A Teacher's Guide to Educational Assessment

(Revised edition)

Iasonas Lamprianou and
James A. Athanasou

SensePublishers

A Teacher's Guide to Educational Assessment

A Teacher's Guide to Educational Assessment

Iasonas Lamprianou
European University-Cyprus
University of Manchester, UK

James A. Athanasou
University of Technology, Sydney, Australia

SENSE PUBLISHERS
ROTTERDAM/BOSTON/TAIPEI

A C.I.P. record for this book is available from the Library of Congress.

ISBN 978-90-8790-912-3 (paperback)
ISBN 978-90-8790-913-0 (hardback)
ISBN 978-90-8790-914-7 (e-book)

Published by: Sense Publishers,
P.O. Box 21858, 3001 AW
Rotterdam, The Netherlands
http://www.sensepublishers.com

Printed on acid-free paper

To international friendship

and collaboration

CONTENTS

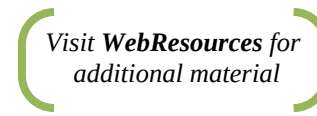
PREFACE TO THE SECOND EDITION

This book is a natural step beyond our earlier text *A Teacher's Guide to Assessment*, which was published almost six years ago. The purpose of this book is to offer a straightforward guide to educational assessment for teachers at all levels of education, including trainers and instructors.

The scope of this book is wider, however, and the targeted audience is broader than the first edition. It is designed to address the needs not only of those taking a first course in educational assessment and measurement but it can also usefully serve students at the post-graduate level, as well as experienced teachers, trainers and instructors who would like to update their knowledge and acquire practical skills using relevant quantitative methods. The book is appropriate for an international audience since it includes material and examples from Australia, the United States and Europe.

In this revised edition we have added new and important material which covers the assessment arrangements necessary for people with special needs, the use of technology for assessment purposes (such as e-assessment and computerised assessment systems); we have elaborated on the dangers of differential item functioning; we have extended the Rasch measurement material; and enriched the book with practical examples using Microsoft Excel.

In some cases, we wanted to offer more examples or more information or additional material on issues that could be of interest to some readers. In order to offer the opportunity for those who are interested to have access to this material without increasing the length of the book too much, we created a dedicated web companion to this book, which we called '*WebResources*'. Whenever additional material is available on the WebResources companion, you will see a sign with some descriptive text, like the one shown to the right of this paragraph. You may access this material at http://www.relabs.org/assessbook.htm.



The main message of the book is that assessment is not based on commonsense but on a huge body of international research and application over many years. Testing is a powerful, vital and large part of a teacher's assessment arsenal because it can be practical, structured and very informative. The correct use of testing, either in its traditional paper-and-pencil form or in its modern technology-based style can be a formidable ally for every teacher who aspires to practise evidence-based teaching and learning.

I am really grateful that my colleague James Athanasou from the University of Technology, Sydney, offered me the opportunity to take the lead on this new endeavour. I was responsible for re-writing much of the existing material, adding new chapters and for removing some of the material that was not deemed necessary today (most of it was moved to the WebResources). James kindly guided and advised me throughout the book, proof-read the manuscript and made sure that we produced a readable and scientifically acceptable book.

Wherever possible we have tried to provide references and every effort has been made to acknowledge or trace sources but if this has not always been possible we apologise for any error or omission. All acknowledgments have been placed directly in the text or in the endnotes.

We trust that this book is useful to you in your study of educational assessment and measurement. It is a fascinating field with rapid innovations and intense research output.

Iasonas Lamprianou
European University, Cyprus
University of Manchester, UK
September 2009

If I may add a few words to those of my colleague and associate Dr Iasonas Lamprianou. Readers may be surprised to learn that we have only ever met on one occasion in faraway Penang in Malaysia at a conference on educational measurement. For more than 8 years we have corresponded and collaborated without ever meeting face-to-face. And so this text is a tribute to international collaboration and friendship.

I am very grateful that Sense Publishers agreed to publish the second edition of *A Teacher's Guide to Assessment* and even more grateful that Iasonas agreed to become the first author. This handing over of the baton ensures that the tradition of this text, which started in 1997 with *Introduction to Educational Testing* (Social Science Press) will continue. Equally, we have decided to continue the tradition of donating the royalties to a charity for children with disabilities – the Estia Foundation in Australia or Radio Marathon in Cyprus.

The changes made to this edition are especially pleasing to me. The emphasis on criterion-referenced assessment and Rasch measurement has been strengthened. This perspective is – if I may say - unique amongst introductory texts on educational assessment. Rasch measurement is a point of view that we both passionately share.

