928–932. doi:10.1111/j.1467-9280.2009.02388.x

Reis, H., & Carmichael, C. (2006). *Married spouses' experiences of intimacy and support.* Unpublished manuscript.

Reis, H. T., & Clark, M. S. (2013). Responsiveness. In J. A. Simpson & L. Campbell (Eds.), *The Oxford handbook of close relationships* (pp. 400–423). New York: Oxford University Press. doi:10.1093/oxfordhb/9780195398694.013.0018

Reis, H. T., Maniaci, M. R., Caprariello, P. A., Eastwick, P. W., & Finkel, E. J. (2011). Familiarity does indeed promote attraction in live interaction. *Journal of Personality and Social Psychology*, *101*, 557–570. doi:10.1037/a0022885

Reis, H. T., Maniaci, M. R., & Rogge, R. D. (2014). The expression of compassionate love in everyday compassionate acts. *Journal of Social and Personal Relationships*, *31*, 652–677. doi:10.1177/0265407513507214

Reis, H. T. & Shaver, P. (1988). Intimacy as an interpersonal process. In S. Duck (Ed.), *Handbook of personal relationships* (pp. 367–389). Chichester, UK: John Wiley & Sons, Ltd.

Rusbult, C. E., Kumashiro, M., & Reis, H. T. (2011). *Perceived partner responsiveness and the Michelangelo Effect.* Unpublished manuscript, Vrije Universiteit Amsterdam.

Sprecher, S. (2014). Initial interactions online-text, online-audio, online-video, or face-to-face: Effects of modality on liking, closeness, and other interpersonal outcomes. *Computers in Human Behavior*, *31*, 190–197. doi:10.1016/j.chb.2013.10.029

Scale

Perceived Partner Responsiveness Scale: Romantic Partner Version

Source: Reis. Reproduced with permission of Harry Reis.

Instructions: Please answer the following questions about your current romantic partner.

Response Categories:

1	2	3	4	5	6	7	8	9
Not at all true		Somewhat true		Moderately true		Very true		Completely true

My partner usually:

General Items
… really listens to me.*
… is responsive to my needs.*

Understanding Items

… is an excellent judge of my character.
… sees the "real" me.*
… sees the same virtues and faults in me as I see in myself.
… "gets the facts right" about me.*
… is aware of what I am thinking and feeling.
… understands me.*
… is on "the same wavelength" with me.*
… knows me well.*

Validation Items

… esteems me, shortcomings and all.*
… values and respects the whole package that is the "real" me.*
… usually seems to focus on the "best side" of me.
… expresses liking and encouragement for me.*
… seems interested in what I am thinking and feeling.*
… seems interested in doing things with me.
… values my abilities and opinions.*
… respects me.

Note: Labels should be removed and items randomized prior to administration. Items marked with an asterisk (*) are those included in the 12-item version of the PPRS.

Profile 52

Profile of Nonverbal Sensitivity (PONS)

(Rosenthal, Hall, DiMatteo, Rogers, & Archer, 2013)

Profiled by: Laura A. Janusik, PhD, MBA

Rockhurst University

Construct

The Profile of Nonverbal Sensitivity (PONS) was designed to measure one's "ability to decode nonverbal cues conveyed by the face, body, and tone of voice" (Rosenthal *et al.*, 2013, p. 1). The PONS is useful in examining individual differences in interpersonal sensitivity as well as detecting differences in channels of communication (Ambady, LaPlante, & Johnson, 2001).

Instrument Type

Cognitive Assessment

Description

There are four versions of the PONS, each of which measures nonverbal decoding accuracy by assessing a participant's ability to identify a female encoder's situation-specific state (Hall, 2001).[1] Each version contains various 2-second segments of pure audio, pure video, or a combination of audio and video (Banziger, Scherer, Hall, & Rosenthal, 2011a). Three

1 It should be noted that four other forms of the test are included in the instructor's manual (Rosenthal *et al.*, 2013): an Audio-Only Version, consisting of both a female and male sender; a Still-Photo version; a 20-item Visual Only; and a 20-item Brief Exposure. Testing of all of these versions, with the exception of the 40-item Face and Body, showed weak results, so these versions will not be addressed. Additionally, some researchers mention the use of the "1/2 PONS" (Ambady, Hallahan, & Rosenthal, 1995; Krendl & Ambady, 2010), which consists of 110 of the 220 items from the Full PONS test, but that form is not included in the manual.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

of the forms (Vocal Expression, Face and Body, and Mini-PONS) are derived from items in the Full PONS Test. All use a similar response system. After watching or listening to the stimuli, the participant selects one of two given responses. For example, after watching a scene of a woman's upper torso and head while she is smiling and looking around, accompanied by indiscernible speech that is fast and high pitched, one would either select "admiring nature" or "expressing motherly love."

Form 1

The Profile of Nonverbal Sensitivity (PONS): Full Test consists of 220 items presented in black and white. Some of the clips are audio only, some video only, and some a combination of video and audio. The test runs for 47 minutes.

Form 2

The Profile of Nonverbal Sensitivity (PONS): Vocal Expression Test consists of 40 items of content-masked speech. The audio was spliced so that full words could not be understood. Subsequently, the participant has to rely on other cues, like tone and volume of voice, to identify the situation-specific state of the encoder. The test runs for 7 minutes and 40 seconds.

Form 3

The Profile of Nonverbal Sensitivity (PONS): Face and Body Test consists of 40 black-and-white video segments. Each segment shows only a face, body, or hand movement. The video clips have no sound. The test runs for 7 minutes and 43 seconds.

Form 4

The MiniPONS Test, released in 2011, contains 64 video items from the original test (Banziger, Klaus, Hall, & Rosenthal, 2011b). The test runs for 12 minutes and 20 seconds.

Administration

All four forms of the profile are distributed through the Northeastern University Libraries at https://repository.library.northeastern.edu. Answer sheets may be found in the instruction manual located on the website. The videos may be shown to a large group, or individuals may access the tests on their own through private computers (Bänziger *et al.*, 2011a). For those with no Internet access at the testing site, the videos may be downloaded for later presentations.

Scoring

Scoring of all tests is the same. After seeing or hearing the stimulus, the answer sheet provides two possible answers reflecting an emotional state, and the participant selects what she or he believes is the most appropriate answer. The instruction manual includes an answer key, and the test is scored to reflect the number of correct answers.

Development

To create the PONS, a research team filmed a 24-year-old woman (the encoder) who acted out several interpersonal "scenes" in an interactional manner (Rosenthal *et al.*, 2013). Thirty-five videotaped scenes representing 30 different interpersonal situations were spliced into 220 different segments representing different channels: full body, no sound (head to knee); face only, no sound; body only, no face or sound; and two audio channels consisting of random spliced voices and low-pass filter sound. Scenes were limited to 2 seconds to optimize test difficulty (i.e., for a two-item choice test, the typical participant would score 75% correct) (Hall, 2001).

The final items on the PONS ($N = 220$) were selected based on a predetermined factorial structure (Hall, 2001; Rosenthal *et al.*, 1979), which "required five scenes in each of four quadrants, crossed by 121 channels which themselves represent the crossing of two vocal channels with three video channels plus each channel in isolation" (Hall, 2001, p. 149). The first five channels are "pure," meaning that they either consist of a video clip with no sound (channels 1–3) or an auditory clip with no video (channels 4 and 5). The remaining six channels are mixtures of the pure channels containing both visual and auditory cues (Rosenthal *et al.*, 2013). Each channel was further divided into 2-second auditory and/or visual segments. The final scenes represent four affective quadrants (positive–dominant, positive–submissive, negative–dominant, and negative–submissive) crossing 11 different channels (Funder & Harris, 1986).

Normative data have been published for a wide range of ages (elementary students to adults) for different forms of the test (Hall, 2001; Rosenthal *et al.*, 2013).

In response to criticism that the PONS led to experienced test fatigue, the MiniPONS Test was created and published in 2011 (Bänziger *et al.*, 2011b). This short, multichannel version consists of 64 of the 2-second scenes from the original PONS and takes no longer than 15 minutes to administer. Researchers have also used parts of the PONS (e.g., audio only) in several studies.

Reliability

Internal consistency has ranged from .86 when computed with a Kuder-Richardson formula 20 (KR-20) (Rosenthal *et al.*, 1979) to .92 when computed with Armor's theta (Rosenthal, 1974, as cited in Rosenthal *et al.*, 2013). Notably, the average interitem correlation was only .03 (Hall, 2001). The PONS has been used widely, and by 1986 it had been tested with more than 7000 participants in 20 different countries (Funder & Harris, 1986; Rosenthal *et al.*, 1979).

The correlations of the subtests within the PONS and the full PONS test are as follows: Self-Administered Still-Photo Version ($r = .64$), Brief Exposure PONS ($r = .54$), Face and Body PONS ($r = .50$), Audio Only PONS ($r = .30$) (Rosenthal *et al.*, 1979, 2013), and the MiniPONS ($r = .70$) (Bänziger *et al.*, 2011a).

Due to low interitem correlations, the shorter forms of the PONS Test (Face and Body, Vocal Expression, and MiniPONS) tend to generate poor reliability estimates: Face and Body, less than .40; and Vocal Expression, .17 to .30 (Hall, 2001).

Hall (2001) argued that the standard psychometric model might not be applicable to nonverbal sensitivity tests. Weak internal consistency exists in many of the other nonverbal decoding tests, including the 15- and 30-item IPT (Interpersonal Perception Task) and the 32-item CARAT (Communication of Affect Receiving Ability), with Cronbach's alphas ranging as low as .06 to .56 (see Hall, 2001).

In six separate studies, the median test–retest reliability was .69 (Rosenthal *et al.*, 1979, 2013), and estimates of the other versions include a test–retest reliability of .64 for the MiniPONS, and .09 to .38 for the Face and Body and Vocal Expressions (Bänziger *et al.*, 2011b).

Validity

Support for construct, convergent, criterion, and discriminant validity has been reported (see Hall, 2001; Rosenthal *et al.*, 1979, 2013).

The initial testing of the PONS instrument was quite extensive, and the researchers were able to identify norms for different groups (Rosenthal *et al.*, 1979, 2013). Subsequent research also investigated the relation of PONS to other constructs such as IQ ($r = .14$), SAT scores ($r = .15$), and cognitive complexity, median ($r = .29$). In addition, in a meta-analysis of 215 studies on interpersonal sensitivity (IS), with the PONS test accounting for a substantial portion of the studies, IS was positively associated with empathy, mean $r = .12$; extraversion, mean $r = .07$; openness, mean $r = .14$; tolerance, mean $r = .18$; and internal locus of control, mean $r = .19$; and negatively associated with neuroticism, mean $r = -.08$ (Hall, Andrzejewski, & Yopchick, 2009).

One's ability to decode in one of the nonverbal channels taped by the PONS could not predict the individual's ability to decode another channel, as the channels represent orthogonal factors (Rosenthal *et al.*, 1979). This issue contributes to the low interitem correlations. The logical question for construct validity is, how does PONS measure up with the other tests of nonverbal interpersonal sensitivity?

A number of studies have examined the relations between versions of the PONS and other nonverbal measures (e.g., Interpersonal Perception Task; Ambady, Hallahan, & Rosenthal, 1995; Bänziger *et al.*, 2011a; Baum & Nowicki, 1998; Nowicki & Duke, 1994; Pitterman & Nowicki, 2004). This discussion focuses on one: the Diagnostic Analysis of Nonverbal Accuracy (DANVA) test. The DANVA focuses on one's ability to identify four emotions (happiness, sadness, anger, and fear) through facial expressions. Several studies report significant correlations between versions of the DANVA (e.g., 2-AF) and PONS tests, including the Face and Body and Vocal Expression (Hall, Roter, Blanch, & Frankel, 2009; Rosip & Hall, 2004). The full version of PONS did not correlate significantly with the DANVA 2-AP (Baum & Nowicki, 1998; Nowicki & Duke, 1994). Similarly, the MiniPONS was not significantly correlated with the DANVA2-AF or DANVA2-AP (Baum & Nowicki 1998; Nowicki & Duke 1994).

Given the time commitment for the PONS, the MiniPONS was created and takes fewer than 15 minutes to complete (Bänziger *et al.*, 2011a). The correlation of the MiniPONS with the Full PONS was .70. When comparing results from the 64 items across versions, the correlation fell to .64. Data reported by Bänziger and colleagues (2011) showed that the MiniPONS correlated significantly with three of the four emotion recognition tests that were used for construct validation: the Multimodal Emotion Recognition Test (MERT; Bänziger, Grandjean, & Scherer, 2009), Japanese and Caucasian Facial Expressions of Emotion (JACFEE) test (Matsumoto & Ekman, 1988), and Emotion Recognition Index (ERI) (Scherer, 2007; Scherer & Scherer, 2011).

Availability

All PONS tests (Full, Audio, Face & Body, and Mini) and their related scoring sheets are available online from the University Libraries of Northeastern University (http://repository.neu.edu/collections/neu:193290/contents/0).

Sample Studies

Because of the volume of studies that have used one or more versions of the PONS test, this section is necessarily selective. In healthcare, good decoders on the PONS received higher ratings of effectiveness from supervisors (Rosenthal *et al.*, 1979), and physicians with more satisfied patients scored higher on the PONS (DiMatteo, Friedman, & Taranta, 1979). In a study of occupational therapy students, those students who scored higher on the face and body PONS received better evaluations in pediatric and psychosocial fieldwork than lower scorers (Tickle-Degnen, 1998). Schizophrenia patients scored significantly lower on the PONS as compared to a healthy control group (Wynn, Sugar, Horan, Kern, & Green, 2010), and higher PONS scorers have been found to be less likely to get depressed (Ambady & Gray, 2002).

Research utilizing the PONS test was the first to report gender difference in decoding nonverbal communication. In the initial 133 samples ($N = 2615$), 80% of the studies showed that females scored slightly higher on nonverbal sensitivity tests than their male counterparts ($d = .62$ for grade school; $d = .49$ for junior high; $d = .57$ for high school; and $d = .44$ for college; $r = .21$ between gender and Full PONS test score; median $d = .42$) (Rosenthal *et al.*, 1979). Similar findings were reported with the MiniPONS, which also is correlated with emotional intelligence (Gulabovska & Leeson, 2014). The Face and Body PONS, however, has shown mixed results with some studies reporting no significant sex-based difference and others suggesting small differences (Loredana & Duduciuc, 2011; Mast, Jonas, & Hall, 2009; Rosip & Hall, 2004; Schmid, Mast, Bombari, & Mast, 2011). Across 11 studies that manipulated the motivation to accurately decode, results were relatively consistent across women and men participants on the Full, Face and Body, and Vocal Expression PONS tests (Hall, Blanch, *et al.*, 2009). In general, if sex-based differences exist, they are small in magnitude.

Finally, in education settings, PONS scores have been associated with IQ (Rosenthal *et al.*, 1979, 2013). A recent meta-analysis of PONS and IQ measures estimated the effect at $r = .19$ (Murphy & Hall, 2011). Related studies suggest that high scorers learn more in interpersonal situations (Bernieri, 1991) and are perceived to be more effective music teachers (Kurkul, 2007).

Critique

The PONS test has been called "a watershed, making its mark as the first systematic, large-scale effort to evaluate individual differences in the ability to decode nonverbal behavior" (Archer *et al.*, 2001, p. 169). It is unique because of its ability to separate vocal and visual interpersonal cues in moving form (Bänziger *et al.*, 2011a). The PONS resulted in three major shifts in the study of nonverbal communication

(Rosenthal, 1979, as cited in Riggio, 2006): (a) a shift from cognitive processes to affective processes, (b) a shift from abstract traits to more concrete abilities, and (c) a shift from a study of inferred traits to that of processes.

Despite its accolades, longevity, and continued use in research studies, several limitations and criticisms should be noted. First, it is possible that PONS scores vary as a function of the encoder. Thus, the use of a single encoder may be problematic (Bänziger *et al.*, 2011a; Riggio, 2006). Second, the test is likely culture specific (Archer *et al.*, 2001; Hall, 2001; Nowicki & Duke, 2001). Initial research included 20 nations outside of the United States and showed that every culture scored better than chance; the cultures most similar to the United States scored the best (Rosenthal *et al.*, 2013). Measurement equivalence across culture has not, however, been demonstrated. It is possible, for instance, that score differences are the result of differential item functioning. Third, participants may lack experience with some portrayed situations (e.g., divorce) (Archer *et al.*, 2001), and thus scores may be less a function of ability and more a function of familiarity. Finally, some scenes are devoid of affect (e.g., ordering a meal), whereas others blend affect, context, and circumstances (e.g., returning a defective item to a store) (Riggio, 2006). These confounds are interesting fodder for future work.

The PONS has also been criticized for its lack of realism (i.e., actors may not accurately express emotion) (Archer *et al.*, 2001; Bernieri, 2001; Riggio, 2006). Support for this claim is seen in the lack of correlation between PONS Face scores and the JACFEE test (Bänziger *et al.*, 2011b).

In terms of listening research, for those that believe that part of listening includes decoding the nonverbal communication of others, the PONS test stands to hold great promise. To date, no work has explored the relation between nonverbal decoding abilities as measured by the PONS and listening traits or behaviors.

Finally, although the PONS test is one of the best known measures of individual ability to decode nonverbal behavior, other excellent measures do exist, including the DANVA (Baum & Nowicki, 1998; Nowicki & Duke, 1994), the IPT (Constanzo & Archer, 1989), the MERT (Bänziger *et al.*, 2009), the JACFEE test (Matsumoto & Ekman, 1988), and the ERI (Scherer, 2007; Scherer & Scherer, 2011).

References

Ambady, N., & Gray, H. M. (2002). On being sad and mistaken: Mood effects on the accuracy of thin-slice judgments. *Journal of Personality and Social Psychology*, *83*, 947–961. doi:10.1037/0022-3514.83.4.947

Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology*, *69*, 518–529. doi:10.1037/0022-3514.69.3.518

Ambady, N., LaPlante, D., & Johnson, E., (2001). Thin-slice judgments as a measure of interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 89–101). Hillsdale, NJ: Erlbaum.

Archer, D., Costanzo, M., & Akert, R. (2001). The Interpersonal Perception Task (IPT): Alternative approaches to problems of theory and design. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 161–182). Hillsdale, NJ: Erlbaum.

Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: The Multimodal Emotion Recognition Test (MERT). *Emotion*, *9*, 691–704. doi:10.1037/a0017088

Bänziger, T., Klaus, S. R., Hall, J. A., & Rosenthal, R. (2011b). Profile of Nonverbal Sensitivity: MiniPONS. Retrieved February 8, 2017, from http://repository.neu.edu/items/neu:193853.

Bänziger, T., Scherer, K. R., Hall, J. A., & Rosenthal, R. (2011a). Introducing the MiniPONS: A short multichannel version of the Profile of Nonverbal Sensitivity (PONS). *Journal of Nonverbal Behavior*, *35*, 189–204. doi:10.1007/s10919-011-0108-3

Baum, K. M., & Nowicki, S., Jr. (1998). Perceptions of emotion: Measuring decoding accuracy of adult prosodic cues varying in intensity. *Journal of Nonverbal Behavior*, *22*, 89–107. doi:10.1023/A:1022954014365

Bernieri, F. J. (1991). Interpersonal sensitivity in teaching interactions. *Personality and Social Psychological Bulletin*, *17*, 98–103. doi:10.1177/0146167291171015

Bernieri, F. J. (2001). Toward a taxonomy of interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 3–20). Hillsdale, NJ: Erlbaum.

Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The interpersonal perception task. *Journal of Nonverbal Behavior*, *13*, 225–245. doi:10.1007/BF00990295

DiMatteo, M. R., Friedman, H. S., & Taranta, A. (1979). Sensitivity to bodily nonverbal communication as a factor in practitioner-patient rapport. *Journal of Nonverbal Behavior*, *4*, 18–26. doi:10.1007/BF00986909

Funder, D. C., & Harris, M. J. (1986). On the several facets of personality assessment: The case of social acuity. *Journal of Personality*, *54*, 529–549. doi:10.1111/j.1467-6494.1986.tb00411.x

Gulabovska, M., & Leeson, P. (2014). Why are women better decoders of nonverbal language? *Gender Issues*, *31*, 202–218. doi:10.1007/s12147-014-9127-9

Hall, J. A. (2001). The PONS test and the psychometric approach to measuring interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 143–160). Hillsdale, NJ: Erlbaum.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior*, *33*, 149–180. doi:10.1007/s10919-009-0070-5

Hall, J. A., Blanch, D. C., Horgan, T. G., Murphy, N. A., Rosip, J. C., & Mast, M. S. (2009). Motivation and interpersonal sensitivity: Does it matter how hard you try? *Motivation and Emotion*, *33*, 291–302. doi:10.1007/s11031-009-9128-2

Hall, J. A., Roter, D. L., Blanch, D. C., & Frankel, R. M. (2009). Nonverbal sensitivity in medical students: Implications for clinical interactions. *Journal of general internal medicine*, *24*, 1217–1222. doi:10.1007/s11606-009-1107-5

Krendl, A. C., & Ambady, N. (2010). Older adults' decoding of emotions: Role of dynamic versus static cues and age-related cognitive decline. *Psychology and Aging*, *25*, 788–793. doi:10.1037/a0020607

Kurkul, W. W. (2007). Nonverbal communication in one-to-one music performance instruction. *Psychology of Music*, *35*, 327–362. doi:10.1177/0305735607070385

Loredana, I., & Duduciuc, A. (2011). Social skills, nonverbal sensitivity and academic success: The key role of centrality in student networks for higher grades achievement. *Revista de Cercetare și Intervenție Socială*, *33*, 151–166.

Mast, M., Jonas, K., & Hall, J. A. (2009). Give a person power and he or she will show interpersonal sensitivity: The phenomenon and its why and when. *Journal of personality and social psychology*, *97*, 835. doi:10.1037/a0016234

Matsumoto, D., & Ekman, P. (1988). *Japanese and Caucasian Facial Expressions of Emotion and Neutral Faces (JACFEE and JACNeuF)*. Human Interaction Laboratory, University of California, San Francisco, CA.

Murphy, N. A., & Hall, J. A. (2011). Intelligence and interpersonal sensitivity: A meta-analysis. *Intelligence*, *39*, 54–63. doi:10.1016/j.intell.2010.10.001

Nowicki, S., Jr., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of Nonverbal Accuracy Scale. *Journal of Nonverbal Behavior*, *18*, 9–35. doi:10.1007/BF02169077

Nowicki, S., & Duke, M. P. (2001). Nonverbal receptivity: The Diagnostic Analysis of Nonverbal Accuracy (DANVA). In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 183–198). Hillsdale, NJ: Erlbaum.

Pitterman, H., & Nowicki, S., Jr. (2004). A test of the ability to identify emotion in human standing and sitting postures: The Diagnostic Analysis of Nonverbal Accuracy-2 Posture Test (DANVA2-POS). *Genetic, Social, and General Psychology Monographs*, *130*, 146–162. doi:10.3200/MONO.130.2.146-162

Riggio, R. E. (2006). Nonverbal skills and abilities. *The Sage handbook of nonverbal communication* (pp. 79–96). Thousand Oaks, CA: Sage.

Rosenthal, R. (Ed.). (1979). *Skill in nonverbal communication*. Cambridge: MA: Oelgeschlager, Gunn, & Hain.

Rosenthal, R., & DePaulo, B. M. (1979). Sex differences in accommodation in nonverbal communication. In R. Rosenthal (Ed.), *Skill in nonverbal communication* (pp. 68–103). Cambridge, MA: Oelgeschlager, Gunn, & Hain.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (2013, January). *Profile of Nonverbal Sensitivity manual*. Retrieved February 8, 2017, from http://libfedora.neu.edu/fedora/objects/neu:193837/datastreams/PDF/content

Rosip, J. C., & Hall, J. A. (2004). Knowledge of nonverbal cues, gender, and nonverbal decoding accuracy. *Journal of Nonverbal Behavior*, *28*, 267–286. doi:10.1007/s10919-004-4159-6

Scherer, K. R. (2007). Component models of emotion can inform the quest for emotional competence. In G. Matthews, M. Zeidner, & R. D. Roberts (Eds.), *The science of emotional intelligence: Knowns and unknowns* (pp. 101–126). New York: Oxford University Press.

Scherer, K. R., & Scherer, U. (2011). Assessing the ability to recognize facial and vocal expressions of emotion: Construction and validation of the Emotion Recognition Index (ERI). *Journal of Nonverbal Behavior*. *35*, 305–326. doi:0.1007/s10919-011-0115-4

Schmid, P. C., Mast, M. S., Bombari, D., & Mast, F. W. (2011). Gender effects in information processing on a nonverbal decoding task. *Sex Roles*, *65*, 102–107. doi:10.1007/s11199-011-9979-3

Tickle-Degnen, L. (1998). Working well with others: The prediction of students' clinical performance. *American Journal of Occupational Therapy*, *52*, 133–142. doi:10.5014/ajot.52.2.133

Wynn, J. K., Sugar, C., Horan, W. P., Kern, R., & Green, M. F. (2010). Mismatch negativity, social cognition, and functioning in schizophrenia patients. *Biological psychiatry*, *67*, 940–947. doi:10.1016/j.biopsych.2009.11.024

Profile 53

Rational-Experiential Inventory-40 (REI-40)

(Pacini & Epstein, 1999)

Profiled by: Shaughan A. Keaton, PhD

Young Harris College

Construct

Cognitive-Experiential Self-Theory (CEST) asserts that individuals process information through two independent but interactive systems, the preconscious experiential system and the conscious rational system (Epstein, 1994). The *rational processing system* is inferential, guided by culturally transmitted rules, characteristically slower, more systematic, primarily verbal, and relatively emotion-free. The *experiential system* is a preconscious system that is more rapid and automatic, holistic, primarily nonverbal, and emotional. The employment of these systems is thought to be partially a function of individual predispositions captured by the Rational-Experiential Inventory (REI-40), which captures four factors underlying these two processing modes: Rational Ability, Rational Engagement, Experiential Ability, and Experiential Engagement (Epstein, Pacini, & Norris, 1998; Pacini & Epstein, 1999).

Instrument Type

Self-Report

Description

The current form of the Rational-Experiential Inventory (REI-40) (Epstein *et al.*, 1998; Pacini & Epstein, 1999) is a 40-item self-report instrument measuring two independent dimensions of human information processing—rational and experiential. Each dimension is assessed using two subscales composed of 10 items each under the factors Rational Ability, Rational Engagement, Experiential Ability, and Experiential Engagement.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

Administered via paper or online, the measure employs a 5-point response scale ranging from 1 (*definitely not true of myself*) to 5 (*definitely true of myself*). The survey takes approximately 10 to 15 minutes to complete.

Scoring

Subscale scores are computed by averaging the 10 composite items. Thus, each respondent receives four scores, one each for Rational Ability, Rational Engagement, Experiential Ability, and Experiential Engagement. Ability and Engagement scores can be further averaged to form two composite scores for Rationality and Experientiality.

Development

A key tenet of CEST is that individuals process information through two parallel, interactive systems: rational and experiential (Epstein, 1991; Epstein, Pacini, Denes-Raj, & Heier, 1996; Pacini & Epstein, 1999). The original REI was developed with this framework in mind (Epstein *et al.*, 1996). The original measure was constructed as a 31-item, self-report instrument with two unipolar scales measuring individual differences in the tendency to employ these two systems. It was composed of 19 items from the Need for Cognition (NFC) scale, representing rational processing, and 12 items from the Faith in Intuition (FI) scale, representing experiential processing.

The REI-40 was developed to address limitations of the original scale, among them a lack of parallel content and internal consistency issues with the NFC items. The NFC items address cognitive activities (engagement), and the FI items refer to making effective intuitive judgments (ability). In addition, the FI items refer to social activities, whereas the NFC items do not. The REI-40 addressed these issues by proposing ability and engagement subscales for each processing model, resulting in two dimensions and four subscales: Rational Ability, Rational Engagement, Experiential Ability, and Experiential Engagement (Epstein *et al.*, 1998; Pacini & Epstein, 1999).

Rational Ability refers to an ability to think logically and analytically. Rational Engagement refers to a reliance on and enjoyment of thinking in an analytical manner. Experiential Ability refers to the ability to trust one's intuition and feelings. Experiential Engagement refers to reliance on and enjoyment of using intuition in decision making.

Reliability

In general, reliability estimates of the REI-40 improved over the original scale, supporting developers' notion that the REI-40 is the preferable scale. The REI-40 has shown evidence of reliability for the two general constructs, Rationality (α ranging from .86 to .91) and Experientiality (α ranging from .87 to .90), and the four subscales: Rational Ability (α ranging from .80 to .85), Rational Engagement (α ranging from .78 to .87), Experiential Ability (α ranging from .77 to .80), and Experiential Engagement (α ranging from .78 to .84).

Validity

The REI began as a construct validation investigation of CEST (Epstein *et al.*, 1996). In Epstein and colleagues' original study, the goal was to develop an individual-difference measure of the rational and experiential processing modes. The 31 items of the original REI were examined with principal component analysis (PCA) across two studies. All items loaded on the appropriate component with item loadings > .30. The NFC and FI scales were not significantly correlated (r = −.07 for the first study and .08 for the second), suggesting that the components are orthogonal. A replication of this procedure with the REI-40 found similar results (Handley, Newstead, & Wright, 2000). Because PCA was utilized, model fit estimates were not provided in either article.

Epstein and Meier (1989) also examined the convergent validity of the original REI, comparing it to the Constructive Thinking Inventory (CTI). NFC was significantly associated with the CTI factors of Global Constructive Thinking, Emotional Coping, and Absence of Negative Overgeneralization and Nonsensitivity. For men, NFC also was related to the Behavioral Copying facet of Positive Thinking, which the authors interpreted as men placing greater importance on NFC to determine coping ability. For female participants, Distrust was more strongly related to NFC and FI. NFC and FI also displayed predictive validity, accounting for significant variance in Action Orientation and Conscientiousness. NFC also was significantly correlated with Dominance (r = .39), Modern Racism (r = −.26), Depression (r = −.24), State-Trait Anxiety (r = −.30), Self-Esteem (r = .35), stress in college life (r = −.13), drinking (r = .09), SAT scores (r = .55), and GPA (r = .13). FI was significantly correlated with Dominance (r = .12), Depression (r = −.09), State-Trait Anxiety (r = −.17), Self-Esteem (r = .18), and stress in college life (r = −.11).

The REI-40 has shown evidence of convergent and divergent validity. Pacini and Epstein (1999) reported rational thinking as positively correlated with Ego Strength (r = .44), Openness (r = .44), and Conscientiousness (r = .32), and negatively correlated with Neuroticism (r = −.38) and Conservative Ideology (r = −.20). In the same study, the experiential thinking style was positively related to Extraversion (r = .21), Agreeableness (r = .18), Favorable Relationship Beliefs (r = .34), and Emotional Expressivity (r = .27) and negatively correlated with Categorical Thinking (r = −.29), Distrust of Others (r = −.23), and Intolerance (r = −.19).

There is evidence of the reliability and validity of the REI-40 in other languages, including Slovak (Mikusková, Hanák, & Cavojová, 2015) and Swedish (Björklund & Bäckström, 2008).

Although the developers advocate for using the REI-40, Akinci and Sadler-Smith (2013) found evidence for a two-dimensional model (rather than four dimensions) in a study of police organizations.

Availability

The REI-40 (Pacini & Epstein, 1999) is presented here, with permission; the original version is located in the initial article published in the *Journal of Personality and Social Psychology* (Epstein *et al.*, 1996). The instrument is free to use for research purposes.

Sample Studies

The REI has been used in a wide range of research. Ares and colleagues (Ares, Mawad, Giménez, & Maiche, 2014) reported that rational and experiential thinking styles affect consumer food choices and dietary patterns. In one study on schizotypy and beliefs about the paranormal, participants scoring higher on both rational and experiential thinking also scored higher on cognitive aspects of schizotypy and self-efficacy (Wolfradt, Oubaid, Straube, Bischoff, & Mischo, 1999). Further, intuitive thinkers scored highest on interpersonal aspects of schizotypy and interpersonal tolerance of ambiguity. Genovese (2005), in a follow-up study, reported similar findings, and concluded that teachers may transmit paranormal beliefs to their students, suggesting a relationship between social learning and thinking style.

The REI-40 also has been used in a variety of research. Feng and Lee (2010) reported that thinking styles (rational and experiential) had an effect on the perceived quality of supportive messages. In particular, more highly rational individuals were likely to positively respond to advice, whereas those with a stronger experiential thinking style rated emotionally supportive messages as greater in quality. Concerning education, McLaughlin and colleagues (2014), using a sample of student pharmacists, found that rational scores were higher than experiential scores, and that rational scores for students under 30 years of age were significantly higher than for those over 30. Buzdar, Ali, and Tariq (2014) administered an adapted version of the REI for adolescents (REI-A), reporting that religious orientations explained a moderate amount of variance in rational thinking of Hindu and Muslim students; the religious orientation of Christian students affected their rational choices minimally.

Berger, Lee, and Johnson (2003) found that men assign greater importance to more specific, base-rate explanations (of increasing world population) than more general, less specific explanations for both positive and negative accounts of the problem. Women were only likely to favor base-rate explanations for negative explanations. High rationals assessed negative non-base-rate accounts of increasing world population (i.e., less specific, more general) as less important. When asked to produce their own examples, rationals tended to bring forth more specific, base-rate explanations for negative explanations (of increasing world population).

Berger (2005) investigated the effects of rational thinking style and variations on the magnitude of threat escalation (shallow, moderate, and steep increases of campus theft) on people's judgments. Highly rational individuals reacted with less apprehension and judged the problem as less substantial, and their responses demonstrated greater variability in response to the different degrees of campus theft. A follow-up experiment confirmed that "high rationals" are more likely to pay attention to evidence that serves to reduce apprehension.

Critique

The two processing systems proposed by CEST seem related to listening styles. Rational processing types call to mind analytical and critical listeners, whereas experiential types seem representative of relational listeners. The REI-40 could be easily applied to listening contexts. Listening researchers interested in education or social support should consider using the REI-40, as it is solidly based in theory. Although there is evidence of convergent validity for the REI-40, however, investigations into the construct validity reviewed in this profile have not included confirmatory factor analysis for model fit. Future researchers are urged to conduct and report findings of their own confirmatory factor analyses.

References

Akinci, C., & Sadler-Smith, E. (2013). Assessing individual differences in experiential (intuitive) and rational (analytical) cognitive styles. *International Journal of Selection and Assessment*, *21*, 211–221. doi:10.1111/ijsa.12030

Ares, G., Mawad, F., Giménez, A., & Maiche, A. (2014). Influence of rational and intuitive thinking styles on food choice: Preliminary evidence from an eye-tracking study with yogurt labels. *Food Quality and Preference*, *31*, 28–37. doi:10.1016/j.foodqual.2013.07.005

Berger, C. R. (2005). Slippery slopes to apprehension: Rationality and graphical depictions of increasingly threatening trends. *Communication Research*, *32*, 3–28. doi:10.1177/0093650204271397

Berger, C. R., Lee, E.-J., & Johnson, J. T. (2003). Gender, rationality, and base-rate explanations for increasing trends. *Communication Research*, *30*, 737–765. doi:10.1177/0093650203258282

Björklund, F., & Bäckström, M. (2008). Individual differences in processing styles: Validity of the Rational–Experiential Inventory. *Scandinavian Journal of Psychology*, *49*, 439–446. doi:10.1111/j.1467-9450.2008.00652.x

Buzdar, M. A., Ali, A., & Tariq, R. U. H. (2014). Religious orientations as a predictor of rational thinking among secondary school students. *Thinking Skills and Creativity*, *16*, 1–8. doi:10.1016/j.tsc.2014.11.005

Epstein, S. (1991). Cognitive-experiential self-theory: An integrative theory of personality. In R. Curtis (Ed.), *The relational self: Theoretical convergences in psychoanalysis and social psychology* (pp. 111–137). New York: Guilford.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*, 709–724. doi: 10.1037/0003-066X.49.8.709

Epstein, S., & Meier, P. (1989). Constructive thinking: A broad coping variable with specific components. *Journal of Personality and Social Psychology*, *57*, 332–350. doi:10.1037/0022-3514.57.2.332

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of Personality and Social Psychology*, *71*, 390–405. doi: 10.1037/0022-3514.71.2.390

Epstein, S., Pacini, R., & Norris, P. (1998). *The Rational-Experiential Inventory, long form*. Unpublished inventory, University of Massachsetts at Amherst.

Feng, B., & Lee, K. J. (2010). The influence of thinking styles on responses to supportive messages. *Communication Studies*, *61*, 224–238. doi: 10.1080/10510971003604000

Genovese, J. E. C. (2005). Paranormal beliefs, schizotypy, and thinking styles among teachers and future teachers. *Personality and Individual Differences*, *39*, 93–102. doi:10.1016/j.paid.2004.12.008

Handley, S. J., Newstead, S. E., & Wright, H. (2000). Rational and experiential thinking: A study of the REI. In R. J. Riding & S. G. Rayner (Eds.), *International perspectives on individual differences* (Vol. *1*, pp. 97–113). Stamford, CT: Ablex.

McLaughlin, J. E., Cox, W. C., Williams, C. R., & Shepherd, G. (2014). Rational and experiential decision-making preferences of third-year student pharmacists. *American Journal of Pharmaceutical Education*, *78*, 1–6. doi:10.5688/ajpe786120

Mikusková, E. B., Hanák, R., & Cavojová, V. (2015). Appropriateness of two inventories measuring intuition (the PID and REI) for Slovak population. *Studia Psychologica*, *57*, 63–82.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, *76*, 972–987. doi:10.1037/0022-3514.76.6.972

Wolfradt, U., Oubaid, V., Straube, E. R., Bischoff, N., & Mischo, J. (1999). Thinking styles, schizotypal traits and anomalous experiences. *Personality and Individual Differences*, *27*, 821–830. doi:10.1016/S0191-8869(99)00031-8

Scale

Rational-Experiential Inventory–40 (Pacini & Epstein, 1999)

Instructions: Using the following scale, please rate the extent that these items refer to you.

1	2	3	4	5
Definitely true of myself				*Definitely not true of myself*

Rationality scale

Rational Ability

1) I'm not that good at figuring out complicated problems.*
2) I am not very good at solving problems that require careful logical analysis.*
3) I am not a very analytical thinker.*
4) Reasoning things out carefully is not one of my strong points.*
5) I don't reason well under pressure.*
6) I am much better at figuring things out logically than most people.
7) I have a logical mind.
8) I have no problem thinking things through carefully.
9) Using logic usually works well for me in figuring out problems in my life.
10) I usually have clear, explainable reasons for my decisions.

Rational Engagement

11) I try to avoid situations that require thinking in depth about something.*
12) I enjoy intellectual challenges.
13) I don't like to have to do a lot of thinking.*
14) I enjoy solving problems that require hard thinking.
15) Thinking is not my idea of an enjoyable activity.*
16) I prefer complex problems to simple problems.
17) Thinking hard and for a long time about something gives me little satisfaction.*
18) I enjoy thinking in abstract terms.
19) Knowing the answer without having to understand the reasoning behind it is good enough for me.*
20) Learning new ways to think would be very appealing to me.

Experientiality scale

Experiential Ability

21) I don't have a very good sense of intuition.*
22) Using my gut feelings usually works well for me in figuring out problems in my life.
23) I believe in trusting my hunches.
24) I trust my initial feelings about people.
25) When it comes to trusting people, I can usually rely on my gut feelings.
26) If I were to rely on my gut feelings, I would often make mistakes.*
27) I hardly ever go wrong when I listen to my deepest gut feelings to find an answer.
28) My snap judgments are probably not as good as most people's.*
29) I can usually feel when a person is right or wrong, even if I can't explain how I know.
30) I suspect my hunches are inaccurate as often as they are accurate.*

Experiential Engagement

31) I like to rely on my intuitive impressions.
32) Intuition can be a very useful way to solve problems.
33) I often go by my instincts when deciding on a course of action.
34) I don't like situations in which I have to rely on intuition.*
35) I think there are times when one should rely on one's intuition.
36) I think it is foolish to make important decisions based on feelings.*
37) I don't think it is a good idea to rely on one's intuition for important decisions.*
38) I generally don't depend on my feelings to help me make decisions.*
39) I would not want to depend on anyone who described himself or herself as intuitive(-).
40) I tend to use my heart as a guide for my actions.

Note: Labels should be removed and items randomized prior to administration. Items marked with an asterisk (*) should be reverse coding prior to scoring. Subscale scores are computed by averaging the 10 composite items.

Profile 54

Relational Framing

(Dillard, Solomon, & Samp, 1996)

Profiled by: Denise Haunani Solomon, PhD and Sara K. Salmon, MA

The Pennsylvania State University

Construct

Relational framing refers to the process through which people draw inferences about social relations during interpersonal interactions. The relational framing scales (RFSs) were designed to measure aspects of the relational framing process, which includes the relevance of relational frames and the intensity of relational judgments in interpersonal interactions.

Instrument Type

Self-Report

Description

The RFSs measure the relevance and intensity of relational judgments in the context of interpersonal interactions (Dillard & Solomon, 2005; Dillard, Solomon, & Samp, 1996). According to relational framing theory (RFT), relational frames are cognitive structures through which individuals perceive social reality. The theory specifies two bipolar relational frames: *dominance–submissiveness*, which indexes the extent to which one person tries to regulate another's behavior; and *affiliation–disaffiliation*, which is the extent to which one holds the other in high regard. These two frames are thought to exist in opposition to each other, such that one or the other is the prevailing lens through which people make sense of a particular interaction episode. Relational inferences are dependent on involvement in the interaction. *Involvement* is a unipolar dimension

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

capturing the degree to which interaction partners are engaged with one another. Involvement judgments, in combination with the salient relational frame, inform judgments about the intensity of dominance and affiliation conveyed during the interaction.

Administration

The relational framing scales are presented within questionnaires that are typically administered in person, although an online variation has been recently developed (Salmon, 2015). Participants complete the scales in regard to either a scenario depicting an interaction between them and another character or a personal experience they have had involving communicating with another person. Because relational frame relevance judgments may be difficult and unfamiliar, Dillard *et al.* (1996) added an extended example centered on perceptions of tactile surface dimensions (Hollins, Faldowski, Rao, & Young, 1993) to analogize relational dimensions. More recently, Salmon (2015) used examples of social interactions, rather than tactile surfaces, to clarify the relational frame relevance judgment. Following the presentation of the extended example, participants rate the extent to which word pairs are *completely irrelevant* (1) to *completely relevant* (5) to the scenario or recalled situation. To measure relational intensity judgments, participants rate the extent to which a relational quality is *absent* (1) or *present* (5) in that scenario or interaction.

Scoring

There are two subscales indexing the relevance of the dominance–submissiveness and affiliation–disaffiliation frames, and five subscales assessing the intensity of involvement, dominance, submissiveness, affiliation, and disaffiliation. The relevance subscales each contain seven word pairs representing opposite poles of the dimensions. The intensity subscales are each composed of seven items; and the involvement intensity subscale is composed of four items (see the Scale section).

Scores for each subscale can be computed by taking the average of the items within that subscale. Higher scores indicate judgments of greater intensity or relevance. Subscales cannot be combined to form an overall score.

Development

RFT was informed by Burgoon and Hale’s (1984, 1987) research on dimensions of relational communication, which proposed 12 relational themes operationalized by the relational message scale (1987). Dillard *et al.* (1996) argued that people draw upon more abstract frames to organize social inferences and that these dimensions subsume the 12 themes articulated by Burgoon and Hale. Based on first- and second-order confirmatory factor analyses of responses to the Burgoon and Hale items, Dillard *et al.* (1996) created the scales measuring frame relevance to evaluate the impact of interaction goals on the relevance of dominance and affiliation relational judgments; early work assessed the “relevance” of involvement to cohere within the other scale items, but theoretically, the intensity of involvement is the variable of interest. Dillard *et al.* (1996) retained four items per subscale, and subsequent investigations added items to each subscale (Lannutti & Monahan, 2002; Tusing, 2001). Dillard, Solomon, and Palmer (1999) measured the intensity of relational judgments using Burgoon and Hale's (1987) relational message scale. Tusing (2001) amended the intensity measures using the anchors of the frame relevance word pairs, and these items informed the contemporary version of the scales (see Dillard & Solomon, 2005).

Reliability

Internal consistencies for the relevance subscales have ranged from $\alpha = .71$ to .84 for the dominance-submissiveness scale and $\alpha = .66$ to .93 for the affiliation–disaffiliation scale. Coefficient alphas for the involvement subscale have been found to range from $\alpha = .68$ to .87. Estimates of the internal consistency of the intensity subscales have been reported as $\alpha = .79$ for dominance, $\alpha = .69$ for submissiveness, $\alpha = .94$ for affiliation, and $\alpha = .87$ for disaffiliation.

Validity

The factor structure of the intensity items support four distinct intensity subscales: dominance, submissiveness, affiliation, and disaffiliation (Tusing, 2001). In addition, one study that included both intensity and relevance measures concluded, using principle axis factor analysis with varimax rotation, that they are empirically distinct (Tusing, 2001). There has, however, been variation found in the dimensional structure of the relational relevance subscales. The factor structure of the affiliation–disaffiliation and dominance–submissiveness scales has remained consistent in that items have loaded on separate factors (CFA procedures) as expected in all previous investigations (Dillard *et al.*, 1996, 1999; Henningsen, Henningsen, Cruz, & Morrill, 2003; McLaren, Dillard, Tusing, & Solomon, 2014; Salmon, 2015; Solomon, Dillard, & Anderson, 2002; Tusing, 2001). Involvement has been found to either load on the same factor as the affiliation items (Dillard *et al.*, 1999) or to comprise a separate factor (Dillard *et al.*, 1996; Solomon *et al.*, 2002; Tusing, 2001). Salmon (2015) found that the dominance–submissiveness items separated into two first-order factors, one relating to persuasion and the other indexing control. With the inconsistencies that have been observed across context and empirical design, further measurement work is needed to understand the dimensionality of the relevance scales.

Availability

Measures adapted from Dillard *et al.* (1996) and Tusing (2001) are provided here with permission. They are free to use for research purposes.

Sample Studies

In the initial studies employing the relational framing scales, Dillard and colleagues (1996) found that the nature of strategic goals influenced frame relevance in compliance gaining and affiliation episodes. Specifically, the dominance frame was perceived as more relevant than the affiliation frame to scenarios illustrating compliance goals ($r^2 = .17$), and the affiliation frame was perceived as more relevant than the dominance frame to scenarios depicting affinity goals ($r^2 = .08$). Solomon *et al.* (2002) considered how individual differences might affect the relevance of relational frames. Their study found attachment anxiety positively associated with both dominance ($r = .19–.23$) and affiliation ($r = .13–.15$) frame relevance. McLaren *et al.* (2014) explored the effects of message directness and relational context in compliance-gaining attempts. Results showed that the dominance frame was perceived as more relevant to direct messages than indirect messages ($\eta^2 = .06$). The dominance frame also was more relevant than the affiliation frame when messages came from a speaker who was portrayed as historically manipulative or competitive, and the affiliation frame was more relevant than the dominance frame when messages came from a speaker who was portrayed as historically friendly ($\eta^2 = .34$). Applications of the relational framing scales have been used to evaluate relational framing processes in small-group interactions (Henningsen *et al.*, 2003), workplace interactions (Solomon, 2006), and communication between potential sexual partners (Lannutti & Monahan, 2002).

Critique

An assumption of the relational framing scales is that the relevance of a frame and the intensity of relational judgments can be measured via self-report. Although this claim may hold for the intensity of substantive judgments, the activity of a cognitive structure in the process of making relational inferences may be less accessible to subjective reporting. Alternative methodologies, such as implicit association testing, are likely to provide better insight into the relevance of relational frames. Relatedly, most studies to date have employed hypothetical scenarios to assess relational judgments, which may provide an inadequate estimation of how people make relational judgments in actual interaction. The degree to which the scales can be adapted to measure how people make inferences in live interaction should be investigated further. A current study is exploring whether trained coders can reliably rate a set of judgments generated by participants asked to watch their videotaped conversations and respond to the "meaning implied by your partner's utterance" (Vickery, 2016).

Another issue concerns the conflation of relational relevance and intensity judgments. Early research in this program failed to reflect conceptual distinctions within the measures employed, which creates ambiguity for researchers using the scales. In addition, a majority of the studies that make use of the scales employ samples of undergraduate students. Because a person's understanding of relational characteristics may change over the lifespan, the lack of variance in age of participants may influence the witnessed properties of the scales.

References

Burgoon, J. K., & Hale, J. L. (1984). The fundamental topoi of relational communication. *Communication Monographs*, *51*, 193–214. doi:10.1080/03637758409390195

Burgoon, J. K., & Hale, J. L. (1987). Validation and measurement of the fundamental themes of relational communication. *Communications Monographs*, *54*, 19–41. doi:10.1080/03637758709390214

Dillard, J. P., & Solomon, D. H. (2005). Measuring the relevance of relational frames: A relational framing theory perspective. In V. Manusov (Ed.), *The sourcebook*

of nonverbal measures: Going beyond words (pp. 325–334). Mahwah, NJ: Erlbaum.
Dillard, J. P., Solomon, D. H., & Palmer, M. T. (1999). Structuring the concept of relational communication. *Communication Monographs*, *66*, 49–65. doi:10.1080/03637759909376462
Dillard, J. P., Solomon, D. H., & Samp, J. A. (1996). Framing social reality: The relevance of relational judgments. *Communication Research*, *23*, 703–723. doi:10.1177/009365096023006004
Henningsen, M. L. M., Henningsen, D. D., Cruz, M. G., & Morrill, J. (2003). Social influence in groups: A comparative application of relational framing theory and the elaboration likelihood model of persuasion. *Communication Monographs*, *70*, 175–197. doi:10.1080/0363775032000167398
Hollins, M., Faldowski, R., Rao, S., & Young, F. (1993). Perceptual dimensions of tactile surface texture: A multidimensional scaling analysis. *Perception & Psychophysics*, *54*, 697–705. doi:10.3758/bf03211795
Lannutti, P. J., & Monahan, J. L. (2002). When the frame paints the picture: Alcohol consumption, relational framing, and sexual communication. *Communication Research*, *29*, 390–421. doi:10.1177/0093650202029004002
McLaren, R. M., Dillard, J. P., Tusing, K. J., & Solomon, D. H. (2014). Relational Framing Theory: Utterance form and relational context as antecedents of frame salience. *Communication Quarterly*, *62*, 518–535. doi:10.1080/01463373.2014.949387
Salmon, S. K. (2015). *Relational meaning in sexual fantasy disclosures*. Unpublished master's thesis. The Pennsylvania State University, Pennsylvania.
Solomon, D. H. (2006). A relational framing perspective on perceptions of social-sexual communication at work. In R. M. Dailey & B. A. LePoire (Eds.), *Applied interpersonal communication matters: Family, health, and community relations* (pp. 271–298). New York: Peter Lang.
Solomon, D. H., Dillard, J. P., & Anderson, J. W. (2002). Episode type, attachment orientation, and frame salience: Evidence for a theory of relational framing. *Human Communication Research*, *28*, 136–152. doi:10.1111/j.1468-2958.2002.tb00801.x
Tusing, K. J. (2001, November). *Intensity and relevance of relational judgments: Same or different concepts?* Paper presented at the annual meeting of the National Communication Association, Seattle, WA.
Vickery, A. J. (2016). *The relational meaning of person-centered emotional support*. Unpublished doctoral dissertation. Louisiana State University and Agricultural & Mechanical College, Baton Rouge, LA.

Scale

Relational Framing: Relevance (Dillard *et al.*, 1996)

Instructions: You have been given several different kinds of materials—wax paper, sandpaper, velvet, a rubber eraser, and a brick—and asked to feel the surface of each of the different materials. [Your task is to judge the relevance of each word pair to making a judgment about the materials.]

1. Rough/smooth	1	2	3	4	5
2. Loud/quiet	1	2	3	4	5
3. Hard/soft	1	2	3	4	5
4. High-pitched/low-pitched	1	2	3	4	5

Most people would say that the rough/smooth and hard/soft dimensions were relevant to the task and that the loud/quiet and high-pitched/low-pitched dimensions were irrelevant. Note that you are NOT evaluating how rough, smooth, loud, quiet, hard, soft, high-pitched, or low-pitched the surfaces are. Instead, you are indicating whether the dimension defined by the word pair is relevant to evaluating those surfaces. Of course, your judgments might be reversed if the task were to judge sounds rather than surfaces in this example. In that case, the rough/smooth and hard/soft dimensions would be irrelevant, and you would probably rate the loud/quiet and high-pitched/low-pitched sounds as relevant.

The following is a list of word-pairs. Each word-pair represents a **dimension**. We would like you to consider the extent to which the following word-pairs are relevant to the behavior of the other person in the scenario that you just read. Remember, we are not asking how much of these qualities is present, but rather we want to know the extent to which you see the dimensions as relevant to understanding the other person's behavior in the situation."

		Completely Irrelevant		Neutral		Completely Relevant
Dominance–Submissiveness						
1.	Persuade/concede	1	2	3	4	5
2.	Influence/comply	1	2	3	4	5
3.	Controlling/yielding	1	2	3	4	5
4.	Dominance/submission	1	2	3	4	5
5.	Convincing/being convinced*	1	2	3	4	5
6.	Coaxing/giving in*	1	2	3	4	5
7.	Demanding/relenting*	1	2	3	4	5
Affiliation–Disaffiliation						
8.	Liking/disliking	1	2	3	4	5
9.	Attraction/aversion	1	2	3	4	5
10.	Affection/disaffection	1	2	3	4	5
11.	Positive regard/negative regard	1	2	3	4	5
12.	Caring/indifference*	1	2	3	4	5
13.	Fondness/lack of fondness*	1	2	3	4	5
14.	Friendly/unfriendly*	1	2	3	4	5

* Added by Lannutti and Monohan (2002).

Note: Items 1–7 index dominance–submissiveness, and items 8–14 index affiliation–disaffiliation. Labels should be removed and items randomized prior to administration. Scores for each subscale can be computed by taking the average of the items within that subscale. Higher scores indicate judgments of greater intensity or relevance. Subscales cannot be combined to form an overall score.

Relational Framing: Intensity (Tusing, 2001)

Source: Tusing (2001). Reproduced with permission of National Communication Association.

Instructions: The following is a list of words that represent judgments that one can make about social situations. We would like you to consider the extent to which the following qualities are present in the behavior of the other person in the scenario that you just read.

	Absent				Present
Dominance					
1. Persuade	1	2	3	4	5
2. Influence	1	2	3	4	5
3. Controlling	1	2	3	4	5
4. Dominance	1	2	3	4	5
5. Convincing*	1	2	3	4	5
6. Coaxing*	1	2	3	4	5
7. Demanding*	1	2	3	4	5
Submissiveness					
8. Concede	1	2	3	4	5
9. Comply	1	2	3	4	5
10. Yielding	1	2	3	4	5
11. Submission	1	2	3	4	5
12. Being convinced*	1	2	3	4	5
13. Giving in*	1	2	3	4	5
14. Relenting*	1	2	3	4	5
Affiliation					
15. Liking	1	2	3	4	5
16. Attraction	1	2	3	4	5
17. Affection	1	2	3	4	5
18. Positive regard	1	2	3	4	5
19. Caring*	1	2	3	4	5
20. Fondness*	1	2	3	4	5
21. Friendly*	1	2	3	4	5
Disaffiliation					
22. Disliking	1	2	3	4	5
23. Aversion	1	2	3	4	5
24. Disaffection	1	2	3	4	5
25. Negative regard	1	2	3	4	5
26. Indifference*	1	2	3	4	5
27. Lack of fondness*	1	2	3	4	5
28. Unfriendly*	1	2	3	4	5

Involvement

29.	Involved	1	2	3	4	5
30.	Active	1	2	3	4	5
31.	Engaged	1	2	3	4	5
32.	Connected**	1	2	3	4	5

* Added by Lannutti and Monohan (2002).
** Added by Tusing (2001).

Note: Items 1–7 index dominance, 8–14 index submission, 15–21 index affiliation, 22–28 index disaffiliation, and 29–32 index involvement. Labels should be removed and items randomized prior to administration. Scores for each subscale can be computed by taking the average of the items within that subscale. Higher scores indicate judgments of greater intensity or relevance. Subscales cannot be combined to form an overall score.

Tusing (2001) also included four items to assess a lack of involvement: *uninvolved*, *inactive*, *withdrawn*, and *disconnected*. Solomon *et al.* (2002) recommended that two additional involvement items that had been previously used, *interested* and *disinterested*, be deleted from the scale because of conflation with affiliation and disaffiliation.

Profile 55

Rhetorical Sensitivity Scale (RHETSEN)

(Hart, Carlson, & Eadie, 1980; Eadie & Powell, 1991)

Profiled by: Shaughan A. Keaton

Young Harris College

Construct

Rhetorical Sensitivity (RS) "is a particular attitude toward encoding spoken messages" (Hart, Carlson, & Eadie, 1980, p. 2). In the most general sense, it refers to the (a) willingness of individuals to adapt their message to an audience and/or (b) the ability to maneuver their interaction partners to adapt to them.

Instrument Type

Self-Report

Description

RS is a relatively stable attitude toward communication that "makes effective social interaction manifestly possible" (Hart & Burks, 1972, p. 75). The original construct was defined as involving five sets of beliefs that "contribute to a rhetorical view of interpersonal encounters" and can be summarized as rhetorically sensitive persons knowing not only "what should be said … [but also] how to say it" (Hart, Carlson, & Eadie, 1980, p. 1). In particular, the five beliefs that underlie rhetorical sensitivity are: (a) an acceptance of individual complexity and diversity, (b) avoidance of scripted behavior, (c) willingness to adapt to the roles of self and interaction partners, (d) recognizing when to speak and when to stay silent, and (e) a commitment to determining the most effective way to state a message (Hart & Burks, 1972). Central to the theoretical understanding of RS is that it places importance on flexibility and adaptability.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Hart *et al.* (1980) maintained that the primary operationalization of rhetorical sensitivity, the Rhetorical Sensitivity (RHETSEN) Scale, is not a competency measure but is designed to distinguish between ideas and feelings individuals apply to communicative scenarios. An individual's level of rhetorical sensitivity as assessed by the RHETSEN is represented by a continuum anchored by the Noble Self (NS) and Rhetorical Reflector (RR). As originally conceived, Rhetorically Sensitive (RS) sits in between the extreme orientations of NSs and RRs. NSs are described as those who see variation from their personal norms as hypocritical and lacking in integrity; it is a more rigid style. In contrast to NSs, RRs are more adaptive, preferring to enact a different identity for each communicative scenario. The RS individual balances personal concerns against those of others in a given situation. People are multifaceted, adaptable communicators who try to elicit certain responses from an equally multifaceted and adaptive other (Ward, Bluman, & Dauria, 1982). In a later conceptual formulation (Eadie & Powell, 1991), RS was strongly and negatively related to NS and negatively but less strongly related to RR.

Administration

The RHETSEN Scale can be administered via paper or online. Participants are asked to indicate to what extent each statement applies to them in general using a 5-point scale that ranges from 1 (*YES!*) to 5 (*NO!*). The scale takes approximately 10 minutes to complete.

Scoring

The items are summed for each orientation, producing a score for each (Hart *et al.*, 1980). These scores can be used in correlational analysis, or it can be used to identify people who are "pure" styles and those who score high or low on two or three of the styles.

Using data from 3023 US college students sampled from 49 four-year institutions, Hart *et al.* reported the following means and SDs – RS ($M = 31.8$, $SD = 7.5$), NS ($M = 15.1$, $SD = 6.3$), RR ($M = 7.0$, $SD = 3.8$) – and the resulting cutoff scores: classic RS = RS ≥ 32, NS < 15, RR < 7; classic NS = RS < 32, NS ≥ 15, RR < 7; classic RR = RS < 32, NS < 15, RR ≥7. They further noted that less than 50% of the sample could be categorized as a "pure" style (RS, 26.9%; NS, 14.5%; RR, 6.7%).

Development

When they introduced the concept of rhetorical sensitivity, Hart and Burks (1972) were attempting to extend rhetorical thought into everyday interpersonal communication encounters. Particularly relevant for listening is the fact that this concept provided for "an active individual other, for whom invention and adaptation must include personal and social, as well as logical, considerations" (Ward *et al.*, 1982, p. 189). In other words, people interact with others who are not mere passive recipients of information but active construers of that information. As such, speakers are required to adapt appropriately to other people in specified contexts or maneuver their interaction partners to adapt to them.

The RHETSEN was first developed as a 40-item scale (Hart *et al.*, 1980) to operationalize the concept of rhetorical sensitivity first discussed by Hart and Burks (1972). The original scale reflected all five of the characteristics that define this trait (see above). Rhetorical sensitivity evolved into a continuum posing the Noble Self (expressive, more rigid) on one end of the spectrum, the Rhetorical Reflector (not as expressive, more adaptive) on the other end, and the Rhetorically Sensitive in the middle (Darnell & Brockriede, 1976).

The original conceptualization of rhetorical sensitivity emerged as a concept that favored confrontation, disclosure, and directness (Hart & Burks, 1972). The original scale, however, did not account for situational limitations or individual differences—it conflated mental states and behaviors. These issues led to the development of a revised scale (RHETSEN2). This newer version is a more balanced, 30-item scale measuring three independent attitudes toward expression, each with 10 items: Rhetorical Sensitivity (RS), Noble Self (NS), and Rhetorically Reflective (RR) (Eadie & Powell, 1991).

Reliability

Early examinations (Eadie & Powell, 1986; Knutson, Komolsevin, Chatiketu, & Smith, 2003) of the RHETSEN provided evidence of modest internal consistency: RS ($.71 < \alpha < .86$), NS ($.70 < \alpha < .81$), and RR ($.54 < \alpha < .69$). Later studies using the RHETSEN2 (Bertoncino, 2010; Knutson *et al.*, 2003) reported similar, if not lower, internal consistency estimates for the three subscales: RS ($.63 < \alpha < .71$), NS ($.71 < \alpha < .76$), and RR ($.56 < \alpha < .66$).

Validity

Knutson *et al.* (2003) submitted the RHETSEN items to a principal component analysis (PCA) and found that no component accounted for a substantial amount of item variance, nor did the items load onto the three anticipated dimensions (rotation method not specified by study authors). Applying the same method to the RHETSEN2 produced the predicted component structure. Although PCA is more appropriate for data reduction and is not technically a factor analytic technique (Tabachnick & Fidell, 2007), these results provide tentative construct validity evidence for the revised scale. The RHETSEN2 was reported as moderately associated with measures of communication apprehension, interaction involvement, and conversational sensitivity (Eadie & Powell, 1991), displaying modest evidence for convergent validity.

In the Hart *et al.* (1980) report, the authors utilized data from 96 nursing students who completed the RHETSEN and were rated by one of five nursing instructors. Ratings and self-reported sensitivity were only moderately correlated with many associations not reaching conventional levels of statistical significance ($-.16 \geq r \leq .42$; $r_{ave} = .08$). Thus, self- and other-perceptions of sensitivity are not isomorphic, calling into question the construct validity of the scale as a measure of actual sensitivity. To date, one study has reported a relation between a 28-item version of the RHETSEN and observed "effective communication behaviors" of students enrolled in an athletic training program who engaged with standardized patients as part of a class activity (Bertoncino, 2010).

Unfortunately, the RHETSEN was scored as unidimensional, making any conclusions about this relationship questionable. No other behavioral data have been reported to date, nor has the scale been used in conjunction with standard measures of listening.

Availability

The RHETSEN is available in the original article by Hart, Carlson, and Eadie (1980). The RHETSEN2 is presented here (Eadie & Powell, 1991), with permission, and is free to use for research purposes.

Sample Studies

The RHETSEN was used in several studies prior to the introduction of the revised scale. Eadie and Paulson (1984) examined the RHETSEN for its ability to distinguish among the three attitudes toward communication. They found that Noble Selves (NS) were more impression leaving, dominant, and less friendly. Hart *et al.* (1980) found pure NS types to be typically younger, politically liberal individuals more likely to reside in the northeastern United States. NS types also have a persuasive communication style that is straightforward and direct, which they typically use to gain power and push for compliance (Eadie & Powell, 1986). In contrast, pure Rhetorical Reflectors (RR) tended to be older, more politically conservative and religious individuals residing in the southern or rural western United States. RR types are more sensitive individuals concerned about relationships and adaptation, placing importance on the needs of others (Knutson *et al.*, 2003). Greenwade (2007) surmised that NS, RS, and RR archetypes would have differing love styles, but these hypotheses were unsupported. Other research suggests Noble Selves tend to be more masculine, Rhetorical Reflectors more feminine, and Rhetorical Sensitives more androgynous (House, Dallinger, & Kilgallen, 1998).

The RHETSEN and the RHETSEN2 have been translated into several languages. In one early study, Ting-Toomey (1988) reported that French communicators preferred a more direct style, Japanese preferred a more moderate rhetorical style, and US Americans preferred a more indirect style. Comparing US Americans and Taiwanese college students, Knutson, Smith, Han, and Hwang (2002) reported that US participants had higher levels of RS and RR than Taiwanese; however, the higher level of RS was opposite of the hypothesized direction. Knutson *et al.* (2003) found the sampled US Americans displayed significantly higher levels of RS and NS than Thais. Thais tended toward higher RR than US Americans. More recently, Dilbeck and McCroskey (2009), using the THAIRHETSEN, found NS was positively related to assertiveness, RS positively correlated to communication competence, and RS negatively associated to NS—although all associations were small.

Critique

Although rhetorical sensitivity is included in a number of communication textbooks, relatively little research has addressed the construct. As seen in this profile, the measure has several empirical and conceptual difficulties (Eadie & Powell, 1997). In the studies that do exist, little evidence of psychometric stability is offered, and there is no direct evidence that self-reported sensitivity results in behavior that is more sensitive to others. The cause for the general lack of interest in this measure is unclear but may be related to the minor results typically reported in assessments of individual differences (Argyle & Little, 1972). The developers of the scale recommended that when measuring rhetorical sensitivity against other communicative and psychological phenomena to use scales with strong evidence of validity and reliability and to have a narrow study focus (Eadie & Powell, 1997). The RHETSEN developers also recommended using the RHETSEN2 (even though a great bulk of the existing research uses the RHETSEN) due to its more balanced form and more extensive reliability and validity portfolio. Notably, Knutson *et al.* (2002) urge caution when using the RHETSEN2 across cultures, suggesting that the measure may be culture specific. Indeed, the studies reporting differences between people of varying nationalities do not report equivalence testing, and so results may be due to different factor structures than true group differences.