Instructors can now also contact the senior author for a free copy of an instructor workbook to accompany the subject. This contains 12 lectures for a basic course in educational assessment together with student exercises and review questions. It covers the essence of the text for an introductory class on assessment. There are also a free copy of PowerPoint slides to accompany the lectures and these are available on request.

In editing the various chapters I had the opportunity in my retirement to reflect on the content. With each page I concluded that this text really expresses what I know and believe to be true about testing. I would say it also summarises most of what I tried to teach generations of students about this wonderful field.

Very few realise that assessment is the cornerstone of educational evaluation and research. It deals with fundamental concepts such as validity. Accordingly it has been a privilege to have been involved in educational and vocational assessment over the last 30 years of my career.

I have now retired from university teaching but I am very proud to be associated with Iasonas Lamprianou and this effort. I look forward to hearing of future editions. I hope that this will remain an applied handbook for teachers involved in assessment and a useful reference for those who are new to testing.

James A. Athanasou
Maroubra, Sydney – Australia
September 2009

CHAPTER 1

INTRODUCTION TO ASSESSMENT

Assessment has grown to become a multi-billion economic sector which shapes the educational and vocational future of millions of people worldwide. In recent years and throughout the world, assessment has been a fundamental component of everyday life in primary, secondary or tertiary education as well as in industrial or commercial training. The frequent exposure of people to assessment has cultivated very different attitudes: some laypersons have unbounded faith in the outcomes of assessment while others are dismissive of their value. Given the significant role that assessments play in modern education as well as their impact on career prospects, we have always considered that there is a need for teachers to be skilled and informed in this area but sadly it has not yet featured in all teacher-preparation or trainer-preparation courses. Many teachers and trainers are still left to acquire their assessment expertise through in-service courses, the assistance of colleagues or by trial-and-error.

It is helpful for you to be knowledgeable about educational assessments because the results may have a significant impact on other people's lives: they are widely used for selection, certification, diagnosis, special instruction or placement. Furthermore tests, exams, quizzes, projects, assignments or portfolios are part and parcel of your own teaching and it is valuable for you to have some knowledge of their development and use. The ability to develop worthwhile assessments does not come naturally but it is a skill that can be acquired and it needs knowledge as well as experience. Moreover assessments demand your attention because they are a sizable chunk of your professional workload. It has been estimated that teachers spend as much as one-third of their time in assessment-related activities[1] and there are indications that this workload is not decreasing. Indeed, some countries are scrapping external examinations so as to give more focus on teacher assessment results – the latest example is England which removed the external examinations for 14-year olds in October 2008 and relied upon teacher assessments.

It is possibly a sign of the times in which we now live that society is prepared to put more faith in your professional opinion. Consciously or otherwise, assessments play a role in your teaching. For instance, your assessments may occur informally during evaluation of your instruction and act as catalyst for helping the learner; or at the other extreme a formal assessment at the end of a course may determine someone's career prospects. This means that you need to be informed about the effects of different forms and methods of assessment. During the course of reading this book you may wish to ponder the extent to which you want assessments to become a seamless component of your teaching or the extent to which you might

want them to become a discrete part of the process. Different people use assessment in different ways, depending on the situation; there are no fixed rules.

Both experienced and novice teachers acknowledge the importance of assessment in learning and instruction for various reasons. They usually focus on assessment as a vehicle for giving some feedback on teaching or assessment as a means of providing evidence of learning or assessment as meeting the requirements of the administration. At the same time many teachers do worry about the quality and the quantity of the testing that they are required to undertake. They worry about issues like the fairness of marking, the accuracy of the questions asked, the suitability of tests for students and the usefulness of the information eventually obtained. In part, this book seeks to address some of these concerns and in the remaining sections of this chapter we shall provide some background on the nature and role of assessment and a brief outline of its development. The discussion concludes with a look at some ethical issues associated with assessment. Throughout this text, you should feel free to skip those sections that are not of direct interest to you.

BACKGROUND TO EDUCATIONAL ASSESSMENT

Assessments are dominant aspects of our culture and are encountered at many points in our life. If you are under 60 years old then you may have undertaken your first assessment within one minute of being born. This is the Apgar test (see Figure 1) which is an assessment (devised by Dr. Virginia Apgar in 1952) based on separate tests that it meets the formal definition of an assessment as we accept it in this book. It gives the general overall condition of a newborn within minutes of birth. The assessment is based on a score of 0-2 for five modalities: heart rate, respiratory effort, muscle tone, reflex irritability and colour.

Heart rate:	absent, <100, >100;
Respiratory effort:	absent, slow, irregular, good;
Muscle tone:	limp, some flexion, good motion;
Reflex irritability (nasal catheter):	no response, grimace, cough or sneeze;
Colour:	blue or pale, body pink extremities blue, completely pink.