Rhetorical sensitivity is a construct that is important in communication and listening. Those leaning toward Rhetorical Reflection seem to tend toward listening types (i.e., relational listeners) who are concerned with interpersonal relationships and the needs of others, whereas Noble Selves seem more intent on engagement and the tasks of power acquisition and compliance (i.e., task-oriented listeners). The construct appears to resonate with many communication and listening scholars, as evidenced in the many publications referencing the construct. Additional research is needed to more fully establish the reliability and validity of the scale.

References

Argyle, M., & Little, B. R. (1972). Do personality traits apply to social behaviour? *Journal for the Theory of Social Behaviour*, *2*, 1–33.

Bertoncino, T. K. (2010). *Linking rhetorical sensitivity with the ability of an athletic training student to succesfully perform a patient medical interview*. Unpublished doctoral dissertation. University of Kansas, Lawrence.

Darnell, D. K., & Brockriede, W. (1976). *Persons communicating*. Englewood Cliffs, NJ: Prentice-Hall.

Dilbeck, K. E., & McCroskey, J. C. (2009). Socio-communicative orientation, communication competence, and rhetorical sensitivity. *Human Communication*, *12*, 255–266.

Eadie, W. F., & Paulson, J. W. (1984). Communicator attitudes, communicator style, and communication competence. *Western Journal of Speech Communication*, *48*, 390–407. doi:10.1080/10570318409374172

Eadie, W. F., & Powell, R. G. (1986). *Rhetorical sensitivity and persuasive communication behavior*. Paper presented at the annual meeting of the Speech Communication Association, Chicago, IL.

Eadie, W. F., & Powell, R. G. (1991). *RHETSEN2: A new measure of rhetorical sensitivity*. Paper presented at the annual Western States Communication Association convention.

Eadie, W. F., & Powell, R. G. (1997). *The utility of rhetorical sensitivity as a contemporary communication concept*. Paper presented at the annual meeting of the National Communication Association, Chicago, IL.

Greenwade, L. (2007). *The relationship between love styles, rhetorical sensitivity, and self-monitoring*. Paper presented at the annual meeting of the National Communication Association, Chicago, IL.

Hart, R. P., & Burks, D. M. (1972). Rhetorical sensitivity and social interaction. *Communication Monographs*, *39*, 75–91. doi:10.1080/03637757209375742

Hart, R. P., Carlson, R. E., & Eadie, W. F. (1980). Attitudes toward communication and the assessment of rhetorical sensitivity. *Communication Monographs*, *47*, 1–22. doi:10.1080/03637758009376016

House, A., Dallinger, J. M., & Kilgallen, D. L. (1998). Androgyny and rhetorical sensitivity: The connection of gender and communicator style. *Communication Reports*, *11*, 11–20. doi:10.1080/08934219809367681

Knutson, T. J., Komolsevin, R., Chatiketu, P., & Smith, V. R. (2003). A cross-cultural comparison of Thai and US American rhetorical sensitivity: Implications for intercultural communication effectiveness. *International Journal of Intercultural Relations*, *27*, 63–78. doi:10.1016/S0147-1767(02)00060-3

Knutson, T. J., Smith, V. R., Han, P-C, & Hwang, J. C. (2002). A comparison of Taiwanese and Amerian samples on rhetorical sensitivity and conflict style. In G.-M. Chen & R. Ma (Eds.), *Chinese conflict management and resolution* (pp. 149–162). Westport, CT: Ablex.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson.

Ting-Toomey, S. (1988). Rhetorical sensitivity style in three cultures: France, Japan, and the United States. *Communication Studies*, *39*, 28–36. doi:10.1080/10510978809363232

Ward, S. A., Bluman, D. L., & Dauria, A. F. (1982). Rhetorical sensitivity recast: Theoretical assumptions of an informal interpersonal rhetoric. *Communication Quarterly*, *30*, 189–195. doi:10.1080/01463378209369448

Scale

Rhetorical Sensitivity Scale (RHETSEN2) (Eadie & Powell, 1991)

Source: Eadie and Powell (1991). Reproduced with permission of William F. Eadie and Robert G. Powell.

Instructions: For each of the following items, please indicate the degree to which it represents your typical attitudes, beliefs, or practices by choosing the number that corresponds to the appropriate response. For instance, if the statement represents you VERY WELL, choose 1. If the statement represents you NOT AT ALL, choose 5.

	1	2	3	4	5	
YES!	|	|	|	|	|	NO!

<u>Rhetorically Sensitive Items</u>

1) Others have told me that I communicate well with difficult people.
2) In an argument, I can get my point across without hurting my relationship with the other person.
3) As a child I communicated easily with adults.
4) I'm good at figuring out the meanings behind what others say.

5) I value my ability to adapt when faced with various communication situations.
6) In group situations, I enjoy offering my opinions about the topic of the discussion.
7) I have been told that I am able to give criticism in a way that does not hurt others.
8) I can disagree with someone in a way that does not damage our relationship.
9) Most of the conflicts I have with others are resolved to everyone's satisfaction.
10) More than a few times I've been told that I communicate well in difficult situations.

Noble Self Items

11) Other people think that I am too direct with them.
12) More than occasionally, I am honest to the point of being blunt when communicating with others.
13) Most of the time I express my opinions, even if they bother others.
14) I like to talk according to my own beliefs, no matter what others may think.
15) Others have often been uncomfortable because I have not hidden my opinions from them.
16) When I disagree with others, I find it difficult to back down.
17) When asked for an opinion, I tend to say the first thing that comes to mind.
18) Regardless of the consequences, I tell my friends what I think.
19) Others have been known to react negatively because I am too honest with them.
20) I like to tell others exactly what I feel.

Rhetorical Reflector Items

21) In conversations, pleasing the other person is an important goal for me.
22) I avoid others rather than risk saying something that might hurt them.
23) I prefer to go along with others rather than openly disagree with them.
24) I feel uncomfortable when people argue in public.
25) In conversations I am most successful when I am able to please the other person.
26) I would rather say nothing than say something that another wouldn't like to hear.
27) If what I would like to say might make others uncomfortable, I keep quiet instead.
28) I usually feel uncomfortable when faced with persuading others.
29) More than occasionally, I have felt that others take advantage of me when we communicate.
30) If at all possible I try to avoid arguing with others.

Note: Labels should be removed and items randomized prior to administration. Items are summed for each of the above areas: rhetorical sensitive, noble self, and rhetorical reflector.

Profile 56

Role Category Questionnaire (RCQ)

(Crockett, 1965)

Profiled by: Andrea J. Vickery, PhD

University of Richmond

Construct

The Role Category Questionnaire (RCQ) was designed to measure the system of personal constructs people use to categorize and describe other people.

Instrument Type

Cognitive Assessment

Description

The RCQ is a measure of how people differentiate, articulate, and integrate cognitive processes related to communication. People use constructs to structure experiences and events, grouping events based on similarities and differences (Delia, O'Keefe, & O'Keefe, 1982; Kelly, 1963). Constructs are "transparent patterns or templates" represented as bipolar dimensions (Kelly, 1963, p. 9). People develop and use constructs to understand phenomena in a variety of domains; the primary focus of the RCQ is on interpersonal constructs, or "the thoughts, behaviors, characteristics, and qualities of other people" (Burleson & Waltman, 1988, p. 3).

Participants completing the RCQ provide impressions of people. These free-response data contain a sample of constructs drawn from the population of interpersonal constructs available for the cognitive domain of interpersonal relationships. People with high scores on the RCQ are said to have a high level of *interpersonal cognitive complexity* (ICC), whereas people scoring lower are said to be cognitively simple in the interpersonal domain.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

The RCQ takes approximately 15 minutes to administer. The required materials include a writing utensil, blank copies of the RCQ, and a timer. Prior to starting the RCQ, participants are instructed to think of two people who are their age—one who they know well and like and one who they know well and dislike. The detailed instructions preceding each writing task ask participants to focus on the habits, mannerisms, and characteristics of each person rather than demographics and other surface-level features (e.g., hair color and height). Participants record their thoughts for each individual on a provided sheet and are allowed to use the back of the page if needed. Participants are timed, writing 5 minutes per person. Combined, the free responses generate impressions of two people. Crockett (1965) provided additional prompts that can be used to extend the two-role RCQ into specific descriptions of people (e.g., liked male peer, disliked male peer, liked female peer, disliked female peer, older male, and older female).

There are two variations in administering the RCQ. The first variation involves recording participants while they orally describe the two individuals; researchers are permitted to prompt participants as necessary to ensure targets are fully described. Participants can also be given an untimed version of the RCQ. Specific recommendations are made in Burleson and Waltman (1988), who outlined factors under which researchers should consider employing an alternative administration.

Scoring

The most common scoring method for the RCQ is construct differentiation, where coders identify and tabulate the number of unique constructs generated in the free-response impressions. Training coders should take approximately 2 hours. The session should permit coders to (a) read through sample responses, preferably from the same or a similar dataset; (b) read and discuss the coding rules with a trained coder; (c) work through examples of coding rules; (d) work through a sample set together; and, finally, (e) work independently to establish intercoder reliability (see Burleson & Waltman, 1988; Crockett, Press, Delia, & Kenney, 1974).

There are six primary rules in scoring impressions for differentiation. For instance, if a person is described as "'domineering, assertive, and aggressive,' all three are scored as constructs because these are similar but not identical concepts (Burleson & Waltman, 1988, p. 26). Not all phrases with multiple words represent distinct constructs; phrases with adjectives or adverbs modifying a phrase, repeated phrases or words, and idiomatic sayings are scored as single constructs. Phrases about the age, physical appearance, or social role(s) of the peer are excluded because these phrases are not appropriate for the nature of the task; they do not represent constructs as bipolar dimensions. The total number of unique constructs listed for each impression should be independently summed. Then, these totals should be added together (across both descriptions) to generate an index of construct differentiation. Lower scores indicate fewer constructs available and relatively simple systems for describing people; higher scores indicate more constructs available and relatively more complex systems for describing people.

The RCQ also can be scored for construct abstractness (for procedures, see Burleson, 1984; Delia, Clark, & Switzer, 1974) and construct system organization (for procedures, see Crockett *et al.*, 1974); both methods have resulted in additional reliability estimates not specified in this profile.

Development

The original RCQ was developed by Walter H. Crockett (1965) based on George A. Kelly's theory of personality and Heinz Warner's developmental perspective to operationalize the concept of cognitive complexity as it influences impression formation and interactions with others. The original measure featured eight roles. The two-role version of the RCQ described in this profile is the most frequently used version of the RCQ. It was used extensively in the 1970s and 1980s in the work of scholars such as Jesse Delia, Ruth Ann Clark, Barbara O'Keefe, Dan O'Keefe, James Applegate, Brant Burleson, and Susan Kline, who were building a case for the theory of human communication known as Constructivism (Delia *et al.*, 1982).

Reliability

Prior uses of the RCQ have generated intercoder reliability statistics (r_{ICC}) at or above .80. Furthermore, the two parts of the RCQ have been found moderately associated, with *r*s between .43 and .67. One study also reported high test–retest reliability, $r = .84$ (O'Keefe, Shepherd, & Streeter, 1982).

Validity

The RCQ has two items, but it is not truly a multi-item test; discriminant, construct, or external validity is more appropriate to analyze versus measurement model validity (Cronbach & Meehl, 1955). There is evidence for discriminant validity based on the small and statistically unsupported associations between the RCQ and various measures, including affinity seeking and attitude toward communication (Weger & Polcar, 2000), argumentativeness (Hample, Gordy, Sellie, Wright, & Zanolla, 2008), and communication apprehension and emotional empathy (Burleson & Samter, 1985).

The RCQ has been scrutinized in terms of construct validity. Because the RCQ elicits free-response data, some scholars have argued the RCQ measures loquacity, not construct differentiation (Allen, Mabry, Banski, Stoneman, & Carter, 1990; Beatty & Payne, 1985; Powers, Jordan, & Street, 1979). Studies in which operationalizations of loquacity and construct differentiation were measured independently, however, counter this claim (Burleson, Applegate, & Neuwirth, 1981; Burleson, Waltman, & Samter, 1987). The developmental perspective recognizes that construct systems change and develop with age as individuals have more social experiences that shape their construct systems. Construct validity has also been represented in studies reporting age-related differences where older children generate more interpersonal constructs than younger children (Burleson, Delia, & Applegate, 1992; Delia & Clark, 1977).

The RCQ has been tested in various populations, including adolescents (Delia & Clark, 1977), college students (Burleson & Samter, 1990; Delia, Clark, & Switzer, 1979), parents of adolescents (Burleson *et al.*, 1981), and counselor trainees (Duys & Hedstrom, 2000). Combined, these populations provide evidence for the external validity of the RCQ.

Availability

The two-role version of the RCQ is available in the appendix of an edited chapter by Burleson and Waltman (1988) and is duplicated here, with permission. The copyright is held by Walter Crockett. It is free to use for research purposes.

Sample Studies

The RCQ has been used to investigate various aspects of constructivist theory that "maintains that all social perception processes occur through the application of interpersonal constructs" (Burleson, 1987, p. 310); these processes include message production, message processing, interaction coordination, and social perception. Highly differentiated construct systems are associated with the ability to select more sophisticated comforting and persuasive message strategies (Burleson, 1983; Delia, Kline, & Burleson, 1979), the ability to generate sophisticated comforting messages (Samter & Burleson, 1984), the ability to discern between and evaluate sophisticated comforting messages (Bodie *et al.*, 2011; Samter, Burleson, & Basden-Murphy, 1989), and the ability to discern and organize detailed impressions of conversational partners (Delia *et al.*, 1974; O'Keefe, 1984). There are, of course, many other studies within this theoretical perspective; the studies described here illustrate the importance of the RCQ and interpersonal constructs in advancing constructivist theory.

There are additional studies outside the constructivist perspective that use the RCQ to investigate individual differences in listening attitudes, behaviors, and outcomes. Higher construct differentiation has been associated with increased listening comprehension (Beatty & Payne, 1984), as well as the general tendency to remember conversations (Neuliep & Hazleton, 1986; see Profile 38, Memory for Conversation). Additionally, construct differentiation has been associated with subscales of the Kentucky Comprehensive Listening Test (Sypher, Bostrom, & Seibert, 1989), along with other individual differences impacting listening such as receiver apprehension (Beatty & Payne, 1981; see the Informational Reception Apprehension profile, Profile 24) and communication apprehension (Neuliep & Hazleton, 1985). Recently, the RCQ has been adapted for use in health communication. The Healthy-Unhealthy Other Index (HUHOI) elicits free-form responses about one healthy and one unhealthy person, and these responses capture the differentiation of health constructs through the number of unique health constructs used to describe others (Bodie *et al.*, 2013).

Critique

Prior critiques of the construct validity of the RCQ have resulted in additional studies comparing the RCQ to other measures of written and verbal abilities. Results from those studies suggest the RCQ is distinct from these measures. Because the RCQ measures how people perceive social information, it is important to recognize that "*persons must use words to express constructs*" (Burleson & Waltman, 1988, p. 19; emphasis in original).

The RCQ purports to measure cognitive complexity, and there are other measures of cognitive complexity, including the role construct repertory test (RCRT) used by Bieri (1955); the paragraph completion test (PCT) developed by Schroder, Driver, and Streufert (1967); as well as the Scott test of cognitive complexity (Scott, 1962). O'Keefe and Sypher (1981) conducted tests comparing the RCPT and RCQ and found the measures to be unrelated. The RCQ has been primarily used in communication-focused empirical studies and by scholars operating from the constructivist perspective. Other instruments measuring cognitive complexity tend to be used in other research domains (e.g., psychology).

The RCQ does require a greater time investment when compared to self-report measures, including administration, training coders, and coding the written impressions. If the particular aims and goals of an empirical study would benefit from the assessment of cognitive complexity as represented by construct differentiation, integration, or abstraction, the RCQ is an appropriate measure.

References

Allen, M., Mabry, E. A., Banski, M., Stoneman, M., & Carter, P. (1990). A thoughtful appraisal of measuring cognition using the Role Category Questionnaire. *Communication Reports*, *3*, 49–57. doi:10.1080/08934219009367504

Beatty, M. J., & Payne, S. K. (1981). Receiver apprehension and cognitive complexity. *Western Journal of Communication*, *45*, 363–369. doi:10.1080/10570318109374057

Beatty, M. J., & Payne, S. K. (1984). Listening comprehension as a function of cognitive complexity: A research note. *Communication Monographs*, *51*, 85–89. doi:10.1080/03637758409390186

Beatty, M. J., & Payne, S. K. (1985). Is construct differentiation loquacity? A motivational perspective. *Human Communication Research*, *11*, 605–612. doi:10.1111/j.1468-2958.1985.tb00063.x

Bieri, J. (1955). Cognitive complexity-simplicity and predicive behavior. *Journal of Abnormal and Social Psychology*, *51*, 263–268. doi:10.1037/h0043308

Bodie, G. D., Burleson, B. R., Holmstrom, A. J., McCullough, J. D., Rack, J. J., Hanasono, L. K., & Rosier, J. G. (2011). Effects of cognitive complexity and emotional upset on processing supportive messages: Two tests of a dual-process theory of supportive communication outcomes. *Human Communication Research*, *37*, 350–376. doi:10.1111/j.1468-2958.2011.01405.x

Bodie, G. D., Collins, W. B., Jensen, J. D., Davis, L. A., Guntzviller, L. M., & King, A. J. (2013). The conceptualization and measurement of cognitive health sophistication. *Journal of Health Communication*, *18*, 246–441. doi:10.1080/10810730.2012.727955

Burleson, B. R. (1983). Social cognition, empathic motivation, and adults' comforting strategies. *Human Communication Research*, *10*, 295–304. doi:10.1111/j.1468-2958.1983.tb00019.x

Burleson, B. R. (1984). Age, social-cognitive development, and the use of comforting strategies. *Communication Monographs*, *51*, 140–153. doi:10.1080/03637758409390190

Burleson, B. R. (1987). Cognitive Complexity. In J. C. McCroskey & J. A. Daly (Eds.), *Personality and Interpersonal Communiction* (pp. 305–349). Thousand Oaks, CA: Sage.

Burleson, B. R., Applegate, J. L., & Neuwirth, C. M. (1981). Is cognitive complexity loquacity? A reply to Powers, Jordan, and Street. *Human Communication Research*, *7*, 212–225. doi:.1111/j.1468-2958.1981.tb00570.x

Burleson, B. R., Delia, J. G., & Applegate, J. L. (1992). Effects of maternal communication and children's social-cognitive and communication skills on children's acceptance by the peer group. *Family Relations*, *41*, 264–272. doi:10.2307/585189

Burleson, B. R., & Samter, W. (1985). Individual differences in the perception of comforting messages: An exploratory investigation. *Central States Speech Journal*, *36*, 39–50. doi:10.1080/10510978509363198

Burleson, B. R., & Samter, W. (1990). Effects of cognitive complexity on the perceived importance of communication skills in friends. *Communication Research*, *17*, 165–182. doi:10.1177/009365090017002002

Burleson, B. R., & Waltman, M. S. (1988). Cognitive complexity: Using the role category questionnaire measure. In C. H. Tardy (Ed.), *A handbook for the study of human communication: Methods and Instruments for observing, measuring and assessing communication processes* (pp. 1–35). Norwood, NJ: Ablex.

Burleson, B. R., Waltman, M. S., & Samter, W. (1987). More evidence that cognitive complexity is not loquacity: A reply to Beatty and Payne. *Communication Quarterly*, *35*, 317–328. doi:10.1080/01463378709369697

Crockett, W. H. (1965). Cognitive complexity and impression formation. In B. A. Maher (Ed.), *Progress in experimental personality research* (Vol. *2*, pp. 47–90). New York: Academic Press.

Crockett, W. H., Press, A. N., Delia, J. G., & Kenney, C. J. (1974). *The structural analysis of the organization of written impressions*. Unpublished manuscript, University of Kansas, Lawrence.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302. doi:10.1037/h0040957

Delia, J. G., & Clark, R. A. (1977). Cognitive complexity, social perception, and the development of listener-adapted communication in six-, eight-, ten- and twelve-year-old boys. *Communication Monographs*, *44*, 326–345. doi:10.1080/03637757709390144

Delia, J. G., Clark, R. A., & Switzer, D. E. (1974). Cognitive complexity and impression formation in informal social interaction. *Communication Monographs*, *46*, 299–308. doi:10.1080/03637757409375854

Delia, J. G., Clark, R. A., & Switzer, D. E. (1979). The content of informal conversations as a function of interactants' interpersonal cognitive complexity. *Communication Monographs*, *46*, 274–281. doi:10.1080/03637757909376012

Delia, J. G., Kline, S. L., & Burleson, B. R. (1979). The development of persuasive communication strategies in kindergarteners through twelfth graders. *Communication Monographs*, *46*, 241–256. doi:10.1080/03637757909376010

Delia, J. G., O'Keefe, B. J., & O'Keefe, D. J. (1982). The constructivist approach to communication. In F. E. X. Dance (Ed.), *Human communication theory: Comparative essays* (pp. 147–191). New York: Harper & Row.

Duys, D. K., & Hedstrom, S. M. (2000). Basic counselor skills training and counselor cognitive complexity. *Counselor Education and Supervision*, *40*, 8–18. doi:10.10002/j.1556-6978.2000.tb01785.x

Hample, D., Gordy, C., Sellie, A., Wright, M., & Zanolla, D. (2008). Inventional repetoires and written messages. *Communication Studies*, *59*, 220–234. doi:10.1080/10510970802257630

Kelly, G. A. (1963). *A theory of personality: The psychology of personal constructs*. New York: W. W. Norton.
Neuliep, J. W., & Hazleton, V., Jr. (1985). Cognitive complexity and apprehension about communication: A preliminary report. *Psychological Reports*, *57*, 1224–1226. doi:10.2466/pr0.1985.57.3 f.1224
Neuliep, J. W., & Hazleton, V., Jr. (1986). Enhanced conversational recall and reduced conversational interference as a function of cognitive complexity. *Human Communication Research*, *13*, 211–224. doi:10.1111/j.1468-2958.1986.tb00103.x
O'Keefe, B. J. (1984). The evolution of impressions in small working groups: Effects of construct differentiation. In H. E. Sypher & J. L. Applegate (Eds.), *Communication by children and adults: Social cognitive and strategic processes* (pp. 262–291). Thousand Oaks, CA: Sage.
O'Keefe, D. J., Shepherd, G. J., & Streeter, T. (1982). Role category questionnaire measures of cognitive complexity: Reliability and comparability of alternative forms. *Central States Speech Journal*, *33*, 333–338. doi:10.1080/03637757409375854
O'Keefe, D. J., & Sypher, H. E. (1981). Cognitive complexity measures and the relationship of cognitive complexity to communication. *Human Communication Research*, *8*, 72–92. doi:10.1111/j.1468-2958.1981.tb0057.x
Powers, W. G., Jordan, W. J., & Street, R. L. (1979). Language indices in the measurement of cognitive complexity: Is complexity loquacity? *Human Communication Research*, *6*, 69–73. doi:10.1111/j.1468-2958.1979.tb00292.x
Samter, W., & Burleson, B. R. (1984). Cognitive and motivational influences on spontaneous comforting behavior. *Human Communication Research*, *11*, 231–260. doi:10.1111/j.1468-2958.1984.tb00047.x
Samter, W., Burleson, B. R., & Basden-Murphy, L. (1989). Behavioral complexity is in the eye of the beholder: Effects of cognitive complexity and message complexity on impressions of the source of comforting messages. *Human Communication Research*, *15*, 612–629. doi:10.1111/j.1468-2958.1989.tb00201.x
Schroder, H. M., Driver, M. J., & Streufert, S. (1967). *Human information processing.* New York: Holt, Rinehart & Winston.
Scott, W. A. (1962). Cognitive complexity and cognitive flexibility. *Sociometry*, *25*, 405–414. doi:10.2307/2785779
Sypher, B. D., Bostrom, R. N., & Seibert, J. H. (1989). Listening, communication abilities, and success at work. *Journal of Business Communication*, *26*, 293–303. doi:10.1177/002194368902600401
Weger, H., Jr., & Polcar, L. E. (2000). Attachment style and the cognitive representation of communication situations. *Communication Studies*, *51*, 149–161. doi:10.1080/10510970009388515

Scale

Role Category Questionnaire (Burleson & Waltman, 1988; Crockett, 1965)

Source: Burleson and Waltman (1988) and Crockett (1965).

Note: The RCQ is distributed as three stapled pieces of paper. The first sheet contains general instructions. The second third sheets contain instructions to write about a person known and liked and a person known but disliked, respectively.

Instructions for the cover sheet: Our interest in this questionnaire is to learn how people describe others whom they know. Our concern here is with the habits, mannerisms—in general, with the personal characteristics rather than the physical traits—that characterize a number of different people.

In order to make sure that you are describing real people, we have set down a list of two different categories of people. In the blank space beside each category below, please write the initials, nicknames, or some other identifying symbol for a person of your acquaintance who fits into that category. Be sure to use a different person for each category.

1) A person your own age whom you like. ___________
2) A person your own age whom you dislike. ___________

Spend a few moments looking over this list, mentally comparing and contrasting the people you have in mind for each category. Think of their habits, their beliefs, their mannerisms, their relations to others, and any other characteristics they have which you might use to describe them to other people.

If you have any questions about the kinds of characteristics we are interested in, please ask us.

Instructions for sheet two: Please look back to the first sheet and place the symbol you have used to designate the person in category 1 here ______.

Now describe this person as fully as you can. Write down as many defining characteristics as you can. Do not simply put down those characteristics that distinguish him/her from others on your list, but include any characteristics that he/she shares with others as well as characteristics that are unique to him/her. Pay particular attention to his/her habits, beliefs, ways of treating others, mannerisms, and similar attributes. Remember, describe him/her as completely as you can, so that a stranger might be able to determine the kind of person he/she is from your description. Use the back of this page if necessary. <u>Please spend only about five (5) minutes describing him/her</u>.

This person is:

[At this point, several blanks are provided for the participant to write.]

Instructions for sheet three: Please look back to the first sheet and place the symbol you have used to designate the person in category 2 here ______.

Now describe this person as fully as you can. Write down as many defining characteristics as you can. Do not simply put down those characteristics that distinguish him/her from others on your list, but include any characteristics that he/she shares with others as well as characteristics that are unique to him/her. Pay particular attention to his/her habits, beliefs, ways of treating others, mannerisms, and similar attributes. Remember, describe him/her as completely as you can, so that a stranger might be able to determine the kind of person he/she is from your description. Use the back of this page if necessary. <u>Please spend only about five (5) minutes describing him/her</u>.

This person is:

[At this point, several blanks are provided for the participant to write.]

Profile 57

Self-Perceived Listening Competence Scale (SPLCS)

(Ford, Wolvin, & Chung, 2000; Mickelson & Welch, 2013)

Profiled by: Margarete Imhof, PhD

Gutenberg University, Mainz, Germany

Construct

The Self-Perceived Listening Competence Scale (SPLCS) provides listeners with a self-assessment of their listening competencies in specific contexts (e.g., work environment, family and friends, and education).

Instrument Type

Self-Report

Description

The original Self-Perceived Listening Competence Scale (SPLCS) was based on the typology of listening outlined by Wolvin and Coakley (1996). Wolvin and Coakley specified five purposes for listening: *discriminative* (being sensitive for verbal and nonverbal information from a source), *comprehensive* (understanding and retaining messages), *therapeutic* (supporting a speaker to talk through an issue and showing empathy), *critical* (evaluating messages for logic and credibility), and *appreciative* (listening for the enjoyment and the sensual experience). The SPLCS also included a subscale addressing attending behaviors (see also Wolvin, Berko, & Wolvin, 1999). Following further validation studies (Mickelson & Welch, 2012, 2013), the instrument (SPLCS-R; Mickelson & Welch, 2013, p. 169) now contains subscales to represent the five aforementioned purposes of listening; the attending behaviors subscale was dropped. Each subscale contains four items resulting in a 20-item scale. All items are worded in a way to address active but internally perceived listening behavior.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

Both the original and the revised version of the questionnaire are self-paced self-assessments. They can be administered as a paper-and-pencil test or as an online survey (Mickelson & Welch, 2013). Test takers are instructed to envision themselves as listeners in a specific situation (e.g., work environment, school, friends, and family) and to rate their behavior using five-point Likert scaling. If necessary, participants may go through the questionnaire several times with a different communication context in mind each time. The items should be presented in random order for each administration. Ford *et al.* (2000) noted that the test should be sensitive to differences between context areas and to changes in listening competence across time, such as a learning period of a one-semester course of listening instruction (Ford *et al.*, 2000).

Scoring

Individual scores are calculated as the means of the items pertaining to a subscale. The values, which were determined in the original sample of $N = 469$ students for four communication contexts (Ford *et al.*, 2000), may be used as a rough and tentative guideline for interpreting the scores. Unfortunately, the authors failed to report the standard deviations associated with the means.

Development

The SPLCS was developed within the context of a basic speech communication course for US undergraduate students. Building on the model of listening proposed by Wolvin and Coakley (1996), the authors generated items to represent the patterns of behavior pertaining to different types of listening included in this model and, in addition, to attending behaviors.

Mickelson and Welch (2012, 2013) and Welch and Mickelson (2013) tested the instrument with different target populations. In their initial study, Mickelson and Welch (2012) found low estimates of internal consistency when the instrument was extended to a more general population beyond college undergraduates (in a sample of business-people, for example, consistency coefficients for the scales ranged from .45 to .76). Also, data did not confirm the factor structure that had been theoretically assumed when developing the instrument.

Building on these results, Mickelson and Welch (2013) invited an expert panel to revise the items of the SPLCS using the following guidelines:

> The theoretically derived structure of the questionnaire needs to be retained; items need to be worded in the active voice (rather than passive voice); items describe an internal action on the part of the listener; items should expand the construct to create a larger item pool, which should be statistically analyzed for appropriate scale characteristics. (p. 160)

As a consequence, a large item pool was generated and double-checked for compliance with the construction guidelines. A typical example for the changes that were made is that the authors rewrote all items from "I can (recognize) …" to "I recognize …" They did this consistently, creating a more accessible version of the instrument.

An unidentified type of exploratory factor analysis was conducted to guide the item selection for the final instrument. A questionnaire based on a 5-factor model was created and investigated for validity using an extended sample of working adults. A confirmatory factor analysis (CFA) was conducted that returned satisfactory fit indices consistent with a plausible interpretation of the factor structure. It is not clear, however, from this report why the scale representing Attention Behavior was eliminated in the process.

Reliability

The authors of both the original and the revised version used Cronbach's α to investigate scale reliability (no published data are available for alternative measures of reliability, e.g., test–retest coefficients). Ford *et al.* (2000) found the estimates acceptable, ranging from .74 to .84 across the subscales. Mickelson and Welch's study (2012) yielded considerably lower reliability scores when they administered the original scale to extended samples, including working adults and respondents from a nationally representative sample in the United States. The revised version of the scale, the SPLCS-R (Mickelson & Welch, 2013), generated acceptable results. Reliability estimates ranged from $.80 \leq \alpha \leq .92$ in a US nationally representative sample.

Validity

The SPLCS-R has been empirically tested for factor structure and consistency. Due to the rigorous revision of the items toward active behavior of the listener, it is relatively safe to assume face validity of the instrument. The factor analyses reported by Mickelson and Welch (2012, 2013) suggest that the revised instrument captures the construct of self-perceived listening behavior reasonably well within the limitations of self-report scales, such as self-serving bias, social desirability, and limited access to automatic internal processes (Keaton & Bodie, 2013; Levine, Hullett, Turner, & Lapinski, 2006). Additional validation studies are needed to explore the criterion validity of the instrument, because little is known about the relationship between self-perceived listening competences and objective measures of listening or between self-perceived listening competences and listening competences perceived by a third person.

Availability

Articles presenting the original and revised versions of the SPLCS are readily available to researchers (Ford *et al.*, 2000; Mickelson & Welch, 2013). The SPLCS-R is presented at the end of this profile, with permission. It is free to use for research purposes.

Sample Studies

Ford *et al.* (2000) used the SPLCS to measure the difference between self-perceived listening competences at the beginning and at the end of a semester in which the undergraduate students had worked through listening instruction. For each student, the researchers collected 48 scores (6 scales × 4 contexts × 2 repeated measures). Contrary to expectations, students reported significantly lower listening competences in 11 of the pre-post comparisons at the end of term. In the interpretation of the study, the authors argued that the students started the course with an "inflated sense of listening effectiveness" (Ford *et al.*, 2000, p. 10); they might have developed a better sense of the listening challenges and might have questioned their overconfidence as they had a chance to learn more about the complexity of listening skills. This study may be taken to illustrate the difficulty and the questionable validity of self-perception and self-report.

Johnson and Long (2007) also used the SPLCS to test if students who had completed a basic communication course had increased listening skills from the beginning to the end of the semester. The course was a hybrid addressing both speaking and listening. The instructors followed a common syllabus, which included three class periods of direct listening instruction. A sample of undergraduate college students ($N = 1059$) was administered both the SPLCS and the Watson-Barker Listening Test (WBLT; see Profile 64) as an objective performance indicator. As expected, self-perceived listening performance had increased significantly over the semester; however, this change was not reflected in the objective performance scores. In a similar vein, the authors could not identify any substantial correlations between the SPLCS and WBLT. At the beginning of the semester, the correlation between these measures was statistically similar to zero. The finding that the same measures yielded significant correlations at around $r = .14$ at the end of the semester needs to be interpreted very carefully because the correlations are indeed small, and the significance of the results might be a rather trivial function of the sample size. Overall, this study suggests that the level of self-perceived listening competence, on the one hand, and objective listening performance, on the other hand, would very likely not overlap (see also Bodie, Jones, Vickery, Hatcher, & Cannava, 2014). However, more research is needed because the measure for objective performance, the WBLT, is notorious for poor reliability scores.

Critique

The SPLCS-R may be useful in all situations in which groups of learners need to be stimulated to think about their listening skills. The instrument offers a theoretically derived operationalization of listening, which helps test takers to go beyond a simplistic concept of listening and to understand the variety of behavioral implications of listening. The revised version has been empirically validated through CFA, thus aligning with current recommended practice (Levine *et al.*, 2006). Authors should continue to submit the SPLCS-R to CFA in future work.

As Keaton and Bodie (2013) noted, caution needs to be used when self-report instruments are administered. It is hard to tell how much self-perceived behavior and skills overlap with objective measures. In particular, the actor–observer difference may lead the listener to overestimate his or her listening skills. Results from studies using the original version of the scale appear to support this assumption (Ford *et al.*, 2000; Johnson & Long, 2007).

The instrument is consistent with a behavior-based approach to listening and stimulates the awareness for forms and components of listening. The revised version of the SPLCS has both face validity and initial evidence of construct validity. Further validation studies are necessary, however, in particular to test for criterion validity (i.e., testing the relationship of SPLCS-R scores and those of other objective measures of listening behavior and performance). Mickelson and Welch (2013) also recommend investigating the generalizability of the scale to a variety of distinct populations and contexts (Welch & Mickelson, 2013). Given that the instrument was developed to identify changes in listening skills over an extended period of time, the sensitivity of the instrument for changes and differences within and across individuals needs to be investigated more closely. Finally, the scale should be translated into other languages and tested for cross-cultural variability.

References

Bodie, G. D., Jones, S. M., Vickery, A. J., Hatcher, L., & Cannava, K. (2014). Examining the construct validity of enacted support: A multitrait-multimethod analysis of four perspectives for judging immediacy and listening behaviors. *Communication Monographs, 80*, 495–523. doi:10.1080/03637751.2014.957223

Ford, W. S. Z., Wolvin, A. D., & Chung, S. (2000). Students' self-perceived listening competencies in the basic speech communication course. *International Journal of Listening, 14*, 1–13. doi:10.1080/10904018.2000.10499032

Johnson, D. I., & Long, K. M. (2007). Student gains in the basic communication course: A comparison of self-report and performance-based competence measures. *International Journal of Listening, 21*, 92–101. doi:10.1080/10904010701301990

Keaton, S. A., & Bodie, G. D. (2013). The statistical and methodological acuity of scholarship appearing in the *International Journal of Listening*. *International Journal of Listening, 27*, 115–135. doi:10.1080/10904018.2013.813206

Levine, T., Hullett, C. R., Turner, M. M., & Lapinski, M. K. (2006). The desirability of using confirmatory factor analysis on published scales. *Communication Research Reports, 23*, 309–314. doi:10.1080/08824090600962698

Mickelson, W. T., & Welch, S. A. (2012). Factor analytic validation of the Ford, Wolvin, and Chung Listening Competence Scale. *International Journal of Listening, 26*, 29–39. doi:10.1080/10904018.2012.639646

Mickelson, W. T., & Welch, S. A. (2013). Improving the performance of the Listening Competency Scale: Revision and validation. *International Journal of Listening, 27*, 157–171. doi:10.1080/10904018.2013.821355

Welch, S. A., & Mickelson, W. T. (2013). A listening competence comparison of working professionals. *International Journal of Listening, 27*, 85–99. doi:10.1080/10904018.2013.783344

Wolvin, A. D., Berko, R. M., & Wolvin, D. R. (1999). *The public speaker/the public listener*. Los Angeles: Roxbury.

Wolvin, A. D., & Coakley, C. G. (1996). *Listening*. Dubuque, IA: Brown & Benchmark.

Scale

Self-Perceived Listening Competency Scale–Revised (Mickelson & Welch, 2013)

Source: Mickelson and Welch (2013). Reproduced with permission of Taylor & Francis.

Instructions: Envision yourself in your (researcher specified) environment. Using the following scale, rate the extent to which you engage in each behavior presented in these statements:

1 = *strongly disagree* 2 = disagree 3 = neither agree/disagree 4 = agree 5 = *strongly agree*

<u>Discriminative Listening</u>

1) I identify someone's feelings when s/he is speaking to me.
2) I recognize when someone is not telling the truth.
3) I interpret someone's facial expression.
4) I recognize when someone is withholding information from me.

<u>Therapeutic Listening</u>

5) I listen patiently when someone is upset.
6) I give someone time to express their feelings.
7) I am understanding with someone who is upset.
8) I encourage people to share their feelings with me.

<u>Critical Listening</u>

9) I evaluate a message on how the person develops his/her line of reasoning.
10) I critically evaluate the content of information that is presented to me.
11) I carefully assess information as it is being shared with me.
12) I determine if a person has a credible message.

<u>Comprehensive Listening</u>

13) I correctly recall information after hearing it.
14) I correctly recall information a few minutes after I hear it.
15) I correctly construct a person's message after I hear it.
16) I understand messages I have just heard.

<u>Appreciative Listening</u>

17) I appreciate hearing another's point of view.
18) I enjoy listening to other people.
19) I listen with an open mind to what others have to say.
20) I appreciate hearing what others have to say.

Note: Labels should be removed and items randomly ordered prior to administration.

Profile 58

Talkaholic Scale (Compulsive Communication Scale) (TAS)

(McCroskey & Richmond, 1993)

Profiled by: Debra L. Worthington, PhD

Auburn University

Construct

Talkaholicism, more typically referenced as compulsive communication (CC), refers to an individual's tendency to talk beyond a socially acceptable norm.

Instrument Type

Self-Report

Description

The 16-item Talkaholic Scale (TAS) measures an individual's self-perceived tendencies to overcommunicate. Ten items measure an individual's self-perceived CC; the remaining six are distractor or filler items. Scores can range between 10 and 50, with higher scores indicating greater compulsiveness. Although scoring techniques have varied somewhat, McCroskey and Richmond (1993, 1995) described those with scores over 40 (2 or more standard deviations above the mean) as compulsive communicators; those between 30 and 39 were described as borderline talkaholics, who occasionally face difficulty controlling their compulsive tendencies.

Administration

The TAS is a self-administered questionnaire that takes approximately 5 minutes to complete. Individuals respond to 16 items using 5-point Likert scaling.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Scoring

As noted above, of the 16 items, only 10 are scored; six items act as distractors. Scoring follows a multistep process: First, scores for items 2, 3, 5, 7, 8, 10, 11, and 14 are totaled; next, scores for items 13 and 16 are summed. The total score is calculated using the following formula: 12 + Total from Step 1 – Total from Step 2. As noted in the Description section, scores range from 10 to 50. Based on normed data collected by McCroskey and Richmond (1993), individuals scoring two standard deviations above the mean (>40) were identified as compulsive communicators.

Development

At least in the United States, people who are willing to express themselves are generally viewed more favorably than their quieter counterparts (Cain, 2013). More talkative individuals are more likely to be viewed as leaders, seen as more competent, and generally viewed more positively. Some individuals, however, engage in excessive communication, negating these benefits.

As a measure of CC, the TAS is based on the premise that some people are driven to excessively talk (McCroskey & Richmond, 1993, 1995). Originally described as *talkaholicism* (suggesting a kinship to the impulsiveness associated with alcoholics and workaholics), CC is typically applied to those who do more than just talk with great frequency (i.e., quantity). It applies to those who express themselves in qualitatively different ways from their noncompulsive counterparts.

McCroskey and Richmond (1993) identified four qualities associated with CC: (a) compulsive high verbalization, (b) self-awareness that talking behavior is perceived as excessive, (c) excessive talking that occurs consistently across a variety of communication contexts, and (d) continued talking even when against their own best interests. Thus, these individuals typically say things others do not want them to say and are perceived to be ineffective communicators. These highly verbal individuals may have great difficulty and little desire to be quiet when with others. McCroskey and Richmond initially included 25 items to measure CC, which when submitted to a principal component analysis (PCA; Kaiser's measure = .90) revealed two components. The first included 12 items representing "talkaholic" behavior. The second composed the remaining items and reflected avoidance behavior. Ten of the 12 items with the highest loadings on the talkaholic factor were selected to make up the TAS and thus to measure CC.

The scale has been used in a number of studies since its introduction. Most of these studies have not, however, tested the dimensionality of the scale.

More recently, an observer measure of CC was developed (TAS-O; Long, Fortney, & Johnson, 2009). The TAS-O provides a means of exploring aspects of CC that individuals may not recognize in themselves. It can also be useful in assessing possible differences between self-report and observer assessments of CC and provides a means of providing feedback to the talkaholic on his or her communication behaviors.

Reliability

Strong internal reliability estimates are consistently reported, with Cronbach's alphas regularly above .85 for US and non-US samples (e.g., McCroskey and Richmond, 1995, reported $\alpha = .92$; Martin & Myers, 2006, reported $\alpha = .95$; for an exception, see

McPherson & Liang, 2007). Both Richmond, McCroskey, McCroskey, and Fayer (2008) and O'Mara, Long, and Allen (2003) reported internal reliability estimates for Spanish versions of the scale of .87. Hackman, Johnson, and Barthel-Hackman (1995) reported a slightly lower reliability estimate with New Zealand students ($\alpha = .74$). McCroskey and Richmond (1993) reported test–retest reliability after a 13-week delay of .76.

O'Mara *et al.* (2003) reported norms across a number of studies and across several cultures (United States, Puerto Rico, and New Zealand), with means ranging from a low of 22.9 ($SD = 8.5$) to a high of 25.7 ($SD = 7.6$) (Ifert, Long, & Fortney, 1998). O'Mara *et al.* also reported that approximately 5 to 6% of the 3228 participants scored as compulsive communicators (i.e., scores > 40). Researchers in at least one study (Ifert *et al.*, 1998) chose to use this percentage to classify participants in their sample as compulsive communicators rather than the cutoff value of 40.

Validity

McCroskey and Richmond (1993) submitted TAS scores to PCA along with items for the CCS and the communication responsiveness scale (Richmond & McCroskey, 1992) and found items loaded on separate components. Early studies reported small to moderate positive correlations between CC scores and assertiveness (.28), neuroticism (.16), and willingness to communicate (.19); negative relationships were found with communication apprehension (−.24), introversion (−.29), and shyness (−.62) (McCroskey & Richmond, 1995; see also Fortney, Ifert, & Long, 1998). Similar, albeit weaker, findings were reported with a New Zealand student population (Hackman *et al.*, 1995).

McCroskey, Heisel, and Richmond (2001), in a study of the relation between communication traits and Eysenck's (1947, 1990) Big Three personality structure (i.e., extraversion, neuroticism, and psychoticism), reported that neurotic psychotic extraverts were more likely to be compulsive communicators with a greater acceptance for disagreement.

Bodie (2011) found a weak relationship ($r = .13$) between CC and Active-Empathic Listening-Processing, but not AELS-Sensing or AELS-Responding. Bodie suggested this association may indicate a greater need to talk when actively tracking conversational points to achieve listening goals.

Long *et al.* (2000) reported a weak association ($r = .15$) between TAS scores and TAS-Observer ratings, suggesting the need for more research exploring how best to assess CC.

Sample Studies

Most studies into CC (as measured by the TAS) fall into two broad categories: intercultural and instructional.

Intercultural studies often were designed to extend and build upon the previously reviewed validity portfolio. For example, in results of a study of communication traits on first and second languages, McCroskey *et al.* (2008) suggested that strong traits are not impacted by situational factors, such as moving from one's native language to a second language (in the case of this study, from Spanish to English for bilingual participants). Talkaholicism scores were only slightly higher ($\eta^2 = .03$) for the participants' first language (Spanish, $M = 24.8$, $SD = 8.8$) compared to their second language (English, $M = 21.9$, $SD = 8.0$). In a New Zealand study, Hackman *et al.* (1995) reported similar findings ($M = 24.4$, $SD = 7.6$). They also reported a significant difference between male ($M = 22.9$) and female students ($M = 25.9$), although the effect size, as with most biological sex difference work, was small (3%).

In the instructional context, Ifert *et al.* (1998) examined intravariations in compulsive communicators in a classroom population. More specifically, self-reported compulsive communicators who believed that they were also competent communicators were less apprehensive and scored higher on argumentativeness. In contrast, those who saw themselves as less skilled rated themselves higher on communication apprehension and lower on trait argumentativeness.

Fortney, Johnson, and Long (2001) studied the effect that compulsive communicators may have on their classmates (in a general communication class). Presence of a self-reported compulsive communicator depressed typical increases in self-perceived communication competence (SPCC) scores over the course of a semester. Instructional strategies that move student attention from compulsive communicator peers to a reflection on their personal achievements and improvements can, however, mitigate score depression. Of note, students expect instructors to handle their compulsive peers (McPherson & Liang, 2007), preferably using prosocial strategies. When instructors do, students rate them higher on credibility and view them more favorably. Martin and Myers (2006) reported that CC was not associated with out-of-class communication between students and instructors.

Critique

As seen in this profile, self- and other-reports of talkaholicism are not highly correlated (Long *et al.*, 2009). Thus, differences in perspective may influence the results of other reports. McCroskey and Richmond (1993) noted that a layperson's perceptions of CC may be skewed by the perceived "quality" of the conversation. Thus, a person may be viewed as a talkaholic when it may be that the receiver does not particularly care for what is being said. Other possible influences include personality (i.e., will introverts score individuals differently than extraverts?) and the nature of the interaction (i.e., the topic, and evaluating one vs. multiple conversations).

Additional study into the relation between competency and context is needed. Fortney *et al.*'s (2001) findings suggest that regular contact with someone who is a compulsive communicator can affect personal competency assessment in the classroom context. The question is how these findings may play out in other professional contexts.

Some researchers have suggested that identifying someone as a compulsive communicator based on scale scores is problematic, in part because the scale does not account for related predispositions such as dominance, inhabitation, and frequency of talk (Bostrom, Prather, & Harrington, 1998) or other factors such as predisposition toward verbal communication, self-esteem, and locus of control (Bostrom & Harrington, 1999). Other researchers have suggested the construct may be part of a larger communication addiction disorder and that individuals may deny their behavior (Walther, 1999).

As noted above, Ifert *et al.* (1998) conducted one of the few studies examining intravariations in compulsive communicators. Their findings are a caution for academics conducting research in communication traits to resist the temptation to paint compulsive individuals with a broad "personality" brush.

References

Bodie, G. D. (2011). The Active-Empathic Listening Scale (AELS): Conceptualization and evidence of validity with the interpersonal domain. *Communication Quarterly, 59*, 277–295. doi:10.1080/01463373.2011.583495

Bostrom, R. N., & Harrington, N. G. (1999). An exploratory investigation of the characteristics of compulsive talkers. *Communication Education, 48*, 73–80. doi:10.1080/03634529909379154

Bostrom, R., Prather, M., & Harrington, N. G. (1998, July). *Birth order and self-reports of communicative behavior*. Paper presented at the annual meeting of the International Communication Association, Jerusalem, Israel.

Cain, S. (2013). *Quiet: The power of introverts in a world that can't stop talking*. New York: Broadway.

Eysenck, H. J. (1947). *Dimensions of personality*. New York: Praeger.

Eysenck, H. J. (1990). Biological dimensions of personality. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 244–276). New York: Guilford.

Fortney, S. D., Ifert, D., & Long, K. M. (1998, April). *Transcending boundaries advancing the construct: Differences between compulsive and non-compulsive communicators*. Paper presented at the annual meeting of the Eastern Communication Association, Saratoga Springs, NY.

Fortney, S. D., Johnson, D. I., & Long, K. M. (2001). The impact of compulsive communicators on the self-perceived competence of classroom peers: An investigation and test of instructional strategies, *Communication Education, 50*, 357–373. doi:10.1080/03634520109379261

Hackman, M. Z., Johnson, C. E., & Barthel-Hackman, T. (1995). Correlates of talkaholism in New Zealand: An intracultural analysis of the compulsive communication construct. *Communication Research Reports, 12*, 53–60. doi:10.1080/08824099509362039

Ifert, D. E., Long, K. M., & Fortney, S. D. (1998). Examining variation among compulsive communicators in classroom settings. *Communication Research Reports, 15*, 426–435. doi:10.1080/08824099809362142

Long, K. M., Fortney, S. D., & Johnson, D. I. (2000). An observer measure of compulsive communication. *Communication Research Reports, 17*, 349–356. doi:10.1080/08824090009388783

Martin, M. M., & Myers, S. A. (2006). Students' communication traits and their out-of-class communication with their instructors. *Communication Research Reports, 23*, 283–289. doi:10.1080/08824090600962599

McCroskey, J. C., Heisel, A. D., & Richmond, V. P. (2001). Eysenck's BIG THREE and communication traits: Three correlational studies. *Communication Monographs, 68*, 360–366. doi:10.1080/03637750128068

McCroskey, J. C., & Richmond, V. P. (1993). Identifying compulsive communicators: The talkaholic scale. *Communication Research Reports, 10*, 107–114. doi:10.1080/08824099309359924

McCroskey, J. C., & Richmond, V. P. (1995). Correlates of compulsive communication: Quantitative and qualitative characteristics. *Communication Quarterly, 43*, 39–52. doi:10.1080/01463379509369954

McPherson, M. B., & Liang, Y. (2007). Students' reactions to teacher's management of compulsive communicators. *Communication Education, 56*, 18–33. doi:10.1080/0363452060101678

O'Mara, J., Long, K. M., & Allen, J. L. (2003). Compulsive communication and competency: An intercultural perspective. *Journal of Intercultural Communication Research*, *32*, 205–225.
Richmond, V. P., & McCroskey, J. C. (1992). *Communication: Apprehension, avoidance, and effectiveness* (3rd ed.). Scottsdale, AZ: Gorsuch Scarisbrick.
Richmond, V. P., McCroskey, J. C., McCroskey, L. L., & Fayer, J. M. (2008). Communication traits in first and second languages: Puerto Rico. *Journal of Intercultural Communication Research*, *37*, 65–73. doi:10.1080/17475750802533331
Walther, J. B. (1999, August). *Communication addiction disorder: Concern over media, behavior and effects*. Paper presented at the annual meeting of the American Psychological Association, Boston, MA.

Scale

The Talkaholicism/Compulsive Communication Scale (McCroskey & Richmond, 1993)

Source: McCroskey and Richmond (1993). Reproduced with permission of Taylor & Francis.

Read the following questions and select the answer that corresponds with what you would do in most situations. Do not be concerned if some of the items appear similar. Please use the scale below to rate the degree to which each statement applies to you. Use the following responses:

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

1) Often I keep quiet when I should talk.
2) I talk more than I should sometimes.
3) Often, I talk when I know I should keep quiet.
4) Sometimes I keep quiet when I know it would be to my advantage to talk.
5) I am a "talkaholic."
6) Sometimes I feel compelled to keep quiet.
7) In general, I talk more than I should.
8) I am a compulsive talker.
9) I am not a talker; rarely do I talk in communication situations.
10) Quite a few people have said I talk too much.
11) I just can't stop talking too much.
12) In general, I talk less than I should.
13) I am not a "talkaholic."
14) Sometimes I talk when I know it would be to my advantage to keep quiet.
15) I talk less than I should sometimes.
16) I am not a compulsive talker.

Scoring: To determine the score on the Talkaholic Scale, complete the following steps:

Step 1: Add the scores for items 2, 3, 5, 7, 8, 10, 11, and 14.
Step 2: Add the scores for items 13 and 16.
Step 3: Total Score = 12 + Total from Step 1 - Total from Step 2.

Note: Items 1, 4, 6, 9, 12, and 15 are filler items and are not scored. Compulsive Communicators are identified as those scoring 40 or higher. Items should be randomized prior to administration.

Profile 59

Team Listening Environment (TLE)

(Johnston & Reed, 2014)

Profiled by: Michelle K. Johnston, PhD and Kendra Reed, PhD

Loyola University New Orleans

Construct

Team Listening Environment (TLE) refers to individual perceptions of communication behaviors that demonstrate genuine attention and understanding from team members in the workplace (Johnston, Reed, & Lawrence, 2011).

Instrument Type

Self-Report

Description

The TLE scale taps the extent to which an individual perceives his or her team members' communicative actions as supportive (i.e., demonstrating genuine attention and understanding), thereby capturing individual workers' perceptions of how the organization exchanges ideas and shares feedback. The construct fits within a nomological network of organizational design and performance as either an antecedent or a consequence. An individual's score provides a quantitative indicator of his or her affective perception of the TLE within a particular organization or work team.

Administration

The TLE scale is a self-administered, 5-item questionnaire that takes no more than 2 minutes to complete. Participants rate their level of agreement for each item, using 5-point Likert scaling.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Scoring

Scoring consists of calculating the mean of participant responses to the five scale items. Generally, an average score of 4.0 or higher indicates a positively perceived team listening environment.

Development

Although researchers have explored the impact of communication satisfaction on simulated organizational results, some research suggests that the complexity of communication satisfaction may confound results (Johnston, Reed, Lawrence, & Onken, 2007). More specifically, most research into communication satisfaction has not addressed a worker's impression of coworker interactions (e.g., feeling understood). In their early research, Johnston *et al.* (2007) identified a subdimension of communication satisfaction that focused on indicators of others' listening. Johnston *et al.* (2007) operationalized TLE with three items from Hecht's (1978) Group Communication Satisfaction scale: (a) The other group members listened to what I had to say, (b) the other group members understood me, and (c) the other group members seemed to be attentive to what others had to say.

For the first follow-up study, Johnston *et al.* (2011) reviewed current listening inventories in academic journals and consulted with experts in business communication on item inclusiveness and redundancy. Next, they gathered data from a sample of full- and part-time MBA students ($N = 101$). Results suggested that the addition of two items increased the reliability estimate: (a) "The other group members paid attention to me," and (b) "The other group members genuinely wanted to hear my point of view." The conclusive 5-item scale measuring the TLE construct demonstrated reasonable fit via confirmatory factor analysis: CFI = 0.95, RMSEA = 0.09 (CI 90% = 0.03, 0.2).

The second follow-up study (Johnston *et al.*, 2011) employed a sample of 66 employees, representing all levels in a manufacturing company with multiple locations. The results of a CFA with these data again suggested a reasonable fit for the data: CFI = 0.96, RMSEA = 0.09 (CI 90% = 0.03, 0.14).

Reliability

Internal consistency for the five items has reached acceptable levels in all published studies, with Cronbach's alphas > .90 (Johnston & Reed, 2014; Reed *et al.*, 2014) and strong fit indices from CFA (see the Development section) (Johnston *et al.*, 2011; Reed, Goolsby, & Johnston, 2016).

Validity

With a seemingly stable 5-item scale, Johnston *et al.* (2011) pursued empirical evidence for convergent and discriminant validity by comparing the TLE items to items from other listening-related scales, namely, (a) the Listening Styles Inventory (Pearce, Johnson, & Barker, 2003), (b) the Small Group Socialization Scale (Riddle, Anderson, & Martin, 2000), and (c) the Relationship Satisfaction Scale (Anderson, Martin, & Riddle, 2001). Employing a sample of 101 MBA students, the comprehensive measurement model included 18 items that were specified to load on one of four latent constructs. Fit statistics provided mixed support for the model, with the CFI value estimated at 0.88, RMSEA estimated at .073, with the 90% confidence interval ranging from .05 to .08; χ^2 (129) = 225.19, $p < .001$. Estimates of item loadings were all above .50, suggesting each item is a valid indicator of its underlying construct. In general, this study suggested that the items measuring TLE are distinct from those measuring similar constructs.

Later studies used the TLE measure in organizational studies with employee samples and provided empirical evidence that the TLE construct fits within a nomological network of organizational design and performance as an antecedent or a consequence. Johnston *et al.* (2011) utilized structural equation modeling to demonstrate a statistically positive impact of TLE on employee organization commitment. Reed *et al.* (2014), using strong theoretical arguments and regression analysis, revealed TLE to be an antecedent to not only employee organizational commitment but also employee organizational identification. Finally, Johnston and Reed (2014) recently were able to establish a positive relation between TLE and an index of organizational financial performance. Thus, in general, the TLE scale appears to be a useful measure for examining listening while examining the communication processes and behaviors in the workplace.

Availability

The TLE scale is provided at the end of this profile, with permission, and is free to use with appropriate citation.

Sample Studies

The TLE scale has been utilized with a wide range of study participants: (a) undergraduate students, (b) working MBA students, (c) manufacturing employees and management, (d) media employees and management, and (e) retail employees and management.

Empirical studies of TLE with employee samples have consistently supported a positive impact of TLE as an antecedent to financial indicators and employee attachments, as well as TLE being a consequence of management actions. Johnston and Reed (2014) recently uncovered an association between TLE and organizational financial indicators of performance. Specifically, the results showed TLE is positively associated with percentage changes in organizational net income. Using this relative indicator of financial performance provided a particularly salient measure of organizational performance based on a relative (i.e., percentage change) rather than absolute number, thus making comparisons to other organizations more generalizable. Moreover, a relative measure of net income can provide a more robust comparison to companies of different sizes and industries.

Reed *et al.* (2014) found that TLE fueled with organizational information positively relates to organizational commitment and identity. Employees who received attention and understanding from team members felt more attached to their organizations. By creating a culture strong in skilled listening, management can build a solid foundation for sharing ideas (Walters, 2005) and drive outcomes aligned with organizational goals.

Critique

A relatively new measure, the TLE scale shows promise as a scale tapping the listening climate inside organizations. Resulting from a systematic design process, including operational testing on a wide range of employees, the TLE scale can be used by researchers with confidence. Having demonstrated evidence of discriminant and criterion validity, the TLE scale is capable of accurately and reliably capturing phenomena related to a relatively silent subdimension of the organizational communication climate, particularly concerning the willingness of employees and leaders to listen to the concerns of team members. The scale has been shown in empirical studies to be predictive of performance and attitudes in organizations, as well as yielding practical results that can be operationalized into policy changes.

A key limitation of the TLE scale is its recent development and application across a limited sample of four unique organizations. Initial results do, however, suggest the measure may act as a key predictor of organizational performance and climate and team member attitudes and beliefs. Additional study across a wide range of organizations will provide evidence to support this belief. More research is needed to investigate whether management can foster a TLE in different types of organizations and among diverse types of organizational members. Also, an association between the TLE scale and more established measures of organizational culture needs to be established.

References

Anderson, C. M., Martin, M. M., & Riddle, B. L. (2001). Small group relational satisfaction scale: Development, reliability and validity. *Communication Studies*, *52*, 220–233. doi:10.1177/104649640003100503

Hecht, M. L. (1978). The conceptualization and measurement of interpersonal communication satisfaction. *Human Communication Research*, *4*, 253–264. doi:10.1111/j.1468-2958.1978.tb00614.x

Johnston, M. K., & Reed, K. (2017). Listening environment and the bottom line: How a positive listening environment can improve financial outcomes. *International Journal of Listening 31*, 71–79. doi:10.1080/10904018.2014.965391

Johnston, M. K., Reed, K., & Lawrence, K. (2011). Team Listening Environment (TLE) scale: Development and validation. *Journal of Business Communication*, *38*, 3–26. doi:10.1177/0021943610385655

Johnston, M. K., Reed, K., Lawrence, K., & Onken, M. (2007). The link between communication and financial performance in simulated organizational teams. *Journal of Managerial Issues*, *19*, 536–553.

Pearce, C. G., Johnson, I. W., & Barker, R. T. (2003). Assessment of the Listening Styles Inventory. *Journal of Business and Technical Communication*, *17*, 84–113. doi:10.1177/1050651902238546

Reed, K., Goolsby, J. R., & Johnston, M. K. (2014). Extracting meaning and relevance from work: The potential connection between the listening environment and employee's organizational identification and commitment. *International Journal of Business Communication*, *53*, 326–342. doi:10.1177/2329488414525465

Reed, K., Goolsby, J. R., & Johnston, M. K. (2016). Listening in and out: Listening to customers and employees to strengthen an integrated market-oriented system. *Journal of Business Research*, *69*, 3591–3599.
Riddle, B. L., Anderson, C. M., & Martin, M. M. (2000). Small Group Socialization Scale. *Small Group Research*, *31*, 554–572. doi:10.1177/104649640003100503
Walters, J. (2005). Fostering a culture of deep inquiry and listening. *Journal for Quality and Participation*, *28*, 4–7.

Scale

Team Listening Environment (Johnston & Reed, 2015)

Please circle your level of agreement with the statements listed below using the following scale:

5 = Strongly Agree 4 = Agree 3 = Undecided 2 = Disagree 1 = Strongly Disagree

1) The other group members paid attention to me.
2) The other group members genuinely wanted to hear my point of view.
3) The other group members listened to what I had to say.
4) The other group members understood me.
5) The other group members seemed to be attentive to what others had to say.

Note: Items are worded with reference to a particular work team interaction. For a more general measure, items can be slightly rephrased to be current tense or more fitting to the situational vernacular (e.g., “Other group members pay attention to me” or “Other team/department/plant members pay attention to me”). To calculate the TLE variable score per participant, average the participant’s responses to all five items.

Profile 60

Time Studies

Profiled by: Graham D. Bodie, PhD

Meek School of Journalism and New Media, The University of Mississippi

Construct

The construct of interest in time studies is the amount of time people spend in communication-related activities. Although the specific activities of interest have varied across studies, most lists include one or more types of listening in order to compare how much time people spend listening versus speaking, writing, reading, and consuming media.

Instrument Type

Self-Report; Behavioral Observation

Description

Time-use studies attempt to estimate the amount of time people spend doing particular activities, usually breaking down estimates for various 24-hour periods and reporting percentages of time spent on specific tasks. These studies have used a variety of approaches. Paul Rankin (1926) conducted the first time study focusing on communication activity. Using a convenience sample of 21 people (12 teachers, 2 housewives, 2 stenographers, 1 student, 3 researcher workers, and 1 nurse; 18 women and 2 men), he asked participants to log their communication activities for one or more days in

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

15-minute increments between 6:00 a.m. and midnight.[1] Later studies have attempted to replicate and extend these results to specific populations (e.g., college students, scientists, and engineers; see below). In contrast, some studies have used retrospective self-reporting methods in the place of time sampling procedures. For instance, Barker, Edwards, Gaines, Gladney, and Holley (1980) asked participants to report on their communication activity over the previous 24 hours. More recent studies have collected data related to the use of various technologies (e.g., television and the Internet; see Janusik & Wolvin, 2009). A final method for estimating time spent in communication activities is to observe behavior directly (Imhof, 2008).

Administration

Time study methods include time sampling procedures, survey methods, and observation. Time sampling involves the administration of some type of log sheet that contains spaces for indicating the communication activity engaged and how long it was engaged. Rankin asked participants to report engaged activities every 15 minutes. Similar logs were used by Bohlken (1999) and Samovar, Brooks, and Porter (1969); Perras and Weitzel (1981) used half-hour increments. Bird (1953) asked students to "keep a running record of minutes spent in listening, reading, speaking, and writing" (p. 127), and, thus, it is unclear whether he asked participants to record in time intervals or when they were asked to record (e.g., at the end of the day or throughout the day; also see Weinrauch & Swanda, 1975). Hinrichs (1964) provided respondents with five data collection forms with a randomly generated time of day printed on each. Respondents set alarms on a self-administered wristwatch to signal them when to fill out the form throughout the day. In a bit more extensive log-based study, Burns (1954) asked four members of an organization to track their time each work day for 5 weeks:

> [Each] recorder … noted whether [he] or someone else initiated an interaction and classified it according to whether it involved obtaining, giving, systematizing, or recording information … the time of [the] episode, and whether it took the form of a conversation, a telephone call, a letter or memorandum (written or read), or whether it took place without verbal communication (e.g., drawing, watching operations, etc.). (pp. 76–77)

For survey-based studies, participants have been asked to estimate the percentage of their time spent in various communication activities (Klemmer & Snyder, 1972) or to estimate their communication activity over the past 24-hour period (Barker *et al.*, 1980; Janusik & Wolvin, 2009).

For retrospective self-report studies, researchers can expect students to spend as little as 10 minutes at one sitting. For time samples, perhaps as much as 10 minutes per day

1 Although it is common in literature reviews and textbooks to claim that respondents reported on their daily activity for multiple days, 6 participants reported for only 1 day, another 6 people reported for 2 days, 8 reported for 3 days, and 1 reported for 18 days. Thus, when literature reviews claim that "21 people reported on their communication activity for 60 days" the "60 days" refers to the total number of days reported, not to the number of days that each participant was asked to report. Most time studies that measure communication activity do not track people over a long period of time (for an exception, see Burns, 1954).

for up to a month is needed. Strategies for the successful employment of time sampling studies are elaborated by Reis, Gable, and Maniachi (2014), and several computer programs currently exist to assist researchers with data collection (e.g., Snap Survey, www.survey-snap.com).

A few time studies have included behavioral observation either as the primary data collection method (Imhof, 2008) or to provide validity evidence for self-reporting of communication activities (Hinrichs, 1964; Klemmer & Snyder, 1972; Wilt, 1966). Doing so requires a larger time commitment on the part of the researcher as well as access to willing organizations and participants.