A score of 10 is perfect; 1-3 is severe depression; 4-7 moderate depression; and 8-10 no depression

Figure 1. Apgar Test.

In education, the word ‘assessment’ is used in a special way that is derived from but different to its ordinary, everyday meaning. The word ‘assessment‘ comes originally from the Latin *assessare* and meant to impose a tax or to set a rate and modern dictionary definitions refer to the valuation or financial meaning of assessment. This word has crept into the field of educational testing largely through

psychology. For instance, the term assessment was used during World War II to describe a program of the Office of Strategic Services, which involved the selection of personnel for secret service assignments. This program involved situational tests and staff rated candidates on many traits. In these assessments, unlike the results of a single test, the assessor usually combined data from different sources.

In the last decade, the word 'assessment' has taken over from terms such as 'testing'. Firstly, it was seen as a broader term than 'test'. It now encompasses many different educational practices, such as portfolios, case studies, presentations, simulations or computer-based activities. Secondly, it also took into account divergent processes of assessment such as teacher assessment, self-assessment and peer-assessment. Thirdly, it gave some expression to more liberal views in education that were opposed to the oppressive, mechanical and unthinking use of tests. For many people the word test may conjure up images of three-hour examinations or endless multiple-choice questions that bring back recollections of exam anxiety. Using the word assessment avoided many of the negative connotations of the word 'test'. Even children in kindergarten are now familiar with the term assessment and one wonders whether in the future it will eventually share some of the same negative associations as its predecessor.

At the outset, we have tried to give you a broad working definition of assessment and one that typifies its use in modern educational circles:

> Assessment is the process of collecting and organising information from purposeful activities (e.g., tests on performance or learning) with a view to drawing inferences about teaching and learning, as well as about persons, often making comparisons against established criteria.

Apparently, you are expected to collect some information from purposeful (not random) activities. This implies that assessment should be part of your planning and cannot be left to happen in a haphazard way. It is also implies that this information has to be organised – just collecting and storing information is not enough. And all this must serve a specific purpose.

This definition also means that you are free to include many different types of activities (assignments, exercises, projects, quizzes, simulations) under the umbrella of assessment. It also means that assessment is principally a professional process of collation, comparison and judgement and inferences are drawn as a result of the process.

Although this is one aspect that we would like to stress, it is recognised that 'assessment' is also being used daily by students and teachers to refer in a shorthand type of way to a particular task that may have been assigned. As an example, the Department of Education and Training in New South Wales (Australia) referred to the State-wide English Language and Literacy Assessment, which is administered to some 150,000 Year 7 and 8 students, as a test of reading, writing and language.[2] Similarly the Victorian Curriculum and Assessment Authority defined an assessment as: 'A task set by the teacher to assess students' achievements of unit outcomes.'[3]

This everyday use of the word to refer to tasks or tests is not a problem for us as long as we recognise that underlying all of this is the systematic process of collecting information. A key part of this process is the various assessment tasks or events.

HISTORY OF EDUCATIONAL ASSESSMENT

Assessments have not sprung from some historical vacuum but have evolved from experience over many thousands of years. In fact, some of the earliest classrooms were those of the Sumerian civilisation from which the clay tablets on which students practised their writing have been recovered. As you will see shortly, current testing practices can be traced back over some 4,000 years and provide a substantial knowledge base about assessment.

The triposes are the formal examinations at the University of Cambridge, in which undergraduates are required to obtain honours in order to qualify for the degree of Bachelor of Arts.

The word *tripos* comes from the fact that the examiner, the 'Ould Bachilour' of the University, sat on a three-legged stool. The examination took the form of a debate or wrangle and concentrated on Grammar, Logic and Rhetoric.

The Mathematical Tripos is probably the most well-known of the triposes. At the time of Sir Isaac Newton's discoveries, mathematics dominated other studies and 'tripos' came to mean the examination in mathematics. The tradition became that one had to pass the mathematical tripos before being able to specialise in classics or other studies. This tradition continued until around 1850.

The mathematical tripos is now a written examination but some of the traditions surrounding the exam have been preserved. The final year results continue to be announced from the balcony of the Senate House and the top students are still called Wranglers.

Sources: Faculty of Mathematics, Faculty of Engineering, University of Cambridge

Figure 2. The tripos examination.

A long history of educational assessment can be dated from at least 2200 BC when the Mandarins set up a civil-service testing program. For the most part oral examinations (*viva voce*) were used until the late 1800s to evaluate achievement. A famous example of a rigorous traditional oral examination was the tripos (see Figure 2).

*Visit **WebResources** for a description of additional relevant examples*

In practical fields, there were examples of competency assessments such as those in the middle ages, where an apprentice was required to complete a master piece before acceptance into a guild.[4] The master piece is described in Figure 3.