Scoring

Scoring for a time study varies as a function of data collection (as discussed in this profile). Most often, interest lies with what Rankin (1966, p. 52) called "four chief modes of communication"—talking (which some label *speaking*), listening, reading, and writing. Time spent in each of these activities is reported as a percentage of total communication time; that is, activities add up to 100% and represent the proportion of communication time spent in various communication modes as opposed to a proportion of one's day spent communicating in these modes. For instance, Rankin included a *no communication* category and a *miscellaneous* category that, when included in estimates of time spent throughout an average day, change the commonly reported estimates (see Table P60.1).

Moreover, although Rankin's (1966) results are often reported as the percentage of time spent in the four communication modes, a distinction he originally reported, his data were more nuanced. The four chief modes of communication are actually aggregate measures that were calculated by summing values reported for subcategories, as detailed in Table P60.2.

Particularly important for estimates of time spent listening versus talking is the fact that Rankin (1966) reported that "it was found easier to record as conversation and divide into talking and listening afterwards" (p. 51) and that "The time listed as conversation was divided in two, and half recorded under talking and half under listening, on the assumption that, in general, one listens about half the time and talks about half the time during a conversation" (p. 52).

Table P60.1 Comparison of time estimates in Rankin's study with and without the *no communication* (NC) category.

Mode	With "NC" category	Without "NC" category
Talking	21.7%	31.9%
Writing	6.9%	11.0%
Listening	29.7%	42.1%
Reading	10.1%	15.0%
Miscellaneous	2.7%	n/a
None	29.6%	n/a

Table P60.2 Components contributing to aggregate measures of talking, writing, listening, and speaking as reported by Rankin.

Communication mode	Components contributing to time spent
Talking	Half of the time reported for conversation (high and low order)
Writing	Forms of writing not stated
Listening	Half of the time reported for conversation (high and low order) PLUS listening to conference, directions, formal talks, memorizations, oral reading, and vocal music
Reading	Forms of reading not stated

Barker *et al.* (1980) and Emanuel *et al.* (2008) provided aggregate data on these four categories as well. The Barker team computed time spent listening by adding categories for classroom, formal, and interpersonal; time spent listening to mass media was placed in a separate category. When the Barker *et al.* study is reported on, however, the tendency is to combine mass-mediated and classroom/formal/interpersonal listening (see Table 1 in Janusik & Wolvin, 2009). The Emanuel team computed time spent listening by adding categories for classroom listening, phone listening, face-to-face listening outside of class, and "listening" to instant messages.

To derive estimates of the four "chief modes of communication" from their data, Janusik and Wolvin (2009) computed "daily % … from total time reported in each activity summed, and then divided by the total time reported for each context" (p. 113).

Weinrauch and Swanda (1975) reported an additional statistic, "breaking down the responses in terms of a typical hour of communication" (p. 28). The average hour of communication was reported in minutes that totaled 60 by summing minutes spent reading, writing, speaking, and listening; they additionally reported these data for the total work week, and separately for morning and afternoon and for each of the five workweek days (i.e., Monday–Friday).

Imhof (2008) derived estimates of time spent listening for her observational rubric based on "listening opportunities contained in different instructional methods," such as "teacher-directed class interaction, instructions for guided student practice, presentations by the teacher, guided practice in pairs, independent group work, student presentations … games, listening to media, [and] independent seat-work" (p. 4). In addition to providing data for these smaller categories, Imhof also provided estimates of the duration of total listening requirement (in minutes) for a 45-minute class period as well as total minutes that required listening compared to minutes spent on activities that did not.

Development

As noted here, Paul Rankin (1926) conducted the first known time study focused on communication activities as part of his doctoral dissertation, written at the University of Michigan under the guidance of faculty in the Department of Education. The activities included in Rankin's survey were (a) conversation, (b) writing, (c) reading,

(d) listening, (e) miscellaneous, and (f) no communication.[2] Conversation was subdivided into high and low order. High-order conversation was that which "dealt with important or complex matters, such as the significance of the Locarno treaties, or appropriate ways of meeting a race problem, or the wisdom of building on a certain lot" (Rankin, 1926, p. 51). Low-order conversation included more trivial matters, "such as the breakfast-table talk about how well different ones slept, or about the furniture of the people who moved in next door" (Rankin, 1926, p. 52). Listening was subdivided into listening to a conference, oral reading, formal talks, directions, vocal music, and memorizations. The estimate for talking was computed by dividing time reported in the conversation categories (high and low order) in half; the other half was used for the calculation of listening, to which was added estimates for listening to a conference, oral reading, formal talks, directions, vocal music, and memorizations.

Most basic course texts cite Rankin along with one or more additional studies in their justification for why studying human communication is so vital to student success (Janusik & Wolvin, 2002). Although these claims often cite time studies as a single type of study, there are important differences in method that have evolved since Rankin's report. Studies prior to 1980 primarily utilized one or more forms of time sampling procedures, asking respondents to report at various times of the day what communication activities they were engaged in. Rankin asked respondents to report every 15 minutes, and these logs of time spent were recorded for between 1 and 18 days (see note 1). Perras and Weitzel (1981) used a similar method with reports every 30 minutes of waking time. Bird (1953) reported having students keep "a running record of minutes spent" in the four modes, and Weinrauch and Swanda (1975) asked respondents "to keep a careful record of their time spent in communication" (p. 27). Hinrichs (1964) used a primitive form of signal-contingent recording, asking participants to set an alarm at five random times during the day and report on their communication up to that point in time.

There is also a large literature on how scientists and engineers use their time, starting with the Case Institute of Technology study conducted in 1958 but also including research by Graham, Wagner, Gloege, and Zavala (1967) commissioned by the American Institutes for Research. The Graham *et al.* study was concerned with "informal scientific and technical communication … those that involve person-to-person interactions" (p. 3).

In her 1949 doctoral dissertation, Miriam Wilt asked elementary school teachers to estimate the amount of time they believed children spend learning through reading, speaking, listening, and writing. Similar retrospective reports have been used in other published studies. The Barker *et al.* (1980) study was particularly influential to subsequent time studies because it introduced two primary instrument modifications that have been utilized in most published work reporting on listening behavior since. First, participants were asked "to think back over the last 24 hours and answer the questions based on [this] reflection" (p. 103). Second, the list of activities was modified and organized into six sections. The first five sections constituted the activities students were asked to estimate: speaking and listening in the classroom (formal lecture, group discussion), writing (in class, extracurricular, and personal pleasure), reading (class, nonclass), interpersonal

2 Although most reports of Rankin's study suggest *talking* was a category, it was not recorded by participants but derived from data reported on high- and low-order conversation, taking half of that reported time.

speaking and listening (face-to-face, telephone), and mass communication activities (radio, television, records/taped/live music). The sixth section asked students "to rate how typical their last 24 hours, with respect to communication activities, had been" (p. 103). An additional modification to the instrument was made by Janusik and Wolvin (2009), who asked undergraduate students to estimate the amount of time they spent writing, reading, speaking (face-to-face), listening (face-to-face), and on the telephone, email, and Internet. Each of these activities was estimated in four contexts (school/school work, with friends, at work, and with family) using a scale that ranged from *never* to *9+ hours*.

Only a few studies have included observational data in estimates of time spent listening, with only one study that has used behavioral data as the sole means of estimation (Imhof, 2008). More commonly, observations are included to check the veracity of retrospective self-report (Hinrichs, 1964; Klemmer & Snyder, 1972; Wilt, 1966).

Reliability

Estimating the homogeneity and stability of time study data differs as a function of how those data are collected. For studies that ask people to report over the immediate past 24 hours, researchers have reported test–retest reliability. Barker *et al.* reported $r = .81$ in a pilot study of 106 participants and $r = .76$ in the primary study of 645 participants. Emanuel *et al.* (2008) reported test–retest coefficients of $r = .86$ after 24 hours and $r = .77$ after 48 hours; these values come from two separate samples, each comprising 5% of the original 842 participants. What these numbers tell us, however, is unclear as there is no real reason to suspect people are highly consistent in how they spend their time from day to day. No measure of reliability was reported in the Janusik and Wolvin (2009) study. Likewise, no reliability estimates have been reported in time sampling studies, although researchers who decide to employ this method are advised to consult Reis *et al.* (2014) for the appropriate statistics to report. Imhof (2008) reported agreement of 93% between observers who recorded listening behaviors of students in a German school system.

Validity

The primary criterion for validity in time studies is the degree to which studies produce an adequate inference for the population; that is, to what degree can we be confident that the estimated time spent in various communication activities is true for some larger population? No time studies can be said to adhere to standards for making this type of inference. Rankin's (1926) study estimated time spent listening for 21 individuals who were chosen because they were convenient to sample. In attempted replications of Rankin, Samovar *et al.* (1969) mailed time logs to "randomly selected adults in the metropolitan San Diego area" (p. 302), and Weinrauch and Swanda (1975) sampled from "South Bend, Indiana area business personnel" (p. 27). Studies interested in populations of scientists or engineers tend to sample from a single organization (Burns, 1954; Klemmer & Snyder, 1972) or utilize snowball and other convenience sampling techniques (Graham *et al.*, 1967). Studies making claims about college students have most typically sampled from a single college campus (Barker *et al.*, 1980; Bird, 1953; Janusik & Wolvin, 2009; Perras & Weitzel, 1981), although Emanuel *et al.* (2008) sampled from four “different colleges and universities in the Southeast” (p. 19); all students were enrolled in a basic communication course.

Others have compared student self-reports with observational data, with findings suggesting differential patterns of under- and overreporting of communication activities. Wilt’s (1966) study asked teachers to estimate the amount of time students spend listening and compared these estimates to observations made in schools within “several types of communities” (p. 65). Both methods suggested that students are expected to listen for most of the school day, although teachers did tend to underestimate how much time students are expected to listen. Similarly, data from a communications research and development laboratory reported by Klemmer and Snyder (1972) suggested that people underestimate time spent in face-to-face conversations but overestimate reading and writing time. Although differences emerged, the general claim is that people, regardless of occupation, spend a large part of their day in communication-focused activities and much of it in face-to-face interaction with others. More recent work that includes technology has yet to show that face-to-face interaction is drastically reduced, at least for college students (Janusik & Wolvin, 2009).

Availability

Multiple versions of time study instruments are available in the literature. We provide instruments used by Janusik and Wolvin (2009), Klemmer and Snyder (1972), Hinrichs (1964), and Emanuel *et al.* (2008) at the end of this profile as examples of how one might construct a time series questionnaire. All reproductions of text are in line with STM Permissions Guidelines; all other extracts are used with permission. All forms can be used in research at no cost, given appropriate citation.

Sample Studies

The primary purpose of Rankin’s study was to estimate the amount of time people spend in various communication activities. Since his study, others have used his list and time sampling methods with samples of elementary school (Imhof, 2008; Wilt, 1966) and college students (Barker *et al.*, 1980; Bird, 1953; Emanuel *et al.*, 2008; Janusik & Wolvin, 2009) and samples of working adults (Hinrichs, 1964; Klemmer & Snyder, 1972; Weinrauch & Swanda, 1975). There is also a literature that focuses exclusively on time spent in information communication activities of scientists (Case Institute of Technology, 1958; Graham *et al.*, 1967).

Although different sampling techniques and methods have been employed, estimates have been quite similar across these studies. Overall, research suggests that people spend a good deal of their waking hours in some form of informal interaction with others, usually involving face-to-face interaction. In general, students are expected to listen for about two thirds of classroom time, and executive personnel as well as scientists spend about the same amount of their day in some form of interpersonal interaction (with much of that time likely spent listening).

Critique

Time studies provide a rationale for textbook authors, practitioners, and educators at all levels regarding the need for training or teaching of communication skills. In their analysis of 17 basic course texts, Janusik and Wolvin (2002) found that all books used a time study, usually Rankin or Barker *et al.*, to establish the importance of listening in daily life. Thus, time studies provide administrators and practitioners with evidence that listening is important and that training is vital to student and professional success.

In general, these studies focus attention on the percentages of time spent in various communication activities. Their design, however, makes reported estimates questionable. For the most part, no study has sought to sample from a diverse or nationally representative sample, instead focusing on more localized populations and convenience samples. Bird's data came from a single semester, using students enrolled in his listening class at the time. Although the Barker team reported selecting students "according to a proportional sampling plan from the entire student population" (p. 102), no population-level demographic information was provided to demonstrate that this goal was achieved. Moreover, students were "approached in classroom buildings, student lounges, cafeterias, dormitories, apartments, and offices" (p. 103) rather than being selected from a randomly generated list of the entire student body population. Janusik and Wolvin sampled from "students enrolled in a basic communication course" (presumably in one academic semester) who were approached in the classroom. Moreover, classes were assigned to report on a "specific day of the week" and not individual students, making the class the unit of sampling for random assignment rather than the individual.

Even when the title of the article limits the scope of estimation, the claim is still more universal than what the data can actually produce. For example, even though Emanuel *et al.*'s (2008) title is "How College Students Spend Their Time Communicating," their data come from only four institutions, and all their participants were enrolled in communication courses. Thus, a more appropriate title would be, "How College Students Enrolled in Communication Studies Courses at Universities in the Southeastern United States Spend Their Time Communicating." When a new time study is conducted, the rationale generally involves some sort of argument regarding shifts in the development and use of technology or the applicability beyond the population sampled, so researchers generally understand this criticism. Still, these studies do not remedy the problem by taking the easy road and administering log sheets to convenience samples of students or people from a single organization.

In addition to sampling people, concern is also raised with respect to the sampling of communication activities. Rankin's categories of talking, listening, reading, and writing are used most often, but Rankin's categories may have helped skew listening as more important. There were more categories to report for listening and not a single "talking" category. Instead, talking was derived from conversation data divided into two, one part for listening and one part for speaking.

Perhaps more troubling is the idea that speaking and listening can be separated, even for analytical purposes. Barker *et al.* (1980) and Janusik and Wolvin (2009) had people report on speaking and listening separately, but we are left to wonder about the accuracy of that recall, especially when reporting over the past 24-hour period as opposed to after every conversation in a time sampling or diary-based study. Can people provide estimates of the time they spent listening and speaking in a conversation when conversations are fluid, dynamic, and not often marked by clear roles?

Dichotomizing speaking and listening is problematic on theoretical and practical grounds. Theoretically, these acts are not separate but part of a larger process of interaction (Berger, 2011). Practically, the researcher is asking participants to potentially recall dozens if not hundreds of conversations among friends and strangers (e.g., a cashier at the local grocery) and then estimate the amount of time spent talking versus listening to the other. Most conversations involve overlapping speech and no clear distinction between speaking and listening roles, thus calling into question exactly what estimates of speaking and listening in an interactional context really tell us. Estimating these roles separately might be more accurate when observing conversation, but researchers would have to carefully operationalize speaking and listening as, for instance, time spent in the front and back channel of speech. Even here, who is to say that the listener is, in fact, listening—that is, processing the contribution of the other person—as opposed to daydreaming or otherwise thinking about what he or she is going to say next? It is thus difficult to rely on estimates of listening time in interactional contexts (face-to-face or mediated). Estimating time spent listening to radio, television, or other forms of media is no more simple a task, as attention waxes and wanes even when the intent is to focus on a program or song. Does a researcher try to dissect media listening time into time spent heavily concentrating versus daydreaming while it is in the background? Or, is a basic estimate good enough for most purposes?

With log sheets, we do not really know when people fill them out—so they can still be retrospective data, although closer to the actual activity than if you asked someone on Wednesday what they did last Thursday. With the advent of computer technology and smartphone proliferation, researchers can use services that timestamp logs (much like the alarm clock strategy employed by Hinrichs, 1964).

In general, then, we are left with a striking picture of the importance of communication (and, by extension, listening) for a variety of people and situations. The more nuanced critiques of these estimates should not take away from the consistency in findings that people spend a great deal of their days interacting with others. Communication practitioners and educators should thus be comforted by a secure role in helping people become better communicators in most life domains.

References

Barker, L., Gladney, K., Edwards, R., Holley, F., & Gaines, C. (1980). An investigation of proportional time spent in various communication activities by college students. *Journal of Applied Communication Research*, *8*, 101–109. doi:10.1080/00909888009360275

Berger, C. R. (2011). Listening is for acting. *International Journal of Listening*, *25*, 104–110. doi:10.1080/10904018.2011.536477

Bird, D. E. (1953). Teaching listening comprehension. *Journal of Communication*, *3*, 127–130.

Bohlken, B. (1999). Substantiating the fact that listening is proportionately most used language skill. *The Listening Post*, *70*, 5.

Burns, T. (1954). The directions of activity and communication in a departmental executive group: A quantitative study in a British engineering factory with a self-recording technique. *Human Relations*, *7*, 73–97. doi:10.1177/001872675400700104

Case Institute of Technology. (1958). *An operations research study of the scientific activity of chemists*. Cleveland, OH: Operations Research Group, Case Institute of Technology.

Emanuel, R., Adams, J., Baker, K., Daufin, E. K., Ellington, C., Fitts, E., … Okeowo, D. (2008). How college students spend their time communicating. *International Journal of Listening*, *22*, 13–28. doi:10.1080/10904010701802139

Graham, W. R., Wagner, C. B., Gloege, W. P., & Zavala, A. (1967). *Exploration of oral/informal technical communications behavior*. Washington, DC: American Institutes for Research.

Hinrichs, J. R. (1964). Communications activity of industrial research personnel. *Personnel Psychology*, *17*, 193–204.

Imhof, M. (2008). What have you listened to in school today? *International Journal of Listening*, *22*, 1–22. doi:10.1080/10904010701802121

Janusik, L. A., & Wolvin, A. D. (2002). Listening treatment in the basic communication course text. *Basic Communication Course Annual*, *14*, 164–210.

Janusik, L. A., & Wolvin, A. D. (2009). 24 hours in a day: A listening update to the time studies. *International Journal of Listening*, *23*, 104–120. doi:10.1080/10904010903014442

Klemmer, E. T., & Snyder, F. W. (1972). Measurement of time spent communicating. *Journal of Communication*, *22*, 142–158. doi:10.1111/j.1460-2466.1972.tb00141.x

Perras, M. T., & Weitzel, A. R. (1981). Measuring daily communication activities. *The Florida Speech Communication Journal*, *9*, 19–23.

Rankin, P. T. (1926). The measurement of the ability to understand spoken language. *Dissertation Abstracts*, *12*(1952): 847.

Rankin, P. T. (1966). Frequency of use of listening [Material excerpted from a 1926 University of Michigan doctoral thesis entitled *Measurement of the ability to understand spoken language*]. In S. Duker (Ed.), *Listening readings* (pp. 5–-62). New York: Scarecrow Press.

Reis, H. T., Gable, S. L., & Maniachi, M. R. (2014). Methods for studying everyday experiences in its natural context. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 373–403). Cambridge: Cambridge University Press.

Samovar, L. A., Brooks, R. D., & Porter, R. E. (1969). A survey of adult communication activities. *Journal of Communication*, *19*, 301–307. doi:10.1111/j.1460-2466.1969.tb00852.x

Weinrauch, J. D., & Swanda, J. R., Jr. (1975). Examining the significance of listening: An exploratory study of contemporary management. *The Journal of Business Communication*, *13*, 25–32. doi:10.1177/002194367501300103

Wilt, M. E. (1966). Demands on the listening skills of elementary school children [Excerpts from a 1949 Pennsylvania State College doctoral dissertation entitled *A study of teacher awareness of listening as a factor in elementary education*, pp. 115–136]. In S. Duker (Ed.), *Listening: Readings* (pp. 63–80). New York: Scarecrow Press.

Time Study Instruments

Example 1 (Klemmer & Snyder, 1972)

What percentage of time of the working week do you spend:

1) On the telephone? Include all telephone calls you make or receive while you are at work.
2) Talking face-to-face? Include all conversations and conferences during the working day regardless of topic.

3) Reading or looking at any printed, written, or drawn material?
4) Writing with pencil or pen or doing any drawing, programming or tallying, etc. with pencil or pen?
5) Using a typewriter, keypunch, adding machine, desk calculator, copying machine, etc., or any computer or computer terminal?
6) Working with laboratory apparatus or any equipment not included above?
7) On other activities such as attending lectures, thinking (only), walking, traveling, _______. (Please underline or add words describing the principal activities included.)

Example 2 (Janusik & Wolvin, 2009)

Yesterday, how much time did you spend:

	0 minutes	15 minutes	30 minutes	45 minutes	Other
1. Speaking in class?					
2. Listening in class?					
3. In conversation on the phone?					
4. In face-to-face conversation outside of class?					
5. Instant-messaging on the Internet?					
6. Reading school-related material *not* on the Internet?					
7. Reading school-related material on the Internet?					
8. On personal reading *not* on the Internet?					
9. On personal reading on the Internet?					
10. Writing/typing school-related material?					
11. On personal writing/typing *not* on the Internet?					
12. On personal writing/typing on the Internet including e-mail?					
13. Listening to music?					
14. Watching/listening to TV?					

Time Study Example 3 (Hinrichs, 1964)

Source: Hinrichs (1964). Reproduced with permission of Taylor & Francis.

NAME
DATE
TIME

1. WHERE ARE YOU?
1☐ Your Desk/Office
2☐ Your Laboratory
3☐ Elsewhere in Your Division
4☐ Other Division
5☐ Other Affiliate
6☐ Other

WHAT TIME IS IT NOW? ☐

WHAT ARE YOU DOING?

2. NOT COMMUNICATING – 1☐ Analysis or Calculations Not in Draft or Final Form to be Transmitted to Others
2☐ In Transit 3☐ Filing 4☐ Looking for Info. 5☐ Waiting 6☐ Lab Work 7☐ Other
IF YOU CHECK ANY OF THE ABOVE DO NOT GO ON – WIND YOUR WATCH AND SET IT FOR THE NEXT SCHEDULED TIME

3. COMMUNICATING – IF YOU CHECK ANY OF THE FOLLOWING CONTINUE IN THAT COLUMN TO REVERSE SIDE

1 ☐ Listening, Speaking	2☐ Writing 3☐ Dictating 4☐ Editing Written work of Others 5☐ Editing Written Work of Your Own 6☐ Plotting or Listing Data in Draft or Final Form to Be Transmitted to Others	7☐ Reading Something Written By Others
4. 1☐ On Telephone 2☐ Meeting – Sch'd. Periodic Intervals 3☐ Special Meeting – Not Periodic but Planned at Least One Day Ahead 4☐ Contact, Gathering or Bull Session Not Planned One Day Ahead	1☐ Letter 2☐ Book, Article, or Manual 3☐ Note or Informal Memo 4☐ Formulas, Data 5☐ Formal Memo 6☐ Progress Report	1☐ Letter 2☐ Book, Article, or Manual 3☐ Note or Informal Memo 4☐ Formulas, Data 5☐ Formal Memo 6☐ Progress Report
5. How Many People Are Included In This Communication, Including Yourself? 1☐ 2 1☐ 7–10 2☐ 3–6 4☐ Over 10	Is This Material Prepared at Periodic Intervals; For Example, Weekly, Monthly, Quarterly? 1☐ Yes 2☐ No	Is This Material Prepared at Periodic Intervals; For Example, Weekly, Monthly, Quarterly? 1☐ Yes 2☐ No
6. Who Started This Communication? 1☐ Yourself 2☐ Someone Above Your Level 3☐ Someone at Your Level 4☐ Someone Below Your Level 5☐ Don't Know or Not Applicable	Is This a Redraft? 1☐ First Draft 2☐ Second Draft 3☐ Third Draft 4☐ Fourth Draft and More 5☐ Don't Know	
7. Are the Other People Primarily From: 1☐ Outside the Company 2☐ Other Affiliate 3☐ Research Company 4☐ Don't Know	Is This Written Matter Directed Primarily to People From: 1☐ Outside the Company 2☐ Other Affiliate 3☐ Research Company 4☐ Don't Know	Are You Reading Material Prepared by People From: 1☐ Outside the Company 2☐ Other Affiliate 3☐ Research Company 4☐ Don't Know
8. What Is the Level of Most of the Participants (Other Then Yourself)?	To What Level Is the Written Matter Primarily Directed?	What Is the Level of the Writer or Writers?

1☐ Above Director Level
2☐ Director Level
3☐ Assoc. or Asst. Director or Sr. Res. Assoc.
4☐ Section Head or Assoc.
5☐ Group Head or Senior Eng./Chemist
6☐ Professional – Non-Supervisory
7☐ Non-Professional
8☐ Mixed, Not primarily Any one of the Above
9☐ Don't Know or Not Applicable

COMPLETE ALL OF THE REMAINING QUESTIONS

What Is the Main Subject of This Communication? (Answer either 9 or 10)

9. TECHNICAL
1☐ Technical Service to Affiliates or Other Divisions
2☐ Technical Service to Non-Affiliates
3☐ Exploratory Research or Engineering
4☐ Development or Normal Research or Engineering
5☐ Other Technical

10. NON-TECHNICAL
1☐ Employee Matters: Interviewing, Employment, Training and Employee Matters Other Than Supervision of Technical Work.
2☐ Budget and Finance: Budget Review, Cost Data on Non-Technical work
3☐ Materials: Supplies, Purchases, Sales, Inventory, Inspection, Ordering
4☐ Other Non-Technical Company Business
5☐ Personal: All Personal Time

11. Does the Subject of This Communication Require Action? 1☐ By You 2☐ By Another Participant 3☐ By Someone Not Participating 4☐ By Several Persons 5☐ No Action to Be Taken 6☐ Don't Know	12. Does This Communication Deal With An Idea Which: 1☐ You Alone Originated 2☐ You Originated With Others 3☐ Someone ELSE Originated 4☐ Don't Know	13. How Long Has This Communication Been Taking Place? 1☐ Less Than 5 Minutes 2☐ 6–15 Minutes 3☐ 16–30 Minutes 4☐ 31–60 Minutes 5☐ Over 60 Minutes
14. What Is You Major Function In This Communication? 1☐ Giving Information 2☐ Receiving Information 3☐ Exchanging Information	15. Do You Think This Communication is 1☐ Necessary to Your Own or Someone Else's Work 2☐ Of Only Minor Importance 3☐ Not Necessary to the Work of Anyone in the Company 4☐ Don't Know	PLEASE WIND YOUR WATCH AND SET IT FOR THE NEXT SCHEDULED TIME.

Time Study Example 4 (Emanuel *et al.*, 2008)

Source: Emanuel *et al.* (2008). Reproduced with permission of Taylor & Francis.

SURVEY INSTRUCTIONS

This survey is designed to measure the amount of time you normally spend in different kinds of communication activities. Think back to **yesterday** as you answer the following questions. Check only ONE box per question.

How to record "doing two things at once" → EXAMPLE: If you listened to a lecture AND took notes at the same time, count the time as BOTH listening time AND writing time.

How to record your answers →

If you have spent no time in an activity, you would record it like this:

_____ hours ☑ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

If you have spent 15 minutes in an activity, you would record it like this:

_____ hours ☐ 0 minutes ☑ 15 minutes ☐ 30 minutes ☐ 45 minutes

If you have spent 1 hour in an activity, you would record it like this:

___**1**___ hours ☑ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

If you have spent 2 hours and 30 minutes in an activity, you would record it like this:

__**2**___ hours ☐ 0 minutes ☐ 15 minutes ☑ 30 minutes ☐ 45 minutes

And so on. Use only the time choices provided. Round **UP** if you need to.
The survey begins on the next page. There are only 20 questions and it takes less than 5 minutes to complete. Remember, check only ONE box per question.

Yesterday, how much time did you spend:

1) **speaking in class**?

_____ hours ☐ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

2) **listening in class**?

_____ hours ☐ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

3) in **conversation using a phone**?

_____ hours ☐ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

4) in **face-to-face conversation outside of class**?

_____ hours ☐ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

5) **instant-messaging**?

_____ hours ☐ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

6) **reading *school-related* material NOT using a computer/smartphone**?

_____ hours ☐ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

7) **reading *school-related* material using a computer/smartphone**?

_____ hours ☐ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

8) on ***personal*** **reading NOT using a computer/smartphone**?

_____ hours ☐ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

9) on ***personal*** **reading using a computer/smartphone**?

_____ hours ☐ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

10) **writing/typing *school-related* material NOT using a computer/smartphone**?

_____ hours ☐ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

<u>Yesterday</u>, how much time did you spend:

11) **writing/typing *school-related* material using a computer/smartphone**?

_____ hours ☐ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

12) on ***personal*** **writing/typing NOT using a computer/smartphone**?

_____ hours ☐ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

13) on ***personal*** **writing/typing using a computer/smartphone**?

_____ hours ☐ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

14) **listening to music**?

_____ hours ☐ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

15) **watching TV**?

_____ hours ☐ 0 minutes ☐ 15 minutes ☐ 30 minutes ☐ 45 minutes

16) How typical a day was yesterday? That is, how much was yesterday like any other day?

[1] ☐ very typical [2] ☐ somewhat typical [3] ☐ not very typical [4] ☐ not typical at all

17) What is your gender? ☐ Female ☐ Male

18) What is your year in school?

☐ Freshman ☐ Sophomore ☐ Junior ☐ Senior ☐ Graduate

19) How many credit hours are you enrolled in this semester?

☐ less than 12 ☐ 12–18 ☐ more than 18

If you answered "**0 minutes**" to **question #5, do not answer question #20**.

20) How would you describe instant-messaging? ☐ As talking **OR** ☐ As typing

Thank you for completing the survey.

Profile 61

The Listening Test of the Internet-Based Test of English as a Foreign Language (TOEFL iBT)

Profiled by: Vahid Aryadoust, PhD

National Institute of Education, Nanyang Technological University, Singapore

Construct

The TOEFL iBT was designed by the Educational Testing Service (ETS) to assess the ability of English as a second language (ESL) learners to use academic English at a tertiary education level. The test consists of four sections: reading, listening, speaking, and writing. This report primarily focuses on the listening section of the TOEFL iBT; where relevant, the integrated speaking and writing sections are also discussed.

Instrument Type

Cognitive Assessment

Description

The TOEFL iBT consists of four sections assessing the reading, listening, speaking, and writing skills of ESL learners. Listening skills are an integral part in all but the reading section. The listening section was developed based on an approach that recognizes both task-based and construct-based test development (Chapelle, Enright, & Jamieson, 2008). The listening test is composed of three primary tasks—listening to lectures, listening to classroom or on-campus conversations, and listening to in-class discussions—that tap listeners' understanding of major ideas, specific support, speakers' attitudes and intentions, and their ability to draw inferences (Educational Testing Service [ETS], 2015).

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

The TOEFL iBT listening test is a post-listening performance (PLP) test, meaning that test takers must first listen to the text passages and take notes. The test items are subsequently presented, and test takers are allowed to use their notes to answer the questions. The PLP format of the test likely precludes taxing test takers' working memory, which could happen if listening, item reading, and answering are performed simultaneously—although, as will be discussed here, there is still some cause for concern (Aryadoust, 2012).

Listening skills are also instrumental in the integrated speaking and writing tasks of the TOEFL iBT. As described by Jamieson, Jones, Kirsch, Mosenthal, and Taylor (2000), listening in these tasks is "contextualized," "linked thematically," and consists of "performance-based" tasks (p. 5). In the integrated speaking test, test takers listen to and/or read several texts to help them prepare their oral responses. Similarly, in the integrated writing test, test takers listen to the text, read a related written text, and then provide a written response. A factor analytic study reported by Sawaki, Stricker, and Oranje (2008) showed that listening ability plays a statistically significant, but minimal, part in the integrated speaking tasks.

Listening comprehension of ESL learners is tested directly in the listening subtest of the TOEFL iBT and indirectly in the speaking and writing subtests. The listening subtest of the TOEFL iBT was designed based on the listening framework developed by Bejar, Douglas, Jamieson, Nissan, and Turner (2000); this framework was further elaborated in later phases of test development. The framework predicts that task characteristics (e.g., situation, text materials, and test rubrics) and task variables (e.g., participants and setting) influence test takers' performance. The framework further proposes multiple features that the developers of the TOEFL iBT should consider in test design. Based on these requirements and guidelines, the test developers developed a 60- to 90-minute listening test comprising 34 to 51 items. As noted here, the items engage listeners' ability to understand surface information, draw inferences, and identify speakers' attitudes and perceptions.

The integrated speaking and writing tasks engage listeners in multiple language modalities. In the speaking test, test takers listen to multiple short and lengthy listening texts and conversations, read relevant texts (in some tasks), and formulate oral responses to tasks. Headsets available at the testing centers are used to play the listening texts and to record test takers' oral responses. Similarly, in the integrated writing section, test takers listen to a short lecture, read a relevant short text, and type a response to a task, using their understanding of the written and oral stimuli. As such, computer competency is a fairly important requirement of taking the TOEFL iBT.

Administration

All subtests of the TOEFL iBT are administered via the Internet at certified test administration sites throughout the world. The test is administered between 15 and 30 times annually, and the centers can choose what time to administer the test. The listening test must be taken with the other subtests on the same day at the same testing center. Mock tests and preparation materials are available from the ETS website, which can be used by test candidates to prepare for the exam (see https://www.ets.org/toefl/ibt/scores/ for further information).

Scoring

Multiple-choice questions are administered in the listening section with one or more correct options that allow partial credit scoring. The scores for each subtest of the TOEFL iBT are scaled and range between 0 and 30. For the listening section, test scores from 0 to 14 are low, 15 to 21 are medium, and 22 to 30 are high (ETS, 2015). The listening test is scored by computer, and test takers are able to see their scores online approximately 10 days after taking the test. At the request of the test takers, scores are sent to the academic institutions to which they apply (ETS, 2015).

Oral and written responses are utilized in the integrated speaking and writing sections, as it is important that test takers demonstrate comprehension of the listening text as well as the ability to thoroughly note the relevant parts and then use them in the written response. The written and spoken responses are marked by human and machine raters. As mentioned earlier, Sawaki *et al.*'s (2008) study suggests that test takers' performance on these sections does not rely heavily on their listening skills.

Development

The TOEFL iBT was developed using the evidence-based approach to assessment (Chapelle *et al.*, 2008), which was designed by Mislevy, Almond, and Lukas (2003). Evidence gathered from multiple papers and research projects was used by ETS to reformulate the test design and develop a system to assess the academic communication skills of the test takers. All of the test items and tasks are developed using the test development cycle of ETS. The items and tasks are written by experienced item writers who are supposedly competent in item writing and have a good understanding of the fairness guidelines of ETS. After this stage, the content, fairness, and formatting of the items are reviewed by ETS assessment specialists, and the approved items are pretested. Pretesting is performed by incorporating the newly developed items into actual tests, but the candidates are unaware of the items being pretested (Chapelle *et al.*, 2008).

Reliability

As TOEFL iBT tests are administered globally, numerous test formats must be developed, and their comparability and reliability have to be ascertained by ETS. According to ETS (2011a), five methods are used to check and maintain test score reliability, and listening tests also go through this process: (a) standardization of the administration and security process, (b) use of precise test specification guidelines, (c) use of proper scales for score reporting, (d) use of test equating techniques, and (e) monitoring of reliability and generalizability of test scores. ETS has attempted to reduce the standard error of measurement of the test forms using item response theory (IRT) modeling and generalizability theory. The average reliability estimate of the listening test in 2007 was .85 with a standard error of measurement of 3.20, indicating that, on average, 15% (3.20) of an obtained listening score is attributed to error of measurement. For example, the actual listening ability score of a test candidate achieving 25/30 would fall within 25 ± 3.20. This does not seem to be a sufficient level of precision for a high-stakes test of listening.

Similar results were recently obtained by Sawaki and Sinharay (2013), who examined the factor structure, reliability, and generalizability of the TOEFL iBT across Arabic, Korean, and Spanish-speaking test takers. Sawaki and Sinharay calculated the proportional reduction in mean squared error (PRMSE), which is similar to Cronbach's alpha internal consistency index and ranges from zero to one, with higher indices indicating higher precision. They found that the listening section had a PRMSE index between .81 and .87. As the PRMSE of the listening section was significantly lower than that of the overall test, Sawaki and Sinharay argued that the listening score hardly contributed any added value to candidates' total TOEFL iBT scores. By contrast, the speaking subsection of the test had a much larger added value. The generalizability coefficient of the listening section was found to be satisfactory (> .80), but there was evidence that the listening texts were of varying difficulty across test forms and test-taking groups.

Validity

The most extensive exploration of validity evidence for the TOEFL iBT is by Chapelle *et al.* (2008), and a shorter report is available from ETS (2011b). Chapelle *et al.* drew on Kane's (2006) validity argument framework and examined several inferences drawn using the data and research results. In the validation project, Bejar *et al.*'s (2000) listening framework was extended and revised to accommodate several listening abilities such as "basic understanding," "pragmatic understanding," and "integrating information" (Enright *et al.*, 2008, p. 101). Enright *et al.* alluded to small-scale listening studies that showed the pragmatic test items were as difficult as expected; however, the difficulty level of the items per se did not demonstrate what language skills were tapped by the items.

Soon after Chapelle *et al.*'s (2008) study, Sawaki and Nissan (2009) developed a test of academic listening comprehension consisting of three lengthy academic lectures (each 30 minutes) administered to 120 undergraduate and 64 graduate students at four American universities. They found medium correlations between participants' listening TOEFL iBT test scores and their performance on the three lectures (correlations of .56 for undergraduate participants and .64 for graduate students). The correlation coefficients are not strong enough to support the validity argument of the listening test, although the results should not be completely dismissed because they still show some relation between the two measures. For high-stakes tests such as the TOEFL iBT, obtaining a correlation coefficient greater than .80 would be much more desirable.

Another stream of TOEFL iBT research involves the factor structure of the test. Although it has been assumed that the test sections are discrete, Stricker, Rock, and Lee (2005) found that listening, reading, and writing loaded on one general factor, which seems to attenuate the argument for discriminant validity of these separate sections. In contrast, Sawaki and Sinharay (2013) found that listening loads on a separate factor, thereby contradicting Sticker *et al.*'s finding and supporting the presence of multidivisible language skills. In sum, the TOEFL iBT listening test has been rigorously researched, but the findings are mixed and do not always support the measurement model proposed by the test developers.

Availability

The actual listening tests are only made available to candidates who register for the test. Test registration can be done online (https://www.ets.org/toefl/ibt/register), by email, or by phone. ETS does, however, provide ample practice test materials, some of which are free, and some of which are available for a fee. Candidates who wish to sit for a practice test that uses actual test materials can register on the ETS website (http://toeflpractice.ets.org/). Other test preparation materials developed by ETS and other organizations are available for purchase online.

Sample Studies

As seen in the studies described above, the TOEFL iBT has a rich research history, and the listening section in particular has attracted researchers' attention. For example, Rosenfeld, Leung, and Oltman (2001) drew upon several groups of potential stakeholders of the TOEFL iBT, including undergraduate and graduate programs in 21 major American universities and one Canadian university. Overall, both faculty members and students rated some listening tasks much lower than others. Examples of tasks receiving low ratings included: distinguishing among "communicative functions" such as offer, advice, and suggestion; understanding the use of "examples, anecdotes, jokes, and digressions"; and recognizing different tones such as sarcasm and jokes (p. 17). In contrast, tasks such as understanding facts, details, and information as well as main ideas, instructions, and specific terminology were rated relatively highly. Although inference-making skills were not always among the most highly endorsed skills, they were still recognized as significant listening skills by many participants. The authors argued that the results would lend support to the theoretical framework of the TOEFL iBT, where the highly endorsed listening skills were primarily tapped by the test.

One of the components of the TOEFL iBT listening test is note taking. Although Rosenfeld *et al.* (2001) found consistency between university students' and professors' perception of significant skills with the theoretical framework of the listening test of the TOEFL iBT, Carrell (2007) found that participants took notes on a significantly smaller proportion of the recognized important information, they used fewer symbols and abbreviations in their notes, and the notes taken by all participants contained answers to approximately 20% of the listening test items. Carrell suggested that these findings likely indicate a discrepancy between what test takers and test developers recognize as important in the tests. A brief intervention helped increase the use of note-taking strategies with modest effect size, although it did not have a significant impact on the participants' listening scores. Similarly, the intervention changed the perception of the participants regarding the use of note taking in listening, but their perception regarding the usefulness of note taking remained the same. Carrell argued that listening and note taking are discrete skills and teachers who are preparing students for the TOEFL iBT should spend significant time on note taking. Although focusing on note taking can have advantages for students, Carrell suggested that test developers should exercise caution when integrating listening and note taking, because the simultaneous use of these skills can tax listeners' cognitive resources, thereby affecting the test with likely construct-irrelevant variance.

Readers interested in further information regarding the TOEFL iBT listening test are referred to Chapelle *et al.*'s (2008) book that synthesizes the available literature and incorporates recent evidence into a validity argument framework.

Critique

Three plausible causes for concern regarding the listening test of the TOEFL iBT are: (a) the aforementioned low reliability and generalizability of the test scores, (b) the lack of differential item functioning (DIF) analysis, and (c) the need for test takers to answer several trial items that do not count toward their performance. DIF analysis is one of the most effective techniques for examining test fairness. DIF analysis seems to be particularly relevant to the listening test of the TOEFL iBT because, as noted in this profile, Sawaki and Sinharay (2013) found that the reliability and generalizability of test scores fluctuated across different subsamples.

Although incorporating trial items into the TOEFL iBT is one way to ascertain the reliability of test item scores before putting them to use, requiring candidates to answer extra items can cause fatigue and introduce construct-irrelevant factors that influence performance. As a result of this practice, a compromise has been made that seems to affect test takers. It is important that test developers keep the number of trial items to a minimum to prevent the undesirable impact of fatigue.

References

Aryadoust, V. (2012). Differential item functioning in while-listening performance tests: The case of IELTS listening test. *International Journal of Listening*, *26*, 40–60. doi:10.1080/10904018.2012.639649

Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (2000). *TOEFL® 2000 listening framework: A working paper* (TOEFL® Monograph No. MS-19). Princeton, NJ: ETS.

Carrell, P. L. (2007). *Notetaking strategies and their relationship to performance on listening comprehension and communicative assessment tasks* (TOEFL® Monograph No. RR-07-01, TOEFL-MS-35). Princeton, NJ: ETS.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.) (2008). *Building a validity argument for the Test of English as a Foreign Language*. New York: Routledge.

Enright, M. K., Bridgeman, B., Eignor, D., Kantor, R. N., Mollaun, P., Nissan, S., Powers, D. E., & Schedl, M. (2008). Prototyping new assessment tasks. In C. A. Chapelle, J. Jamieson, & M. K. Enright (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 145–186). London: Routledge.

Educational Testing Service (ETS). (2011a). *TOEFL iBTTM research insight: Reliability and comparability of TOEFL iBT scores*. Retrieved from https://www.ets.org/s/toefl/pdf/toefl_ibt_research_s1v3.pdf

Educational Testing Service (ETS). (2011b). *TOEFL iBTTM research insight: Validity evidence supporting the interpretation and use of TOEFL iBT scores* (Series 1, Vol. 4). Retrieved from https://www.ets.org/s/toefl/pdf/toefl_ibt_insight_s1v4.pdf

Educational Testing Service (ETS). (2015). *TOEFL iBT® test questions*. Retrieved from https://www.ets.org/Media/Tests/TOEFL/pdf/SampleQuestions.pdf

Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (2000). *TOEFL 2000 framework: A working paper* (TOEFL Monograph Series Report No. 16). Princeton, NJ: Educational Testing Service.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Washington, DC: The National Council on Measurement in Education & the American Council on Education.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design* (Research Report 03-16). Princeton, NJ: ETS.

Rosenfeld, M., Leung, S., Oltman, P. K. (2001). *The reading, writing, speaking, and listening tasks important for academic success at the undergraduate and graduate levels* (TOEFL iBT™ Report No. RM-01-03, TOEFL-MS-21). Princeton, NJ: ETS.

Sawaki, Y., & Nissan, S. (2009). *Criterion-related validity of the TOEFL® iBT listening section* (TOEFL iBT™ Report No. iBT-08). Princeton, NJ: ETS.

Sawaki, Y., & Sinharay, S. (2013). *Investigating the value of section scores for the TOEFL iBT® test* (TOEFL iBTTM Report No. iBT-21). Princeton, NJ: ETS.

Sawaki, Y., Stricker, L. J., & Oranje, A. (2008). *Factor structure of the TOEFL® Internet-based test (iBT): Exploration in a field trial sample* (TOEFL iBT™ Report No. *iBT-04).* Princeton, NJ: ETS.

Stricker, L. J., Rock, D. A., & Lee, Y.-W. (2005). *Factor structure of the LanguEdge™ test across language groups* (TOEFL® Monograph No. MS-32). Princeton, NJ: ETS.

Profile 62

Verbal Aggressiveness Scale (VAS)

(Infante & Wigley, 1986)

Profiled by: Timothy R. Levine, PhD

University of Alabama at Birmingham

Construct

The Verbal Aggressiveness Scale (VAS) was designed to measure trait verbal aggressiveness, or the predisposition to attack the self-concept of others.

Instrument Type

Self-Report

Description

The VAS is a self-report instrument designed to measure trait verbal aggressiveness. The scale was intended to be a unidimensional, 20-item measure with 10 reverse-scored items. The scale uses a 5-point response format anchored by *almost never true* on the low end and *almost always true* on the high aggressiveness end. Higher scores indicate a predilection toward aggressive, hurtful communication.

Administration

Brief instructions ask respondents to rate each of 20 items on a 5-point scale reflecting how often each statement is true for the respondent when he or she is trying to influence other people. Following instructions, a 5-point response scale is provided with 1 = *almost never true*, 2 = *rarely true*, 3 = *occasionally true*, 4 = *often true*, and 5 = *almost always true*.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

The 20 items follow. Respondents rate each of the 20 items using the 5-point scale. Administration time is less than 5 minutes.

Scoring

The scale was intended to be unidimensional across all 20 items. The original intended scoring involves first reverse scoring the 10 items that depict nonaggressive communication, and then summing all items. Some authors have argued that only the 10 aggressively worded items should be scored by summing the 10 items (Levine *et al.*, 2004). A recent study suggested that verbal aggressiveness is best measured by only four items (Beatty, Pascual-Ferra, & Levine, 2015).

Development

The VAS was first introduced by Infante and Wigley (1986) to empirically distinguish between verbally aggressive and argumentative communication. The original article described the item selection and retention process, tested and replicated reliability, reported factor analytic results, correlated the scale with other dimensions of personality, and found that the scale predicted the self-reported ratings of verbal aggressive messages. Since this study, the scale has been used widely and has come under close scrutiny. Currently, best practice suggests using only 10 of the original 20 items.

Reliability

Infante and Wigley (1986) reported reliabilities (coefficient alpha) of .81 in both of the first two original studies. The average reliability for the 20-item scoring is .84 (Hamilton & Mineo, 2001). The reliability for scoring the 10 aggressive items is .82 (Levine *et al.*, 2004). There is general agreement that the scale has acceptable reliability for the original 20-item and the 10-item scoring.

Validity

The validity of the VAA has been the subject of considerable debate (Beatty *et al.*; Beatty, Rudd, & Valencic, 1999; Infante, Rancer, & Wigley, 2011; Kotowski, Levine, Baker, & Bolt, 2009; Levine *et al.*, 2004; Levine, Kotowski, Beatty, & Van Kelegom, 2012). In self-report research, the VAS appears to be construct valid, consistently correlating with other self-reported measures in a coherent and predictable manner. The controversy has centered on two issues: the dimension of the scale and predictive-convergent validity.

The scale was originally intended to be scored as unidimensional. Several studies (e.g., Beatty *et al.*, 1999; Infante & Wigley, 1986; Kotowski *et al.*, 2009; Levine *et al.*, 2004), however, including the original validation work, have found two factors with the aggressively worded items on one factor and the reflected items on a second factor.

Infante and colleagues (Infante *et al.*, 2011; Infante & Wigley, 1986) dismiss the second factor as an artifact, interpreting the scale as unidimensional even though two factors were found. In contrast, Beatty *et al.* (1999), Levine *et al.* (2004), and Kotowski *et al.* (2009) contended that the reflected items measure benevolent communication rather than a mere absence of aggressiveness. Most recently, Beatty *et al.* (2015) provided evidence that neither the one-factor model nor the two-factor model provides a clean fit to the data, and that both models are confounded by some unknown factors apart from the artifact originally suggested by Infante and Wigley (1986). Their report suggests that only four of the original items are valid indicators of verbal aggressiveness.

A second point of contention is the extent to which scores on the VAS predict and converge with observations of verbally aggressive communication (Kotowski *et al.*, 2009). Infante *et al.* (2011) claimed that the scale predicts behavior as well as any trait–behavior association, but a meta-analysis by Levine *et al.* (2012) showed that the scale only predicts self-reported communication and not actual aggressive behavior.

Availability

The scale, instructions, and scoring instructions are provided in the original Infante and Wigley (1986) published article. These elements are reproduced here with permission. The scale is free to use for research purposes.

Sample Studies

The VAS has been widely used in self-report research in interpersonal, organizational, and instructional communication. For example, Rocca and McCroskey (1999), using a modified version of the scale, found verbal aggressiveness was negatively associated with ratings of similarity and attractiveness in instructional settings. As a second example, Boster and Levine (1988) reported that scores on the VAS were positively associated with the endorsements of compliance-gaining strategies, but the nature of the impact was situationally dependent.

With respect to listening research, Schrodt and Wheeless (2001) found that verbal aggressiveness was not associated with reports of listening apprehension. As we might expect, Villaume and Bodie (2008) found that verbal aggressiveness was associated with lower levels of people-oriented listening (as measured by the Listening Style Profile [LSP-16]). Similarly, Worthington (2005) reported that verbal aggressiveness was associated with reduced preference for people and content listening. It should be noted, however, that the validity and reliability of the LSP-16 have been questioned (see Profile 36).

Critique

The VAS has been widely used in communication research, and it attempts to measure an important construct. It is easy to administer, and it is applicable to a variety of contexts. It consistently produces internally consistent scores, it typically correlates as expected with other self-report measures, and it has produced a coherent research literature.

Although the scale was developed as a unidimensional measure, it is clear that the original 20 items are not unidimensional. A two-factor solution consistently fits the data better than the intended unidimensional model, and obtaining a close fit either requires correlating error terms or removing items. Thus, the VAS is not psychometrically sound.

Perhaps the most worrisome findings are the failure of the VAS to predict observed verbally aggressive behaviors. The scale correlates much better with self-reported actions than actual behavior. Because verbal aggressiveness has been conceptualized as a behavioral predisposition, scores do not appear to measure the intended construct (Levine *et al.*, 2012). Thus, the VAS lacks convergent validity with behavioral measures.

References

Beatty, M. J., Pascual-Ferra, P., & Levine, T. R. (2015). Two studies examining the error theory underlying the measurement model of the Verbal Aggressiveness Scale. *Human Communication Research*, *41*, 55–81. doi:10.1111/hcre.12039

Beatty, M. J., Rudd, J. E., & Valencic, K. M. (1999). A re-examination of the Verbal Aggressiveness Scale: One factor or two? *Communication Research Reports*, *16*, 10–17. doi:10.1080/08824099909388696

Boster, F. J., & Levine, T. R. (1988). Individual differences and compliance gaining message selection: The effects of verbal aggressiveness, argumentativeness, dogmatism, and negativism. *Communication Research Reports*, *5*, 114–119. doi:10.1080/08824098809359811

Hamilton, M. A., & Mineo, P. J. (2001). Argumentativeness and its effect on verbal aggressiveness: A meta-analytic review. In M. Allen, R. W. Preiss, B. M. Gayle, & N. Burrell (Eds.), *Interpersonal communication research: Advances through meta-analysis* (pp. 281–314). Mahwah, NJ: LEA.

Infante, D. A., Rancer, A. S., & Wigley, C. J. (2011). In defense of the argumentativeness and verbal aggressiveness scales. *Communication Quarterly*, *59*, 145–154. doi:10.1080/01463373.2011.563439

Infante, D. A., & Wigley, C. J., III. (1986). Verbal aggressiveness: An interpersonal model and measure. *Communication Monographs*, *53*, 63–69. doi:10.1080/03637758609376126

Kotowski, M. R., Levine, T. R., Baker, C. R., & Bolt, J. M. (2009). A multitrait-multimethod validity assessment of the verbal aggressiveness and argumentativeness scales. *Communication Monographs*, *76*, 443–462. doi:10.1080/03637750903300247

Levine, T. R., Beatty, M. J., Limon, S., Hamilton, M. A., Buck, R., & Chory-Assad, R. M. (2004). The dimensionality of the Verbal Aggressiveness Scale. *Communication Monographs*, *71*, 245–268. doi:10.1080/0363452042000299911

Levine, T. R., Kotowski, M. R., Beatty, M. J., & Van Kelegom, M. J. (2012). A meta-analysis of trait-behavior correlations in argumentativeness and verbal aggression. *Journal of Language and Social Psychology*, *31*, 95–111. doi:10.1177/0261927X11425037

Rocca, K. A., & McCroskey, J. C. (1999). The interrelationship of student ratings of instructor's immediacy, verbal aggressiveness, homophily, and interpersonal attraction. *Communication Education*, *48*, 308–316. doi:10.1080/03634529909379181

Schrodt, P., & Wheeless, L. R. (2001). Aggressive communication and informational reception apprehension: The influence of listening anxiety and intellectual inflexibility on train argumentativeness and verbal aggression. *Communication Quarterly*, *49*, 53–69. doi:10.1080/01463370109385614

Villaume, W. A., & Bodie, G. D. (2008). Discovering the listener within us: The impact of trait-like personality variables and communicator style on preferences for listening style. *International Journal of Listening, 21*, 102–123. doi:10.1080/10904010701302006
Worthington, D. L. (2005). Exploring the relationship between listening style preference and verbal aggressiveness. *International Journal of Listening, 19*, 3–11. doi:10.1080/10904018.2005.10499069

Scale

Verbal Aggressiveness Scale (Infante & Wigley, 1986)

Instructions: This survey is concerned with how we try to get people to comply with our wishes. Indicate how often each statement is true for you personally when you try to influence other persons. Use the following scale:

1 = Almost never true
2 = Rarely true
3 = Occasionally true
4 = Often true
5 = Almost always true

<u>Benevolent Items</u>

1) I am extremely careful to avoid attacking individuals' intelligence when I attach their ideas.
2) I try very hard to avoid having other people feel bad about themselves when I try to influence them.
3) When others do things I regard as stupid, I try to be extremely gentle with them.
4) I try to make people feel good about themselves even when their ideas are stupid.
5) When people criticize me shortcomings, I take it in good humor and do not try to get back at them.
6) When I dislike individuals greatly, I try not to show it in what I say or how I say it.
7) When I attack persons' ideas, I try not to damage their self-concepts.
8) When I try to influence people, I make efforts not to offend them.
9) I refuse to participate in arguments when they involve personal attacks.
10) When an arguments shifts to personal attacks, I try very hard to change the subject.

<u>Verbally Aggressive Items</u>

11) When individuals are very stubborn, I use insults to soften the stubbornness.*
12) When people refuse to do a task I know is important, without good research, I tell them they are unreasonable.
13) If individuals I am trying to influence really deserve it, I attack their character.*
14) When people behave in ways that are in very poor taste, I insult them in order to shock them into proper behavior.
15) When people will not budge on a matter of importance I lose my temper and say rather strong things to them.
16) When individuals insult me, I get a lot of pleasure out of really telling them off.

17) I like poking fun at people who do things which are very stupid in order to stimulate their intelligence.
18) When people do things that are mean or cruel, I attack their character in order to help correct their behavior.*
19) When nothing seems to work in trying to influence others, I yell and scream in order to weaken their positions.*
20) When I am not able to refute others' positions, I try to make them feel defensive in order to weaken their positions.

Note: The original scale was intended to be unidimensional across all 20 items, thus scoring involves first reverse coding the 10 benevolent items, and then summing all items. Labels should be removed and items randomized prior to administration. Items marked with an asterisk (*) were found free of confounds in Beatty *et al.* (2015).

Profile 63

Taxonomy of Verbal Response Modes (VRM)

Profiled by: William B. Stiles, PhD

Miami University and Appalachian State University

Construct

The Verbal Response Modes (VRM) taxonomy (Stiles, 1992) is a general-purpose classification of speech acts. It concerns what people do when they say something rather than the content of what they say. It can be used to describe the relationship of speaker to other (i.e., author to addressee) in any sort of discourse.

Instrument Type

Behavioral Assessment; Verbal Coding System

Description

Listening competence involves verbal as well as nonverbal behaviors that show how a listener is paying attention and understanding his or her interlocutor (Bodie, St. Cyr, Pence, Rold, & Honeycutt, 2012). The VRM system is a classification of verbal responses in spoken language that is sensitive to this aspect of relationships. See, for example, the discussion of attentiveness later.

The VRM coding unit is the utterance, defined as a simple sentence; independent clause; nonrestrictive dependent clause; multiple predicate; or term of acknowledgment, evaluation, or address within a specified text. Each utterance is coded as Reflection (R), Acknowledgment (K), Interpretation (I), Question (Q), Confirmation

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

(C), Edification (E), Advisement (A), or Disclosure (D). These VRM codes are assigned according to three principles of classification, each of which can take the value of other or speaker:

1) *Source of experience*: whether the utterance's topic is information held by the other or by the speaker;
2) *Frame of reference*: whether the utterance is expressed from a point of view shared with the other or from the speaker's own point of view; and
3) *Presumption*: whether the speaker presumes knowledge of what the other's experience or frame of reference is, was, will be, or should be (other), or instead uses knowledge only of his or her own experience and frame of reference (speaker).

As shown in Table P63.1, these three forced choices place every utterance into one of the eight mutually exclusive categories, which are exhaustive in the sense that every comprehensible utterance can be coded. The designation *uncodable* (U) is used only for utterances that are incomprehensible.

In VRM coding, each utterance is coded twice, once for its grammatical form and once for its pragmatic intent, using the same eight categories. Form and intent definitions are shown in Table P63.1. An utterance that has the same form and intent is called a *pure mode*. For example, "I have pain when I move my legs" would be coded as disclosure form (first-person singular) and disclosure intent (reveals subjective experience), abbreviated DD. Alternatively, an utterance can be a *mixed mode*, having the form of one mode and the intent of another. For example, "I went to the emergency room last week" would be coded as disclosure form (first-person singular) and edification intent (transmits objective information), abbreviated DE. "Would you close the window?" is question form with advisement intent (QA).

Administration

VRM can be coded from written documents, verbatim transcripts, audio or video recordings, or live interactions. Coding of complex or rapidly moving interaction is difficult; coders working from recordings often need to replay them several times to catch all utterances.

VRM form codes are based on grammatical features, so grammatical, non-elliptical utterances can be coded in isolation. In natural speech, elliptical, incomplete, and ungrammatical utterances require reference to context, although usually the immediately preceding few utterances are sufficient. VRM intent codes classify the speaker's intended meaning and therefore must always be understood in context. In practice, VRM intent can usually be coded in a context of a few preceding utterances, but some utterances may be understandable only in the context of earlier events (e.g., in longstanding relationships).

The VRM system allows any size of summarizing unit. Most VRM studies have summarized over an encounter or a segment of an encounter. Utterances have most often been aggregated separately for each speaker; however, it is possible to aggregate by dyad or by larger groups. In some applications, codes of single utterances are reported.

Table P63.1 Taxonomy of verbal response modes

Source of experience	**Frame of reference**	**Presumption**	
		Other	**Speaker**
Other	Other	REFLECTION (R) *Form*: Second person; verb implies internal experience or volitional action. *Intent*: Puts other's experience into words; repetitions, restatements, clarifications.	ACKNOWLEDGMENT (K) *Form*: Nonlexical or contentless utterances; terms of address or salutation. *Intent*: Conveys receipt of or receptiveness to other's communication; simple acceptance, salutations.
Other	Speaker	INTERPRETATION (I) *Form*: Second person ("you"); verb implies an attribute or ability of the other; terms of evaluation. *Intent*: Explains or labels the other; judgments or evaluations of other's experience or behavior.	QUESTION (Q) *Form*: Interrogative, with inverted subject–verb order or interrogative words. *Intent*: Requests information or guidance.
Speaker	Other	CONFIRMATION (C) *Form*: First-person plural ("we") where referent includes other. *Intent*: Compares speaker's experience with other's; agreement, disagreement, shared experience or belief.	EDIFICATION (E) *Form*: Declarative; third person (e.g., "he," "she," or "it"). *Intent*: States objective information.
Speaker	Speaker	ADVISEMENT (A) *Form*: Imperative, or second person with verb of permission, prohibition, or obligation. *Intent*: Attempts to guide behavior; suggestions, commands, permission, prohibition.	DISCLOSURE (D) *Form*: Declarative; first-person singular ("I") or first-person plural ("we") where other is not a referent. *Intent*: Reveals thoughts, feelings, wishes, perceptions, or intentions.

Note: Both the form and intent of each utterance are coded. For example, "Would you close the window?" is question form with advisement intent (abbreviated QA).

Scoring

The VRM system offers a variety of aggregate measures for characterizing an encounter, including the following three areas:

1) The frequency or percentage of each form or intent; for example, the frequency of Acknowledgment form aggregated across intents or the percentage of Edification intent aggregated across forms.

2) Three role dimensions, which are labeled as (a) Informativeness versus Attentiveness, (b) Unassumingness versus Presumptuousness, and (c) Directiveness versus Acquiescence, correspond to the proportion of speaker versus other values on source of experience, presumption about experience, and frame of reference, respectively.

Each of the taxonomy's eight basic modes is considered as either informative or attentive, either directive or acquiescent, and either presumptuous or unassuming. For a passage of any length, indexes of each role dimension for each speaker can be calculated as the proportion of the speaker's coded utterances in the designated modes. Role dimension indexes can be calculated separately for form and intent, or averaged across form and intent.

Constituent Verbal Response Modes of Role Dimensions

Informativeness	Confirmation, Edification, Advisement, Disclosure
Attentiveness	Reflection, Acknowledgment, Interpretation, Question
Directiveness	Interpretation, Question, Advisement, Disclosure
Acquiescence	Reflection, Acknowledgment, Confirmation, Edification
Presumptuousness	Reflection, Interpretation, Confirmation, Advisement
Unassumingness	Acknowledgment, Edification, Question, Disclosure

Role Dimensions Are Arranged in Complementary Pairs

Attentiveness =	1 - Informativeness
Acquiescence =	1 - Directiveness
Unassumingness =	1 - Presumptuousness

The role dimensions are parallel to the principles of classification (see Table P63.1). Attentiveness and informativeness are based on the source of experience classification principle. Interpersonally, attentiveness has to do with manifest interest in the other and attempts to ensure that the other's thoughts are expressed and considered in the conversation, whereas informativeness has to do with providing information to the other. Acquiescence and directiveness are based on the frame of reference classification principle. Interpersonally, acquiescence has to do with acceding to the other's viewpoint, whereas directiveness measures the degree to which the speaker guides the conversation by using his or her own viewpoint. Presumptuousness and unassumingness are based on the presumption classification principle. Interpersonally, presumptuousness has to do with higher relative status, knowing the other, or assuming that one is important to the other, whereas unassumingness has to do with lower status and deference.

The VRM coding system thus considers each utterance as simultaneously representing one or the other pole on all three of the role dimensions. For example, an edification such as "The accident was on the ninth of September" (EE) is considered as simultaneously informative, acquiescent, and unassuming. A question, such as "Was it a pretty bad car accident?" (QQ), is considered as attentive, directive, and unassuming. An advisement, such as "Now turn this way" (AA), is considered as informative, directive, and presumptuous.

3) The frequency or percentage of each pure or mixed mode, for example the frequency of KK or the percentage of DE.

Mixed modes offer relatively subtle ways of representing relational aspects of verbal exchange. For example, "Could you scoot forward a bit?" (QA) is directive in both form and intent; it is attentive and unassuming in form (question) but informative and presumptuous in intent (advisement). As a result, it is subtly politer than its pure-mode counterpart, “Scoot forward a bit” (AA).

Development

The VRM taxonomy was elaborated by William B. Stiles (see Stiles, 1992, chap. 1) based on a framework for help-intended interpersonal communication developed by Gerald Goodman (Goodman & Dooley, 1976). Originally developed for research on the process of psychotherapy, it evolved into a system that can be applied to any discourse.

Reliability

VRM coding reliability is assessed as interrater agreement and thus is dependent on coders’ ability, training, and experience and on the nature and variability of the material being coded. Experienced coders can achieve high reliability on most sorts of conversations (for illustrations, see Stiles, 1992, chap. 11). Agreement of 95% of utterances for VRM form codes and 85% for VRM intent codes is common.

Validity

The VRM codes are descriptive categories, and it makes little sense to ask if they are valid (e.g., is the question category a valid measure of being a question?). Some people might define the terms differently, however (e.g., disclosure might be defined differently by different investigators).

It makes more sense to ask about derivative indexes, such as the role dimensions. Studies have supported the role dimensions’ construct validity by showing that people in roles expected to be attentive or informative, acquiescent or directive, or presumptuous or unassuming tend to obtain scores consistent with those expectations (for a review, see Stiles, 1992, chap. 4). For example, Interviewers (e.g., doctors, psychotherapists, and courtroom interrogators) are much more attentive than interviewees (patients, clients, and witnesses). Nondirective therapists are much more acquiescent than directive therapists, and patients are more acquiescent than doctors. In mixed-status dyads (e.g., teacher–student, senior–freshman, psychotherapist–client, and doctor–patient), the higher status member is consistently more presumptuous than the lower status member. Conversely, perhaps because it conveys relative status, conversations between social equals are marked by extremely close concordance in presumptuousness; that is, interactants copy each other’s levels of presumptuousness within very close tolerances (Stiles *et al.*, 1997).

Availability

The book *Describing Talk* (Stiles, 1992) includes a detailed coding manual. This book is out of print but available in many libraries. In addition, a prepublication version is available free online at http://www.users.miamioh.edu/stileswb/archive.htmlx.

A computer-based VRM coder training program offers detailed instruction in VRM principles along with coding and practice using examples from a variety of types of discourse, with utterance-by-utterance feedback. The program is dated, written for 16-bit DOS in the 1980s, and the graphics are primitive, but people who complete this successfully will be competent VRM coders. It will run on 32-bit versions of Windows, and it will run on 64-bit versions of Windows with free DOS-emulator software. The training program and instructions for installing it are also available free online at http://www.users.miamioh.edu/stileswb/archive.htmlx. See Stiles (1992) for some suggested training procedures.

Sample Studies

The VRM coding system has demonstrated applicability in a variety of contexts, including professional service encounters such as medical interaction (Cape & Stiles, 1998; Meeuwesen, Schaap, & van der Staak, 1991; Shaikh, Knobloch, & Stiles, 2001) and psychotherapy (Anderson, Knobloch-Fedders, Stiles, Ordonez, & Heckman, 2012; Stiles & Shapiro, 1995); public discourse, such as presidential speeches (Miller & Stiles, 1986), labor–management negotiations (Hinkle, Stiles, & Taylor, 1988), and radio call-in programs (Henricks & Stiles, 1989); and a wide variety of ordinary conversations. Studies have found systematic relations between VRM indices and a variety of interpersonal relationship variables. For example, in brief initial interactions, both men and women used more disclosure if they were attractive or if their partner was attractive (Stiles, Walz, Schroeder, Williams, & Ickes, 1996). In laboratory conversations, women tended to be more attentive than men under some conditions, particularly within committed relationships such as married or dating couples (Stiles *et al.*, 1997). High and moderate trait-anxious university students (but not the low trait-anxious students) used higher percentages of disclosure when speaking about an anxiety-arousing topic than when speaking about a happy topic (Stiles, Shuster, & Harrigan, 1992).

Critique

The subtlety of the interpersonal processes captured by VRM coding comes at a cost of very careful and detailed coding. Learning the system can take 30–40 hours, including completing the computer-assisted training program described in the Availability section, for the undergraduate volunteers who served as coders in many of the cited VRM studies (Stiles, 1992). Researchers themselves must learn the system in order to write about it well. The complexity is the greatest barrier to using the system. Of course, the counterargument is that the complexity of VRM merely reflects the complexity and subtlety of interpersonal relationships as enacted verbally.

References

Anderson, T., Knobloch-Fedders, L., Stiles, W. B., Ordonez, T., & Heckman, B. D. (2012). The power of subtle interpersonal hostility in psychodynamic psychotherapy: A speech acts analysis. *Psychotherapy Research*, *22*, 348–362. doi:10.1080/10503307.2012.658097

Bodie, G. D., St. Cyr, K., Pence, M., Rold, M., & Honeycutt, J. M. (2012). Listening competence in initial interactions I: Distinguishing between what listening is and what listeners do. *International Journal of Listening*, *26*, 1–28. doi:10.1080/10904018.2012.639645

Cape, J. D., & Stiles, W. B. (1998). Verbal exchange structure of general practice consultations with patients presenting psychological problems. *Journal of Health Psychology*, *3*, 5–21. doi:10.1177/135910539800300101

Goodman, G., & Dooley, D. (1976). A framework for help-intended interpersonal communication. *Psychotherapy: Theory, Research and Practice*, *13*, 106–117. doi:10.1037/h0088322

Henricks, W. H., & Stiles, W. B. (1989). Verbal processes on psychological radio call-in programs: Comparison with other help-intended interactions. *Professional Psychology: Research and Practice*, *20*, 315–321. doi:10.1037/0735-7028.20.5.315

Hinkle, S., Stiles, W. B., & Taylor, L. A. (1988). Verbal processes in a labour/management negotiation. *Journal of Language and Social Psychology*, *7*, 123–136. doi:10.1177/02619 27X8800700203

Meeuwesen, L., Schaap, C., & van der Staak, C. (1991). Verbal analysis of physician-patient communication. *Social Science & Medicine*, *32*, 1143–1150. doi:10.1016/0277-9536(91)90091-P

Miller, N. L., & Stiles, W. B. (1986). Verbal familiarity in American presidential nomination acceptance speeches and inaugural addresses (1920–1981). *Social Psychology Quarterly*, *49*, 72–81.

Shaikh, A., Knobloch, L. M., & Stiles, W. B. (2001). The use of a verbal response mode coding system in determining patient and physician roles in medical interviews. *Health Communication*, *13*, 49–60. doi:10.1207/S15327027HC1301_05

Stiles, W. B. (1992). *Describing talk: A taxonomy of verbal response modes*. Newbury Park, CA: Sage.

Stiles, W. B., Lyall, L. M., Knight, D. P., Ickes, W., Waung, M., Hall, C. L., & Primeau, B. E. (1997). Gender differences in verbal presumptuousness and attentiveness. *Personality and Social Psychology Bulletin*, *23*, 759–772. doi:10.1177/0146167297237009

Stiles, W. B., & Shapiro, D. A. (1995). Verbal exchange structure of brief psychodynamic-interpersonal and cognitive-behavioral psychotherapy. *Journal of Consulting and Clinical Psychology*, *63*, 15–27. doi:10.1037//0022-006X.63.1.15

Stiles, W. B., Shuster, P. L., & Harrigan, J. A. (1992). Disclosure and anxiety: A test of the fever model. *Journal of Personality and Social Psychology*, *63*, 980–988. doi:10.1037/0022-3514.63.6.980

Stiles, W. B., Walz, N. C., Schroeder, M. A. B., Williams, L. L., & Ickes, W. (1996). Attractiveness and disclosure in initial encounters of mixed-sex dyads. *Journal of Social and Personal Relations*, *13*, 305–314. doi:10.1177/0265407596132009

Profile 64

Watson-Barker Listening Test (WBLT)

(Watson & Barker, 1983, 1988)

Profiled by: Debra L. Worthington, PhD

Auburn University

Construct

The Watson-Barker Listening Test was designed to measure listening comprehension as well as short-term and long-term listening ability.

Instrument Type

Cognitive Assessment

Description

The Watson-Barker Listening Test (WBLT) was conceived as a means to measure five facets of adult listening behavior—interpretation of meaning, interpretation of emotion, understanding, recall, and the ability to follow instructions (Watson & Barker, 1988; Watson, Barker, Roberts, & Roberts, 2001). The 40 multiple-choice items that comprise the test are divided evenly among each of the five areas, and two forms of the test are provided to allow for test–retest capability. The instrument incorporates background noise, includes conversations, and incorporates a variety of listening contexts (e.g., conversations, lectures, etc.). Speakers on the video differ by gender and speak with different regional dialects. The first three sections of the test were designed to test a listener's short-term listening ability, and the remaining two test for long-term listening.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

Administration

The most recent version of the WBLT (Forms E and F) are presented via a DVD. For each version, participants view five vignettes. Assessment questions for each of the vignettes are also presented via DVD. The entire test takes approximately 40 minutes to present. Participants use a self-scoring response sheet to answer a series of eight multiple-choice-style questions associated with each vignette.

Scoring

Answers to each of the 40 items are scored as either correct or incorrect. Participants receive subscores for each of the five areas (i.e., interpretation of meaning, interpretation of emotion, understanding, recall, and the ability to follow instructions) as well as an overall listening score.

Development

The WBLT was first introduced by Kittie Watson and Larry Barker (1983) with the goal of creating a standardized listening measure for adults and college students. Watson *et al.* (2001) claimed that the test focuses on the types of listening that adults may face in professional settings. The five parts of the test were developed in recognition that different contexts require different types of listening comprehension ability. Over time, the test has been revised, with Forms E and F being the most current. As noted, each version of the test comes in two forms for test–retest purposes. Early versions (Forms A and B) were presented via audiotape, whereas later versions (Forms C and D and Forms E and F) are presented by video.

Reliability

Watson and Barker (1988) reported that Form A of the test was administered to "several thousand subjects" whose scores were "subjected to factor analyses, item analyses, reliability tests, and descriptive analyses" (p. 25). It is important to note that although mean scores were reported, full descriptive statistics were not. For example, the developers reported that the number of business and professional subjects was "3,000 plus" (Watson & Barker, 1988, p. 27). The descriptive statistics they provided are based on a combination of responses to Forms A and B. Watson and Barker stated that results of early versions of the WBLT were subjected to item analysis, but the nature and results were not fully discussed. They reported alternative form reliability coefficients between Forms A and B using the Kuder-Richardson reliability formula at $r = .42$, suggesting less than 20% shared variability.

Reliability estimates for WBLT data have varied widely (e.g., .14–.73) (see, e.g., Villaume & Brown, 1999; Worthington *et al.*, 2009), and reports of factor analyses by Fitch-Hauser and Hughes (1986) and Villaume and Weaver (1996) failed to confirm the test's structure. Later studies testing newer versions of the WBLT (e.g., Form C)

(Bodie, Worthington, & Fitch-Hauser, 2011; Johnson & Long, 2007; Worthington, Keaton, Fitch-Hauser, Cook, & Powers, 2014) also questioned the potential reliability as well as the dimensional structure of the instrument. For example, Bodie *et al.* (2011) tested the underlying dimensions as outlined by Watson and Barker using confirmatory factor analytic procedures. They found that the WBLT did not conform to any of their three theoretically derived models—five interrelated factors, a second-order factor model, and a unidimensional factor structure. A follow-up exploratory factor analysis (maximum likelihood, varimax rotation) provided further support for claims that Form C of the WBLT lacked internal consistency. In particular, the average inter-item correlation between test items was .03. In other words, the test consists of 40 uncorrelated items.

Validity

DeVellis (2003) noted that fundamental to the construction and validity assessment of any measure is that scale items be at least moderately correlated. As seen in this profile, this minimum requirement has not been met by the WBLT. Although recent assessments of the WBLT's reliability and validity have been performed with Form C of the instrument, it is doubtful that other study results utilizing other earlier versions of the measure perform much better. Watson and Barker released Forms C and D as refined versions of Forms A and B of the test. Forms A and B of the WBLT were 10 items longer and presented in an audio format. Much of the early research utilizing Forms A and B involved correlational assessments, instead of more appropriate confirmatory factor analytic techniques. In addition, this research typically compared the WBLT to other early measures of listening comprehension, which are themselves questionable (e.g., Kentucky Comprehensive Listening Test, STEP II). For example, Applegate and Campbell (1985) tested the relation between the WBLT and the Kentucky Comprehensive Listening Test (KCLT). They reported that scores between the two tests were correlated, but also noted that their results suggested that neither the WBLT nor the KCLT accounted fully for individual listening comprehension (see also Villaume & Weaver, 1996).

Availability

Forms E and F of the WBLT are available on DVD from Innolect, Inc. (www.innolectinc.com). A facilitator guide and self-scoring answer sheets accompany the purchase of the WBLT. Additional self-scoring answer sheets can be purchased.

Sample Studies

Despite questions of the reliability and validity of the WBLT, it has been used frequently in educational and business contexts and in academic studies (see, e.g., Applegate & Campbell, 1985; Bommelje, Houston, & Smither, 2003; Clark, 1989; Fitch-Hauser, Powers, O'Brien, & Hanson, 2007; Roach & Fitch-Hauser, 1984; Vierthaler & Barker, 1985; Villaume & Brown, 1999; Watson & Rhodes, 1988; Worthington *et al.*, 2014).

For instance, early research examined the relation between the WBLT and the Communication Competency Assessment Instrument (CCAI; Profile 10) and between the WBLT and the Receiver Apprehension Test (Profile 24). However, as seen here, the lack of internal consistency of the measure as well as the question of the stability of the factor structure make any such research questionable. In fact, researchers often present their findings with caveats (e.g., Bommelje *et al.*, 2003; Worthington *et al.*, 2014).

Critique

Based on the problematic reliability issues described here and the lack of a viable validity portfolio for the WBLT, its use as a research tool is suspect. Bodie *et al.* (2011) offered several reasons for the low correlations across many of the WBLT items, including dichotomous scoring, which may not reflect the complexity of the listening process, and conflating supposedly distinct attributes of the listening process. They noted, as have others (Applegate & Campbell, 1985; Fitch-Hauser & Hughes, 1992), that the WBLT does not fully assess listening comprehension. Treating a complex, multidimensional communication process such as listening comprehension as a single construct is problematic. Even today, listening scholars continue to disagree on what subskills make up listening comprehension (Bostrom, 2011). Based on this profile, researchers, educators, and trainers are advised against using the Watson-Barker Listening Test.

References

Applegate, J. S., & Campbell, J. K. (1985). A correlation analysis of overall and sub-test scores between the Watson-Barker and the Kentucky comprehensive listening tests. Paper presented at the meeting of the International Listening Association: Orlando, FL.

Bodie, G. D., Worthington, D. L., & Fitch-Hauser, M. (2011). A comparison of four measurement models for the Watson-Barker listening test (WBLT)–Form C. *Communication Research Reports*, *28*, 32–42. doi:10.1080/08824096.2011.540547

Bommelje, R., Houston, J. M., & Smither, R. (2003). Personality characteristics of effective listeners: A five factor perspective. *International Journal of Listening*, *17*, 32–46. doi:10.1080/10904018.2003.10499054

Bostrom, R. N. (2011). Rethinking conceptual approaches to the study of "listening." *International Journal of Listening*, *25*, 10–26. doi:10.1080/10904018.2011.536467

Clark, A. J. (1989). Communication confidence and listening competence: An investigation of the relationships of willingness to communicate, communication apprehension, and receiver apprehension to comprehension of content and emotional meaning in spoken messages. *Communication Education*, *38*, 237–249. doi:10.1080/03634528909378760

Fitch-Hauser, M., & Hughes, M. A. (1986). A factor analytic study of four listening tests. Paper presented at the meeting of the International Listening Association, San Diego, CA.

Fitch-Hauser, M., & Hughes, M. A. (1992). The conceptualization and measurement of listening. *International Journal of Listening*, *6*, 6–22. doi:10.1080/10904018.1992.10499105

Fitch-Hauser, M., Powers, W. G., O'Brien, K., & Hanson, S. (2007). Extending the conceptualization of listening fidelity. *International Journal of Listening*, *21*, 81–91. doi:10.1080/10904010701301982

Roach, D., & Fitch-Hauser, M. (1984). *A comparison of three modes of administering listening tests*. Paper presented at the meeting of the International Listening Association, Scottsdale, AZ.

Vierthaler, D., & Barker, L. (1985). The presence of others: An investigation of its effect on listening test scores. *World Communication*, *14*, 95–101.

Villaume, W. A., & Brown, M. H. (1999). The development and validation of the Vocalic Sensitivity Test. *International Journal of Listening*, *13*, 24–45. doi:10.1080/10904018.1999.10499025

Villaume, W. A., & Weaver, III, J. B. (1996). A factorial approach to establishing reliable listening measures from the WBLT and the DCLT: Full information factor analysis of dichotomous data. *International Journal of Listening*, *10*, 1–20. doi:10.1207/s1932586xijl1001_1

Watson, K. W., & Barker, L. L. (1983). *Watson-Barker listening test*. Tega Cay, SC: SPECTRA.

Watson, K. W., & Barker, L. L. (1988). Listening assessment: The Watson-Barker Listening Test. *International Journal of Listening*, *2*, 20–31. doi:10.1080/10904018.1988.10499095

Watson, K. W., Barker, L. L, Roberts, C. V., & Roberts, J. D. (2001). *Watson-Barker Listening Test: Video version: Facilitator's guide*. Sautee, GA: SPECTRA.

Watson, K. W., & Rhodes, S. C. (1988). *A preliminary study of the effects of gender and a videotaped instructional strategy on listening effectiveness as measured by the Watson-Barker listening test*. Paper presented at the meeting of the International Listening Association, Scottsdale, AZ.

Worthington, D. L., Fitch-Hauser, M., Cook, J., & Powers, W. G. (2009, November). Cultural differences in listening comprehension: An investigation of the Watson–Barker Listening Test. Paper presented at the meeting of the National Communication Association, Chicago, IL.

Worthington, D. L., Keaton, S., Cook, J., Fitch-Hauser, M., &. Powers, W. G. (2014). A comparative analysis of Hispanic and Anglo-American listening comprehension ability as measured by the Watson-Barker listening test. *Journal of Research in Education*, *24*, 192–203. Retrieved from http://www.eeraorganization.org/#!journal/cln9

Further Reading

Aryadoust, V., & Goh, C. (2011). Exploring the relative merits of cognitive diagnostic models and confirmatory factor analysis for assessing listening comprehension. In E. D. Galaczi & C. J. Weir (Eds.), *Studies in language testing: Proceedings from the ALTE Krakow conference, 2011*. Cambridge: Cambridge University Press.

Roberts, C. V. (1985). Preliminary research employing the Watson-Barker listening test: A validation of the Instrument. Paper presented at the meeting of the International Listening Association, Orlando, FL.

Rubin, R. B., & Roberts, C. V. (1987). A comparative examination and analysis of three listening tests. *Communication Education*, *36*, 142–153. doi:10.1080/03634528709378655

Profile 65

Willingness to Listen (WTL)

(Richmond & Hickson, 2001; Roberts & Vinson, 1998)

Profiled by: Andrea J. Vickery, PhD

University of Richmond

Construct

Willingness to Listen (WTL) refers to individual differences in motivation to listen in various relational and situational contexts.

Instrument Type

Self-Report

Description

There are two WTL measures. The first instrument was developed by Roberts and Vinson (1998), and the second instrument was developed by Richmond and Hickson (2001). Both WTL instruments measure how people feel about listening in various social contexts. The Roberts and Vinson (1998) instrument varies these contexts based on the type of relationship between the message sender and listener (e.g., stranger, acquaintance, or friend), physical location (e.g., school or work), type of communicative context (e.g., sales presentation, group discussion, or formal presentation), and communication channel (e.g., face-to-face or via telephone). The statements included in the Roberts and Vinson (1998) WTL instrument include combinations of these contexts (e.g., a sales presentation from a friend or a sales presentation from a stranger).

The second WTL instrument was developed by Richmond and Hickson (2001) and also measures how people feel about listening; its primary focus, however, is listening to speeches or presentations. Respondents are asked to describe their willingness to listen

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

based on speaker characteristics (e.g., the speaker is boring, nonimmediate, not clear, or not credible), speaker content (e.g., if the content is boring and disorganized), and distractions (e.g., if there is background noise during a presentation).

Administration

The WTL developed by Roberts and Vinson (1989) is a self-administered questionnaire that takes approximately 10–15 minutes to complete. Participants are asked to provide an estimate of the percentage of time they would choose to listen in that particular situation on a rating scale ranging from 0 (*never*) to 100 (*always*).

The WTL instrument developed by Richmond and Hickson (2001) is a self-administered questionnaire that takes approximately 10 minutes to complete. Participants are asked to indicate their typical level of agreement (5-point, Likert) with statements about feelings toward particular listening situations.

Scoring

Roberts and Vinson (1989) do not specify how to score their WTL instrument. Their version of the WTL involves data that approximate interval-level scaling, and all 36 items are summed and then averaged to generate a mean score on the 0–100 scaling.

Values on the Richmond and Hickson (2001) WTL are summed together following a three-step formula. The values for Step 1 are generated by summing the response values on items 2, 4, 6, 8, 9, 12, 14, 16, 17, 19, 21, and 23. The values for Step 2 are generated by summing the response values on items 1, 3, 5, 7, 10, 11, 13, 15, 18, 20, 22, and 24. After these subtotals are generated, Step 3 is to employ the following formula: 64 – Step 1 + Step 2. Richmond and Hickson (2001) reported individual scores should range from 24, indicating low willingness to listen, to 120, indicating high willingness to listen, although no normative data have been published.

Development

Charles Roberts and Larry Vinson (1989) developed their WTL instrument in order to capture individual variation in effective listening. The first empirical test provided evidence that willingness to listen differs from willingness to communicate based on differing roles between senders and receivers in communication contexts (Roberts & Vinson, 1989). Roberts and Vinson (1998) suggested individuals may be motivated to listen at habitual levels of listening based on situational factors instead of listening at more effective or optimal levels. Roberts and Vinson (1998) compared listening performance to physical performance, arguing that "just as world-class runners can choose to 'go full out,' or hold back, depending on the situation, so too can listeners 'listen well,' or simply pay attention" (p. 42).

Virginia Richmond and Mark Hickson published their willingness to listen instrument in 2001 in a public speaking textbook. Their WTL instrument is located in their chapter on listening and serves as an assessment activity for readers learning about listening. Within this chapter, readers are given practical instructions based on the differing roles of speaker or listener. Listeners are instructed to be courteous toward speakers, reduce distractions, and listen "as if there will be a test" (Richmond & Hickson, 2001, p. 101). Speakers are reminded of the selectivity of listeners and are instructed to make information clear for listeners.

Reliability

Roberts and Vinson (1989, 1998) reported a reliability estimate of $\alpha = .88$ for the first WTL instrument, generated from 17 of the 36 items. The 19 deleted items were items 1–5, 9–11, 15–19, 23, 26, and 28–31 (all items are presented at the end of this profile). Hayhurst (2002) reported a reliability estimate of $\alpha = .95$ for the 36-item instrument. Vickery and Worthington (2014) reported a reliability estimate of $\alpha = .95$ for the 36-item instrument.

Reliability estimates for the second WTL instrument were not reported in Richmond and Hickson (2001). Hayhurst (2002) reported a reliability estimate of $\alpha = .79$ for the 24-item instrument. Vickery and Worthington (2014) reported a reliability estimate of $\alpha = .86$ for the 24-item instrument.

Validity

Roberts and Vinson (1998) stated a factor analysis was performed on the original WTL instrument and that the results of this factor analysis prompted revisions to their original WTL instrument, including the deletion of items about the topic of communication (Roberts & Vinson, 1989). Unfortunately, the statistical results of this factor analysis were not reported. Vickery and Worthington (2014) reported the measurement model fit for a one-factor unidimensional model, but model fit statistics were below conventional thresholds, $\chi2$ (594) = 1644.97, $p < .001$, CFI = .64, RMSEA = .13 (.12; .14), SRMR = .10.

Richmond and Hickson (2001) did not report any validity evidence for their WTL instrument. Vickery and Worthington (2014) reported the measurement model fit for a one-factor unidimensional model, but model fit statistics were below conventional thresholds, $\chi2$ (252) = 896.33, $p < .001$, CFI = .37, RMSEA = .16 (.15; .17), SRMR = .14.

Availability

Roberts and Vinson (1998) published their WTL instrument in the *International Journal of Listening*. All 36 items are reported here, with permission. This scale is free to use for research purposes with appropriate citation. Richmond and Hickson (2001) presented their WTL instrument in their textbook, *Going Public: A Practical Guide to Public Talk*. All 24 items are reported here, with permission. This scale is free to use for research purposes with appropriate citation.

Sample Studies

WTL has primarily been studied in relation to other individual differences in personality, listening, and communication. Roberts and Vinson (1989) found a negative correlation between WTL and communication apprehension ($r = -.21$). WTL also has been found to negatively correlate with receiver apprehension, dogmatism, and communication competence (Roberts & Vinson, 1998). Hayhurst (2002) investigated the relation between WTL and extraversion, neuroticism, and psychoticism.

Hayhurst (2002) further found the two WTL instruments (R&V and R&H) were negatively correlated, $r = -.33$. Within these correlation analyses, both WTL instruments demonstrated different patterns: For example, the Roberts and Vinson (1998) instrument was positively associated with extraversion, $r = .22$; whereas the Richmond and Hickson (2001) instrument was not associated with extraversion, $r = -.06$, $p = .34$ (Hayhurst, 2002). Vickery and Worthington (2014) found the two WTL instruments were not statistically associated, $r = -.03$, $p = .78$.

Critique

The present empirical evidence suggests that willingness to listen differs from other individual differences in listening such as receiver apprehension (Hayhurst, 2002; Roberts & Vinson, 1989, 1998). Based on the present findings, further evidence for reliability and validity of both versions of the WTL is needed. Researchers are advised against using either scale for research or training purposes until further evidence of validity is offered.

In particular, both WTL instruments are presumed to be unidimensional. If so, each contains many more items than necessary. Examining the content of the items excluded from Roberts and Vinson's (1998) reliability analysis, there may be an unobserved factor representing the type of relationship shared between speaker and listener. There were 12 items representing listening to strangers; 11 of these items were excluded by Roberts and Vinson (1998) in order to achieve higher reliability. Similarly, there are five questions on the Richmond and Hickson (2001) instrument that reference how boring a speaker is perceived to be; it may be possible that these items represent a subscale within their WTL instrument. Once measurement models are initially supported with independent data, these models should be tested in various populations. For the Richmond and Hickson instrument in particular, validity should be investigated with samples gathered from populations other than primarily college students. Many of the items focus on willingness to listen when receiving information from speeches or presentations. Although this is highly appropriate if the focus of research is on public speaking classes, other settings for presentations like the boardroom might not resemble the classroom in meaningful ways.

References

Hayhurst, J. L. (2002). *The big three as related to receiver apprehension and listening behaviors*. Unpublished master's thesis. West Virginia University, Morgantown, WV.

Richmond, V. P., & Hickson, M., III. (2001). *Going public: A practical guide to public talk*. Boston: Allyn & Bacon.

Roberts, C. V., & Vinson, L. (1989). *Willingness to listen: A multifaceted process*. Paper presented at the Speech Communication Association Convention, San Francisco, CA.

Roberts, C. V., & Vinson, L. (1998). Relationship among willingness to listen, receiver apprehension, communication apprehension, communication competence, and dogmatism. *International Journal of Listening, 12*, 40–56. doi:10.1080/10904018.1998.10499018

Vickery, A. J., & Worthington, D. L. (2014, November). *Willingness to Listen*. Paper presented at the meeting of the International Listening Association, Minneapolis, MN.

Scale

Willingness to Listen (Roberts & Vinson, 1998)

Roberts and Vinson (1998) labeled the instrument "Form 1822" for participants.

Directions: Below are some situations in which a person might choose to listen or not to listen. Presume you have completely free choice. What percentage of the time would you choose to listen in each type of situation? Indicate in the space on the left what percent of the time you would choose to listen, ranging from 0 = never, to 100 = always. You can stipulate any percentage between 0% and 100%.

___ 1. Listen to a friend give a presentation at a school function.
___ 2. Listen to a stranger give a sales presentation to you in your home.
___ 3. Listen to a stranger participate in a group discussion on TV.
___ 4. Listen to a stranger participate in a group discussion during lunch.
___ 5. Listen to an acquaintance give a presentation at a business meeting.
___ 6. Listen to a friend participate in a group discussion during a business meeting.[1]
___ 7. Listen to a friend talk to you before a meeting begins at a school.[1]
___ 8. Listen to an acquaintance participate in a group discussion on TV.[1]
___ 9. Listen to a stranger give a formal presentation on TV.
___ 10. Listen to a stranger talk to you before a meeting begins at a school.
___ 11. Listen to a stranger talk to you before a meeting begins at work.
___ 12. Listen to an acquaintance give a presentation at a school function.[1]
___ 13. Listen to a friend participate in a group discussion during a school meeting.[1]
___ 14. Listen to an acquaintance on the phone.[1]
___ 15. Listen to a stranger on a plane.
___ 16. Listen to a stranger give a presentation at a business meeting.
___ 17. Listen to an acquaintance participate in a discussion during a business meeting.
___ 18. Listen to an acquaintance on a plane.
___ 19. Listen to a stranger on the phone.
___ 20. Listen to a friend give a sales presentation to you in your home.[1]
___ 21. Listen to a friend participate in a group discussion on TV.[1]
___ 22. Listen to an acquaintance participate in a group discussion during lunch.[1]
___ 23. Listen to a stranger give a presentation at a school function.
___ 24. Listen to an acquaintance participate in a discussion during a school meeting.[1]
___ 25. Listen to an acquaintance talk to you before a meeting begins at work.[1]
___ 26. Listen to a friend on a plane.
___ 27. Listen to a friend give a formal presentation on TV.[1]
___ 28. Listen to a stranger participate in a discussion during a business meeting.
___ 29. Listen to an acquaintance talk to you before a meeting begins at a school.
___ 30. Listen to a friend participate in a group discussion during lunch.
___ 31. Listen to an acquaintance give a sales presentation to you in your home.
___ 32. Listen to a friend on the phone.[1]
___ 33. Listen to a friend talk to you before a meeting begins at work.[1]
___ 34. Listen to a friend give a presentation at a business meeting.[1]
___ 35. Listen to a stranger participate in a group discussion during a school meeting.[1]
___ 36. Listen to an acquaintance give a formal presentation on TV.[1]

1 Items included in the reliability estimate provided in Roberts and Vinson (1989, 1998).

Willingness to Listen (Richmond & Hickson, 2001)

Directions: The following 24 statements refer to the willingness to listen. Indicate in the space at the left of each item the degree to which the statement applies to you.

1 = Strongly Agree
2 = Agree
3 = Undecided
4 = Disagree
5 = Strongly Disagree

___ 1. I dislike listening to boring speakers.[2]
___ 2. Generally, I can listen to a boring speaker.[1]
___ 3. I am bored and tired while listening to a boring speaker.[2]
___ 4. I will listen when the content of a speech is boring.[1]
___ 5. Listening to boring speakers about boring content makes me tired, sleepy, and bored.[2]
___ 6. I am willing to listen to boring speakers about boring content.[1]
___ 7. Generally, I am unwilling to listen when there is noise during a speaker's presentation.[2]
___ 8. Usually, I am willing to listen when there is noise during a speaker's presentation.[1]
___ 9. I am accepting and willing to listen to speakers who do not adapt to me.[1]
___10. I am unwilling to listen to speakers who do not do some adaptation to me.[2]
___11. Being preoccupied with other things makes me less willing to listen to a speaker.[2]
___12. I am willing to listen to a speaker even if I have other things on my mind.[1]
___13. While being occupied with other things on my mind, I am unwilling to listen to a speaker.[2]
___14. I have a willingness to listen to a speaker, even if other important things are on my mind.[1]
___15. Generally, I will not listen to a speaker who is disorganized.[2]
___16. Generally, I will try to listen to a speaker who is disorganized.[1]
___17. While listening to a non-immediate, non-responsive speaker, I feel relaxed with the speaker.[1]
___18. While listening to a non-immediate, non-response speaker, I feel distant and cold toward that speaker.[2]
___19. I can listen to a non-immediate, non-responsive speaker.[1]
___20. I am unwilling to listen to a non-immediate, non-responsive speaker.[2]
___21. I am willing to listen to a speaker with different views from mine.[1]
___22. I am unwilling to listen to a speaker with views different from mine.[2]
___23. I am willing to listen to a speaker who is not clear about what he or she wants to say.[1]
___24. I am unwilling to listen to a speaker who is not clear, not credible, and abstract.[2]

---

1 Items to be summed in Step 1 in scoring.
2 Items to be summed in Step 2 in scoring.

Scoring

Your score can range from 24 to 120.

Step 1: Add scores for items 2, 4, 6, 8, 9, 12, 14, 16, 17, 19, 21, and 23.
Step 2: Add scores for items 1, 3, 5, 7, 10, 11, 13, 15, 18, 20, 22, and 24.
Step 3: Score = 64 – Total from Step 1 + Total from Step 2.

Index

Page references to Figures or Tables will be in **bold**, while references to Notes will be followed by the letter 'n' and Note number.

a

Academic Listening Self-Rating Questionnaire (ALSQ) 105, 153–160
ACC *see* Audio Content Change (Acc)/ Audio Information Introduced (Aii)
access, gaining 55
ACI *see* Affection Communication Index (ACI)
ACS *see* Attributional Complexity Scale (ACS)
action-oriented listening 252
active listening (AL) 106, 136, **137**, 168
Active Listening Attitude Scale (ALAS) 167–173
Active Listening Observation Scale (ALOS) 174–179, 274, 487
active observation 56
active-empathic listening (AEL) 108, 109, 568
Active-Empathic Listening Scale (AELS) 24, 29, 112, 161–166, 194, 322, 487
adaptive selling technique 194
administration of listening measurement instruments
 see individual measurement profiles
advice questions 61
AELS *see* Active-Empathic Listening Scale (AELS)
Affection Exchange Theory 11
affectionate communication 107, 180, 181, 182
 see also Affectionate Communication Index (ACI)
Affectionate Communication Index (ACI) 11, 111, 186–190
Affectionate Communication Scale *see* Trait Affection Scale (TAS)
affective components of listening 4–8, 12, 70, 97–122, 276
 choice of reporting method 116
 daily sampling of listening experiences 115–117
 diary study 113–115
 event-contingent reporting methods 116–117
 length of reporting instrument 117
 method of reporting 117
 multitrait-multimethod (MTMM) matrix 108, **109**, 110
 self-reporting *see* self-reporting
 situationally derived listening goals, measuring 113
Aii *see* Audio Content Change (Acc)/Audio Information Introduced (Aii)
ALAS *see* Active Listening Attitude Scale (ALAS)
ALOS *see* Active Listening Observation Scale (ALOS)
ALSQ *see* Academic Listening Self-Rating Questionnaire (ALSQ)

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

ambient sound change 200
American Psychological Association (APA) 35, 194
  Committee on Psychological Tests 28
  PsycTESTS searchable electronic database 254
analysis of variance tests 45
analytical listening 403
Andersen, J. F. 140, 483, 484, 485–486, 487
Andersen, P. A. 140, 487
APA *see* American Psychological Association (APA)
appropriate disclosure 204–205
Aryadoust, V. 83, 90, 153–156
attention
  attentional focus 469
  attentional switching 460
  conceptualization 75
  derailments of 89
  emphasizing, in listening research 73
  operationalization 75–77
  selective 75
attentiveness 85, 319, 320
attribute independent variable 23
Attributional Complexity Scale (ACS) 111, 191–197
auding 71, 73
Audio Content Change (Acc)/Audio Information Introduced (Aii) 198–203
auditory communication, receiving 71
Aune, R. K. 497, 498, 500
Automated Working Memory Assessment (AWMA) 396, 397, 398
Avtgis, T. A. 412, 413

b

Barker, L. L. 32, 113, 114, 123, 124, 134, 240, 242, 243, 579, 581–583, 585, 613, 614
Basic Communication Fidelity (BCF) test 379
Bavelas, J. B. 131, 141, 448, 449
Beatty, M. J. 206, 207, 315, 601
Beck Depression Inventory 194
behavioral components of listening 5, 8–9, 12, 70, 123–150
  analysis of behavioral data 142–145
  coding and rating of behaviors 140–142
  competences associated with listening, listed **125–126**
  defining behavior 127–128
  design of a behavioral listening study 128–134
    natural and laboratory observation 128–131
    recording equipment, using to capture behavior for later analysis 132–133
    representing behavior that is captured 133–134
  future trends in behavioral listening research 145–146
  judging behavior 139–140
  level of analysis 137
  listening as overt behavior 124, 126–127
  listening behavior as an internal act 124
  naturalistic settings 130
  perspectives of behavior 138–139
  reliability of rating and coding 141–142
  representing behavior 137–140
  sampling behavior 136–137
  thinking of listening behavior in concrete or abstract terms 134–142
  training of coders/raters 140–141
  types of questions for behavioral listening research 143–145
Behavioral Indicants of Immediacy Scale (BIIS) 482–488, 491–492
  see also nonverbal immediacy (NVI) measures
Behavioral Inhibition System (BIS) 301
BIIS (Behavioral Indicants of Immediacy Scale) 482–488, 491–492
blood oxygenation level-dependent (BOLD) contrast 292, 293
Bodie, G. D. 7–8, 24, 25, 33, 35–38, 40, 87, 90, 97, 104, 105, 108, 113, **135**, 136, 162–165, 175–177, 234, 235, 301, 302, 322, 373–376, 402–403, 405, 406–407, 487, 498, 500, 563, 568, 601, 614, 615
body language 134
Bostrom, R. N. 6, 10, 81, 86, 87
Bouskila-Yam, O. 272, 273, 274, 275
brain imaging 83, 292
  see also functional magnetic resonance imaging (fMRI)

brain stem 75
Brandt, J. 388, 389
bricoleurs 53
Brief Job Stress Questionnaire (BJSQ) 169
Brilhart, B. L. 380
Broadbent, D. E. 460, 500
Brown, J. I. 21, 81
Brown-Carlsen Listening Comprehension Test (BCLCT) 9, 77
Brownell, J. 73, **74**, 75, 76, 77, 127, 307, 309
  Listening: Attitudes, Principles and Skills 308

C

Cacioppo, J. T. 77, 499
Campbell, D. T. 30, 108
Cappella, J. N. 143, 151
Carbaugh, D. 47, 49, 50, 52, 53, 62
CAS *see* Communicative Adaptability Scale (CAS)
Cegala, D. J. 319, 320, 321, 322, 323, 412
CFA *see* confirmatory factor analysis (CFA)
CFQ-30 *see* Communication Functions Questionnaire (CFQ-30)
CHE *see* Couples Helping Exercise (CHE)
classification independent variable 23
classification of listening measurement instruments
  see individual measurement profiles
Client-Focused Evaluations Programme (CFEP) 246n1, 248
Cliffordson, C. 342
CLS *see* Conversational Listening Span (CLS)
Coakley, C. G. 113, 560, 561
coding
  coder drift 142
  ethnography 64–65
  and rating of behaviors 140–142
  reliability of rating and coding 141–142
  training of coders/raters 140–141
Cognitive and Affective Mindfulness Scale (CAMS) 282
cognitive components of listening 5–6, 11–12, 276
  see also models of listening
  attention 75–77
  cognitive processing skills 83, 153
  comprehension 81–83
  lecture structure 83, 153
  linguistic 83
  memory 77–81
  new developments 87–90
cognitive dissonance 47
cognitive processing skills (CPSs) 83, 153, 154, 155
Cognitive-Experiential Self-Theory (CEST) 530, 531, 532
Cohen's kappa 176
cohort studies 105
Comer, L. B. 164
comforting skills 217
common method variance (CMV) 103, 110, 111
communication, ethnography of 52–53
Communication Competency Assessment Instrument (CCAI) 24, 26, 212–216
Communication Functions Questionnaire (CFQ-30) 112, 217–223
Communication Quarterly 206
Communicative Adaptability Scale (CAS) 204–211
competency listening models 8
complete observer 57
complete participant 56
comprehension
  conceptualization 81–82
  functional magnetic resonance imaging (fMRI) 294–295
  knowledge types 83
  measuring 25–26
  operationalization 82–83
  structure building 81
comprehension of aural information 5–6
compulsive communication (CC) 566, 567, 568, 569
  see also Talkaholic Scale (Compulsive Communication Scale) (TAS)
computer-assisted qualitative data analysis software (CAQDAS) 64, 65
conceptual development 64
conceptualization
  attention 75
  comprehension 81–82
  memory 78–79
concurrent validity 29

confessional tales, ethnography 65
confidence intervals, reporting 36–37
confirmatory factor analysis (CFA) 30, 90, 108, 163, 206, 322, 342
  see also factor analyses (FAs)
conflict management skills 217
consensus-oriented listening 252, 254
construct validity 29–31, 103, 412, 448
  generating strong evidence for 107–110
Constructive Thinking Inventory (CTI) 532
constructivism 7, 11
constructs 21–22
  see also specific measurement instruments
  construct differentiation 553
  construct proliferation 104
  construct validity *see* construct validity
  defining 552
  theory of mind 329
content validity 29, 156
contrast questions 61
contrast-to-noise (CNR) ratio 293
convergent validity 30, 163
conversation type 58–59
Conversational Listening Span (CLS) 79–80, 224–231, 396, 419
conversational memory 79, 101, 419–429
  see also memory
Conversational Sensitivity (CS) 76, 85, 107, 232–235
Conversational Sensitivity Scale (CSS) 232–238
Cooper, L. O. 139, 509–510, 511, 512–513
Counselor Verbal Response Scale (CVRS) 336
Couples Helping Exercise (CHE) 34, 115, 130, 239–243
covariance 35
criterion sample 56
criterion-oriented validity 28–29, 248
critique of listening measurement instruments *see* evaluation of listening measurement instruments
Cronbach, L. 28–30
cross-sectional study
  making greater use of 105–106
  survey methodology 54
Crowne, D. 498
Csikszentmihalyi, M. 116
CSS *see* Conversational Sensitivity Scale (CSS)
cues/cued recall 79, 136, 421, 422

d

Daly, J. A. 76, 232–236, 420–422
data
  analysis 50–51
  behavioral 127, 142–145
  coding and analyzing 64–65
  collection 45, 49, 50, 51, 53
  dyadic 145
  examination of 35–36
  interpretation 51
  management of 64
  missing 35
  observational, documentation of 62–63
  ranked 26
  reduction of 64
declarative memory 78
definitional operationalism 107
definitions of listening *see* listening definitions
dependent variables (DVs) 23, 39, 106, 181, 242, 467
  Audio Content Change (Acc)/Audio Information Introduced (Aii) 201, 202
descriptive questions 61
descriptive statistics 34
detecting power 85
Deutsch, K. W. 71, 75, 84
Diagnostic Analysis of Nonverbal Accuracy (DANVA) 525, 527
dialogue 60, 447–448
  see also microanalysis of face-to-face dialogue (MFD)
  behavior 126, 127
  face-to-face 445, 446, 450
  naturalistic 131
diary study
  description 113–114
  reasons for conducting 114–115
dichotic listening task (DLT) 76, 87
differential item functioning (DIF) 157, 597
discriminant validity 30–31, 107, 108
DISQ *see* Doctors' Interpersonal Skills Questionnaire (DISQ)

divergent validity 30, 176, 227, 275, 276, 283, 498, 512, 532
Doctors' Interpersonal Skills Questionnaire (DISQ) 99, 246–251
Drollinger, T. 161, 163, 164
dual coding theory 460
dyadic data 145
dyadic language coordination (DLC) 348, 351

e

ear advantage (EA) 76
Educational Testing Service (ETS) 156, 592, 594
EFA *see* exploratory factor analysis (EFA)
effect sizes, reporting 36–37
Effective Listening and Interactive Communication Scale (ELICS) 252–258
ego support skills 217
ELAN software 446
electroencephalogram (EEG) 76, 295
electronic medical records (EMRs) 45, 48, 49, 51, 54, 55, 58, 60, 63, 64
ELICS *see* Effective Listening and Interactive Communication Scale (ELICS)
embedded randomized block design 86
empathic accuracy 261
Empathic Accuracy-Standard Stimulus Paradigm (EA-SSP) 131, 259–264
Empathic Accuracy-Unstructured Dyadic Interaction Paradigm (EA-UDIP) 131, 260, 262, 265–271
Empathic Understanding in Interpersonal Processes Scale (EUS) 336
empathy 107, 124, 194
  see also Empathic Accuracy-Standard Stimulus Paradigm (EA-SSP); Empathic Accuracy-Unstructured Dyadic Interaction Paradigm (EA-UDIP)
  defining 340
  empathic listening 106
  levels of 169, 469
  measures of 113, 162, 163, 341, 342
  in sales 29, 164
episodic memory 78
error
  measurement 107
  random 27, 103
  systematic 103
  Type I 111
ethnographic interviews 60, 62
ethnography 11, 45
  see also qualitative research
  access, gaining 55
  behavioral perspective 128
  coding and analyzing data 64–65
  content of study 54–55
  critical 53
  deciding when to leave the field 63–65
  documentation of observational data 62–63
  "doing," 53–66
  fieldwork 50, 54, 57–59, 105
  interviews 60–62
  matters to observe while in the field 57–59
  "outsider" status 54
  participant listening 59–60
  participant observation 56–57
  sampling strategies 55–56
  tales 65
  validity threats, addressing 63
  writing report 65–66
ethnography of communication (EOC) 52–53
Euclidean distances 412
  evaluation of listening measurement instruments
    see individual measurement profiles
event-contingent sampling 116
event-related potential (ERP) 77
explanation questions 61
exploratory factor analysis (EFA) 38
exploratory listening 252, 254
Expressive Writing (EW) 330, 331
external responsiveness 31
external validity 115, 118, 192, 243, 449, 554, 555
Eysenck, H. J. 568

f

facial cues 136
Facial Meaning Sensitivity Test (FMST) 499

Facilitated Listening Scale (FLS) 76, 272–280
Fassaert, T. 175–176, 177
feedback 124, 134
Feffer, M. 327–333
FFMQ *see* Five Facet Mindfulness Questionnaire (FFMQ)
Fitch-Hauser, M. E. 37, 84, 87, 381, 613
Five Facet Mindfulness Questionnaire (FFMQ) 281–289
FLS *see* Facilitated Listening Scale (FLS)
fMRI *see* functional magnetic resonance imaging (fMRI)
focus groups 61
format, measurement 25–26
free recall 79
Freiburg Mindfulness Inventory 282
functional magnetic resonance imaging (fMRI) 290–298
  see individual measurement profiles
  animal studies 291
  blocked designs 291
  cognitive components of listening 76–77, 80, 89
  event-related designs 291
  general linear model (GLM) 292
  multitasking, while listening 460
  naturalistic 294
  reliability 293
  resting-state fMRI (rsfMRI) 292
  scanner noise 293–294, 296
  temporal resolution 295
functional near-infrared spectroscopy (fNIRS) 295
fusion modeling (FM) 90

g

Gable, S. L. 116, 518, 580
Gearhart, C. C. 113, 163, 300, 301, 302, 406, 498
Generalized Immediacy Scale (GIS) 483, 484, 485, 493–495
  see also nonverbal immediacy (NVI) measures
GIS (Generalized Immediacy Scale) 483, 484, 485, 493–495
Goh, C. C. M. 83, 90, 153, 154, 431, 433
Gross, G. G. 80

h

health-related quality of life (HRQOL) 413
hearing 71, 73
High, A. C. 242, 243
Highly Sensitive Person Scale (HSPS) 299–305
HLP *see* HURIER Listening Profile (HLP)
HSPS *see* Highly Sensitive Person Scale (HSPS)
Human Speechome Project 145
HURIER Listening Profile (HLP) **74**, 75, 76, 80, 81, 306–312
Husband, R. L. 509–510, 511

i

Ickes, W. 263, 267–270, 350–351, 352
ID *see* Interpersonal Decentering (ID)
Ifert, D. E. 569
II *see* interaction involvement (II)
IIS *see* Interaction Involvement Scale (IIS)
ILA *see* International Listening Association (ILA)
imagined interactions (IIs) 164
Imhof, M. 24, 25, 84, 88, 207, 373–376, 406, 440, 581, 583
implicit memory 78
independent variables (IVs) 23, 39, 106
in-depth interviews 46, 50, 60, 61, 62
  see also interviews
individual differences in listening, reasons for 7
inferences
  conceptualization 85
  inference making 83–87
  inferential statistics 35
  operationalization 85–87
  population 31, 32
  process 32–34
inferential statistics 34, 35
informant interviews 60, 62
Information Processing Theory 441
Informational Reception Apprehension (IRA) 98
Informational Reception Apprehension Test (IRAT) 313–318
  see also affective components of listening
Institutional Review Board (IRB) 55, 99–100

intact classes or groups, avoiding use of 39–40
interaction involvement (II) 107, 206, 215, 319, 478, 547
Interaction Involvement Scale (IIS) 85, 319–326
internal validity 39, 63, 118, 448, 449, 461
International English Language Testing System (IELTS) 154
International Journal of Listening (*IJL*) 35, 36, 38, 97, 619
International Listening Association (ILA) 3, 5, 87
Interpersonal Cognitive Complexity (ICC) 552
Interpersonal Decentering (ID) 327–333
Interpersonal Process Recall (IPR) 334–339
Interpersonal Reactivity Index (IRI) 26, 111, 340–347
Interpersonal Skills Questionnaire (ISQ) 246–249
  see also Doctors' Interpersonal Skills Questionnaire (DISQ)
interviews 45, 53, 60–62
  in-depth 46, 50, 60, 61, 62
  ethnographic 60, 62
  informant 60, 62
  narrative 61
  questions, types 61
  recorded 62
  respondent 60–61
intraclass correlation coefficient (ICC) 293
intracranial recordings 77
Inventive Experimental Learning (IEL) 169–170
IPR *see* Interpersonal Process Recall (IPR)
IRAT *see* Informational Reception Apprehension Test (IRAT)
IRI *see* Interpersonal Reactivity Index (IRI)
item pool, generating 24–25
item response theory (IRT) 594

j

Janusik, L. A. 24, 25, 32, 77–79, 81, 82, 87, 226–228, 373–376, 396, 438–442, 581, 583–585
Job Content Questionnaire (JCQ) 169
Johnston, M. K. 76, 574, 575
Jones, S. M. 130, 164, 175, 176, 177, 483, 487

k

Keaton, S. A. 35–38, 40, 97, 315, 316, 406–407, 438–441, 563
Kelly, C. M. 6, 77
Kelly, G. A. 554
Kentucky Comprehensive Listening Test (KCLT) 6, 77, 396
Kentucky Inventory of Mindfulness Skills (KIMS) 282
Kluger, A. N. 272–275

l

LA *see* listening anxiety (LA)
Lakey, B. 506
Lang, A. 200, 201, 202
Language Style Matching (LSM) 134, 348–353
language use, progression 128
Latent Semantic Analysis (LSA) 350
Latent Semantic Similarity (LSS) 350, 351
LCI/LCI-R *see* Listening Concepts Inventory (LCI and LCI-R)
Leader-Member Exchange 7 Questionnaire (LMX-7) 31, 354–360
left hemisphere, brain 294–295
left-ear advantage (LEA) 76
LF *see* Listening Fidelity (LF)
Liden, R. C. 358
Likert scaling 26, 208, 233, 566
linearity, in research 145
  general linear model (GLM), in fMRI 292
  nonlinear nature of qualitative research 47, 48n2, 51, 64
linguistic components and prosody (LCP) 153, 154
Linguistic Inquiry and Word Count (LIWC) 349
linguistic knowledge 83
Listenability Style Guide (LSG) 23, 361–371
listening 3–17
  action-oriented 252
  active 106, 136, **137**, 168

affective components *see* affective components of listening
analytical 403
and auding 71, 73
auditory communication, receiving 71
behavioral components *see* behavioral components of listening
cognitive components *see* cognitive components of listening
consensus-oriented 252, 254
definitions *see* listening definitions
distinctions 3–9
empathic 106
exploratory 252, 254
individual differences, reasons for 7
as multidimensional construct 3–4
and multitasking *see* multitasking, while listening
positive attitude to 97
receptive 252, 254
relational 403
research *see* research of listening methodology/measurement
supportive 106, 138, 276
task-oriented 403
therapeutic 106
listening anxiety (LA) 102, 103, 105, 107, 108
listening concepts 7, 24, 107, 373, 375, 378, 440
Listening Concepts Inventory (LCI and LCI-R) 24, 25, 39, 372–378
listening definitions 3–9
behavioral perspective 124
conceptual 10
limitations of a single definition 11
revisiting "the" definition 10–11
sample of **4**
searching for "the" definition 9–10
Listening Fidelity (LF) 379–386
listening goals 107, 111, 403, 500, 568
situationally derived 113, 117
Listening Practices Feedback Report (LPFR) 387–393
Listening Span tests 224, 394–401, 419
Listening Styles Profile-Revised (LSP-R) 30, 104, 274, 276, 381, 402–409
longitudinal research 105, 118, 397
LPFR *see* Listening Practices Feedback Report (LPFR)
LSG *see* Listenability Style Guide (LSG)
LSM *see* Language Style Matching (LSM)

m

magneto encephalogram (MEG) 76, 77, 295
MALQ *see* Metacognitive Awareness Listening Questionnaire (MALQ)
Maryland university, US 32
MCCS *see* Medical Communication Competence Scale (MCCS)
meaning, assignment of 83–87
conceptualization 85
operationalization 85–87
measuring listening 11
see also profiles, measurement; research of listening methodology/measurement
constructs 21–22
descriptive statistics 34
effect sizes and confidence intervals, reporting 36–37
factor analytic techniques, correct utilization/reporting 38
format, determining 25–26
implications of study results, clarity as to 38–39
inferential statistics 35
initial measures, developing 24–26
intact classes or groups, avoiding use of 39–40
item pool, generating 24–25
matching sampling to population of interest 38
measurement format, determining 25–26
new techniques 8–9
number and types of items on instrument 22
numerical data, reporting 34–40
operationalization *see* operationalization
profiles *see* profiles, measurement
psychometrical validation of derived scores 37–38
recommendations 35–40
scale length, optimizing 25
self-reporting *see* self-reporting
validity portfolio, developing 27–34
videotaped presentations 6–7

Medical Communication Competence Scale (MCCS) 410–418
MEESS *see* Multidimensional Evaluation of Enacted Social Support (MEESS)
Meldrum, H. 51
memory 10, 77–81
  conceptualization 78–79
  constructive nature of 80
  conversational *see* conversational memory
  episodic 78
  explicit (declarative) 78
  factors affecting 80–81
  "false memories," 7
  implicit (procedural) 78
  operationalization 79–81
  retention rates 423
  short-term 5–6, 395, 396, 397
  working memory (WM) 78, 394, 395, 397
memory and concentration (MC) 153, 154, 155, 157
  see also conversational memory
  Memory for conversation
    see also conversational memory
mentalizing 83
message interpretation 7
metacognition 88, 191
Metacognitive Awareness Listening Questionnaire (MALQ) 82, 430–437, 438, 441
metacognitive listening strategies 82, 434, 438, 440–442
Metacognitive Listening Strategies Instrument (MLSI) 82, 438–444
MFD *see* microanalysis of face-to-face dialogue (MFD)
Mickelson, W. T. 561, 564
microanalysis of face-to-face dialogue (MFD) 22, 445–452
  see also dialogue
  dialogue, analysis of 447–448
  ELAN software 446
microscopic skills/functions 134, 135–136
Mindful Attention Awareness Scale (MAAS) 282
Mindfulness Questionnaire 282
mirroring 83, 84, 97
Mishima, N. 167, 168, 169
models of listening 70–75
  HURIER model (Brownell) 74, 75
  MASTER model (Mills) 73
  process models 70, 71
  Taylor, S. E. 71, **72**, 73
Morman, M. 187
motivation/motivations to listen 97, 112–113
MTMM *see* multitrait-multimethod (MTMM) matrix
Multidimensional Evaluation of Enacted Social Support (MEESS) 26, 29, 243, 453–457
multiple-choice questions 5, 6, 8–9, 24, 26, 65n5, 79, 81
multitasking, while listening 458–464
multitrait-multimethod (MTMM) matrix 30, 108, **109**, 110
multivoxel pattern analysis (MVPA) 292
music onset, in audio message 199

n

Narrative Believability Scale (NBS-12) 25, 465–468
Narrative Engagement Measure (NEM)
narrative understanding 469
National Communication Association (NCA) 3, 124, **125–126**
NBS-12 *see* Narrative Believability Scale (NBS-12)
NCA *see* National Communication Association (NCA)
Need for Cognition Scale (NCS) 496, 531
neocortex 75
Nichols, R. 7, 8, 70, 81
  "Listening Is a 10-Part Skill," 5 6
NIS (Nonverbal Immediacy Scale) 483
NMPS *see* Normative Message Processing Scale (NMPS)
no ear advantage (NoEA) 76
Noise Sensitivity Scale *see* Weinstein, N. D.; Weinstein Noise Sensitivity Scale (WNSS)
nomological network approach 276
  organizational design and performance 573, 575
  validity 30, 103

nonverbal immediacy (NVI) measures 22, 28, 136, 482–495
Nonverbal Immediacy Scale (NIS-S, NIS-O) 483
Nonverbal Sensitivity *see* Profile of Nonverbal Sensitivity (PONS)
Normative Message Processing Scale (NMPS) 496–503
note taking (NT) 153, 154, 155, 157
null hypothesis significance testing (NHST) 35, 36

o
observer-participant 57
Occam's razor 106
OCS *see* Ordinary Conversation Scale (OCS)
OLS *see* Organizational Listening Survey (OLS)
open coding 64
operationalization 22–24
  attention 75–77
  choosing an operationalization 23–24
  comprehension 82–83
  defining 22, 75
  memory 79–81
ordinal scale 26
Ordinary Conversation Scale (OCS) 504–508
Organizational Listening Survey (OLS) 139, 509–515
orienting response (OR) 198

p
panel studies 105
participant listening 53, 59–60
participant observation 45, 49, 52, 53, 56–57
passive observation 56
Patient Emotion Cue Test (PECT) 262
PCA *see* principal component analysis (PCA)
Pearson's coefficient 108
Perceived Partner Responsiveness Scale (PPRS) 112, 516–521
perceptiveness 85, 319, 320
Personal Report of Communication Apprehension (PRCA-24) 301
PONS *see* Profile of Nonverbal Sensitivity (PONS)
population inferences 31, 32
positron emission tomography (PET) 295
post-listening performance (PLP) 593
Powers, W. G. 380, 381, 383
PPRS *see* Perceived Partner Responsiveness Scale (PPRS)
predictive validity 29
primary visual cortex 292
principal component analysis (PCA) 38, 104, 175, 193, 206, 234, 253, 254, 273, 274, 314, 315, 320, 373, 432, 484, 485, 511, 532, 547, 567
procedural memory 78
process inferences 31, 32–34
process models 70, 71
processing 161
production effect onset 199
Profile of Nonverbal Sensitivity (PONS) 85, 522–529
Purdy, M. W 46, 48, 52

q
Quakers, listening practices among 50, 52
qualitative research 11, 45–69
  bias of investigators 51
  characteristics of qualitative inquiry 46–52
  defining 45
  electronic medical records (EMRs) 45, 48, 49, 51, 54, 55, 58, 60, 63, 64
  ethnography *see* ethnography
  gatekeeping 52, 55
  iterative approaches 47, 48n2, 64
  methodologies 49, 50, 52–53
  nonlinear nature of 47, 48n2, 51, 64
  purpose 45
  replication concerns 50
  rigor 45, 47–52
  in situ, taking place in 47
  subjectivity 46–48
  thick description 47
quantitative research 51, 52
questions
  advice 61
  behavioral listening research 143–145
  contrast 61

questions (*cont'd*)
  descriptive 61
  explanation 61
  first-order 144
  multiple-choice 5, 6, 8–9, 24, 26, 65n5, 79, 81
  "once-upon-a-time," 61
  quotation 61
  second-order 144
  third-order 145
  zero-order 143
questions, interviews 61
quotation questions 61

r

random error 27, 103
random sampling 31, 55
randomized controlled trials (RCTs) 105
ranked data 26
Rankin, P. T. 8, 32, 114, 578–583, 585
Raters' Perception of Immediacy Scale (RIS) 483
  see also nonverbal immediacy (NVI) measures
Rating Scale Rasch (RSM) 155
ratio scale 26
Rational-Experiential Inventory-40 (REI-40) 111, 501, 530–536
RCQ *see* Role Category Questionnaire (RCQ)
recall
  audio–video format of measures 81
  cued 79, 421, 422
  free 79
  and retention 6
  short-term 5–6, 77
  of spoken monologue 5
Receiver Apprehension (RA) 98, 381
Receiver Apprehension Test (RAT) 98, 313, 315
receptive listening 252, 254
recognition, assessment by 79
recording equipment, use of 132–133
referential skills 217
regulatory skills 217
REI-40 *see* Rational-Experiential Inventory-40 (REI-40)
Reis, H. T. 116, 516, 517, 519, 580
relational framing scales (RFSs) 537–544
relational framing theory (RFT) 7, 11, 537
relational listening (RL) 403
Relational Maintenance Strategies Measure 183
reliability of listening measurement instruments
  see individual measurement profiles
research of listening methodology/measurement
  see also affective components of listening; behavioral components of listening; cognitive components of listening; measurement of listening; sample studies; statistics; *individual listening research instruments*
  behavioral
    future trends 145–146
    types of questions for 143–145
  early tests 6, 9, 75
  experimental 47
  linearity 47
  longitudinal research 105
  multiple-choice questions 5, 6, 8–9, 24, 26, 65n5
  qualitative *see* ethnography; qualitative research
  retention and recall 6
  self-report
    improvement of 105–111
    uses and abuses of self-report methods 102–111
  statistical data, use of 35
  studies
    atheoretical in nature 9
    cohort 105
    comparison difficulties 7
    content of 54–55
    results, clarity as to implications 38–39
  theory, role in 9–11
respondent interviews 60–61
responsiveness 85, 106, 319
responsiveness validity 31
resting-state fMRI (rsfMRI) 292
retention and recall 6, 423
  see also memory
RFSs *see* relational framing scales (RFSs)

Rhetorical Sensitivity (RS)/Rhetorical Sensitivity Scale (RHETSEN) 545–551
Ridge, A. 127
Riggio, R. E. 343
right-ear advantage (REA) 76
rigor, qualitative research 45, 47–52
RIS (Raters' Perception of Immediacy Scale) 483
Roberts, C. V. 215, 617, 618, 619, 620
Rogers, C. R. 168, 267
Role Category Questionnaire (RCQ) 195, 218, 263, 327, 552–559
Role-Taking Task (RTT) 329, 330
Rost, M. 73, 89
Rubin, D. L. 363, 364
Rubin, R. B. 208, 214, 215, 321
  Communication Research Measures 315

S

sample studies
  see individual measurement profiles
sampling
  atypical-case sampling 56
  behavior 136–137
  criterion sample 56
  daily sampling of listening experiences 115–117
  development sample, administering items to 31–34
  event-contingent 116
  experience 114
  matching to population of interest 38
  maximum variation 56
  population inferences 32
  process inferences 31, 32–34
  purposive 55
  random 31, 55
  signal-contingent 116, 117
  snowball 56
  sparse-sampling 294
  standard procedures 32
  strategies 55–56
  studies *see* sample studies
  time-contingent 116, 117
  typical-case 56
Sawaki, Y. 82, 155, 156, 593, 594, 595, 597
scale length, optimizing 25
Science Technology Engineering and Mathematics-Active Listening Skills Assessment (STEM-ALSA) 170
scoring of listening measurement instruments
  see individual measurement profiles
second language (L2) learning 78–79
secondary task reaction times (STRTs) 201, 202
selective attention 75
selective perception listening model 7
self-appraisal 82, 438, 441
Self-Perceived Listening Competence Scale (SPLCS) 39, 111, 560–565, 569
self-regulation 82, 438, 441
self-report measures 86
  see also affective components of listening; *specific measurement instruments*
  acknowledgement of contribution of self-report scales 110–111
  advantages and disadvantages of using self-reports 98–102, 111–112
  and attention 75
  behavioral perspective 127–128
  beliefs about listening 111–112
  cause–effect relationships 102
  construct validity 29–31, 103
    generating strong evidence for 107–110
    creation of new constructs/measures only when necessary 106–107
  cross-sectional study, making greater use of 105–106
  improvement of self-report listening research 105–111
  internal states, measurement of 111
  limiting reliance on self-report measures 39
  motivations to listen, measuring 112–113
  survey instruments 117
  traditional measures 9, 114
  uses and abuses of, in listening research 102–111
Semantic Differential scales 26
Senese, V. P. 477
sensing 85, 86, 161

sensory-processing sensitivity (SPS) 299, 302
Sequential Tests of Educational Progress (STEP) 77
short-term memory (STM) 5–6, 395, 396, 397
signal-contingent sampling 116, 117
signal-to-noise (SNR) ratio 293
silence onset 199
skilled conversation 59–60
Skilled Counselor Training Model (SCTM) 336
social cognitive information processing 327
Social Management Scale (SMS) 205
sound effect onset 199
sounds, interpreting 5
SPLCS *see* Self-Perceived Listening Competence Scale (SPLCS)
Spradley, J. P. 61
Spunt, R. P. 77, 83–84
Stafford, L. 183, 420, 421, 422
Standards for Reporting of Diagnostic Accuracy (STARD) 413
STEP test 9
STRTs (secondary task reaction times) 201, 202
supportive listening 106, 138, 276
systematic error 103

t
Talkaholic Scale (Compulsive Communication Scale) (TAS) 566–572
talkaholicism 566
TAS *see* Talkaholic Scale (Compulsive Communication Scale) (TAS)
task-oriented listening 403
Team Listening Environment (TLE) 76, 573–577
Teel, J. B. 487
Test of Accurate Perception of Patients' Affect (TAPPA) 262, 263
Test of the Internet-Based Test of English as a Foreign Language (TOEFL iBT) 31, 87, 592–598
thalamus 75
thematic apperceptive technique stories (TATs) 327
theory, role in listening research 9–11
  see also research of listening methodology/measurement; *specific theories*
therapeutic listening 106
Time Studies 32, 114, 129, 578–591
TLE *see* Team Listening Environment (TLE)
TOEFL iBT *see* Test of the Internet-Based Test of English as a Foreign Language (TOEFL iBT)
Trait Affection Scale (TAS) 180–185
  trait given (TAS–G) 180, 181, 182, 183
  trait received (TAS–R) 180, 181, 182, 183
transcranial magnetic stimulation (TMS) 77, 295
triangulation 63
Type I error 111
types of listening measurement instruments *see* classification of listening measurement instruments

u
UDIP *see* Empathic Accuracy-Unstructured Dyadic Interaction Paradigm (EA-UDIP)

v
validity coefficients 109
validity of listening measurement instruments
  see individual measurement profiles
  concurrent validity 29
  construct validity *see* construct validity
  content validity 29, 156
  convergent validity 30, 163
  criterion-oriented 28–29, 247
  development sample, administering items to 31–34
  discriminant validity 30–31, 107, 108
  divergent validity 30, 176, 227, 275, 276, 283, 498, 512, 532
  external 115, 118, 192, 243, 449, 554, 555
  functional magnetic resonance imaging (fMRI) 293–294
  internal 39, 63, 118, 448, 449, 461

nomological network approach 30, 103
nonverbal immediacy (NVI) measures 485–486
portfolio development 27–34
predictive validity 29
psychometrical validation of derived scores 37–38
responsiveness validity 31
structural validity 156
substantive validity 156
threats, addressing 63
validity portfolio, developing 27–34
Välikoski, T.-R. 87, 321, 323, 375
VAS *see* Verbal Aggressiveness Scale (VAS)
Verbal Aggressiveness Scale (VAS) 599–604
Verbal Response Modes (VRM), taxonomy 140, 141, 605–611
Verona Patient-Centered Communication Evaluation (VR-COPE) 170
Vickery, A. J. 136, 164, 619, 620
video recordings 6–7, 81, 133
Villaume, W. A. 601, 613

W

Waldhart, E. S. 6
Watson, K. W. 113, 403, 613, 614
Watson–Barker Listening Test (WBLT) 6, 9, 21, 28, 31, 126, 227, 381, 419, 563, 612–616
WBLT *see* Watson–Barker Listening Test (WBLT)
Weaver, C. 124–125
Human Listening: Process and Behavior 7, 8
Weaver, J. B., III. 113, 613
Wechsler Intelligence Scale for Children (WISC) 396, 397
Wechsler-Bellevue Intelligence Scale (W-BIS) 331
Picture Arrangement (PA) subtest 330
Weinstein Noise Sensitivity Scale (WNSS) 475–481
Wheeless, L. R. 99, 205, 314, 486, 601
Willingness to Listen (WTL) 112, 617–623
Wolvin, A. D. 32, 103–104, 113, 560, 561, 581, 583, 584, 585
Working Memory Index 396
Working Memory Rating Scale 397
working memory (WM) 78, 394, 395, 397
Worthington, D. L. 37, 104, 113, 381, 406, 478, 498, 601, 619, 620

Z

Zohoori, A. 308