> In medieval times parents paid a fee to place their seven-year old son as an apprentice with a master of the guild. Both parties signed a contract called an indenture that the boy would work for the master for seven years as an apprentice and the master promised to train him.
>
> He became a journeyman after the period of apprenticeship. The term comes from the French 'journee', (day), and meant that the journeyman was paid a daily rate for his work.
>
> After several years as a journeyman the craftsman would submit a piece of his best work to the guild for approval. The example that he created was known as his 'master-piece' and if it was considered good enough, he was granted the title of 'Master'.

Figure 3. The master piece.

Formal written testing, as you know it today, dates from around the 1860s. An early example of formal written testing survives from a 1904 college entrance examination (see Figure 4).[5]

> Test-takers were assigned to write a two-page essay on one of the four statements listed below. The test did not identify the works of literature.
>
> 1. Describe the character of Mr. Burchell, and compare or contrast him with Dr. Primrose. How far does he influence the course of events in the story?
> 2. Locksley shoots for the prize.
> 3. The elements of greatness in Shylock's character.
> 4. Describe, from the point at which the Albatross "begins to be avenged", the events that precede the Mariner's being left "alone, on the wide, wide sea".
>
> Source: Education Week, June 16, 1999

Figure 4. 1904 College Entrance Examination Board Test.

The advent of computers has given the opportunity to develop new tests with very desirable characteristics: they are cheaper to administer, they can be more realistic (they may use multimedia and simulations), they are often scored automatically and it is much easier to report the results in many different and informative ways. The second generation of computer-based tests are adaptive, in the sense that they automatically adapt the difficulty or the content of the test to the needs and the characteristics of the person who takes the test. Computer-based and computer-

adaptive tests have achieved widespread use in education and are now a useful tool for both teachers and students. Computer-adaptive testing via the Internet has also been adopted by large testing organisations around the world and today millions of students all over the world have the opportunity to take cheap and state-of-the-art tests for high stakes assessment purposes.

The many landmarks in the history of educational testing indicate the continual development of this field and that assessment practices have reflected the prevailing educational policies. We have summarised some of these developments in Table 1. It is not important that anyone should recall the details of these but that at the very least you are now aware that educational assessment is an ongoing endeavour with a significant history.

The gist of these few background details is that some form of assessment has been a regular feature of Western education but that educational tests of a written nature are a relatively recent invention. Much of our past assessment was informal, oral and practical in nature.

Table 1. Selected events in the history of educational assessment

2200BC	Mandarin civil service testing program
1219AD	University of Bologna holds oral examinations in law
1636	Oral exams for awarding of degrees at Oxford
1845	Printed examinations first used in Boston
1864	Fisher (UK) develops sample questions and answers to grade essays
1897	Rice surveys spelling abilities of US schoolchildren
1908	Objective arithmetic tests
1926	Scholastic Aptitude Test first used (now a common basis for university entry in the US)
1930	Australian Council for Educational Research established (premier educational test distributor in Australia as well as the major research and development agency)
1947	Educational Testing Service (major testing agency) founded in US
1952	Development of Rasch model of measurement
1967	US Supreme Court ruling against group ability tests to stream students
1971	US Supreme Court rules that selection tests must have a direct relationship to job performance
1980	Development of item response theory (a new approach to analysing test scores and responses)
1985	Publication of the Standards for Educational and Psychological Testing
1986	First computerised adaptive tests are given to students in Portland (OR) public schools
1988	Code of Fair Testing Practices in Education
2001	Fourteenth Edition of The Mental Measurements Yearbook published (major listing and review of all tests) first published in 1938 by Oscar Buros

There is now a widespread desire to assess individuals for purposes of grading and evaluating educational outcomes. The advent of large classes and organised education has meant that much of our assessment has changed dramatically. It has become more formal, involved greater standardisation and become more quantitative in nature. It has come to meet other demands (administrative, policy, accountability) as well as meeting the needs of the learner or the teacher. Before discussing the purposes of assessment it might be useful to show you some of the extent of standardised assessment in education in Australia, Europe and the USA. We have included examples in the WebResources companion from the New South Wales Department of Education, the Qualifications and Curriculum Authority (UK) and the American College Testing Program to indicate the widespread use of formal educational assessment in primary and secondary education.

*Visit **WebResources** for more information on the widespread use of standardised assessment*

The increase in formal testing and assessment procedures has been linked with large-scale education systems.[6] Some of the developments in testing are of recent origin and the pace of change has certainly quickened. This reflects (a) the growth of mass schooling and expansion of tertiary education with its need for large-scale written rather than oral assessments; (b) the increasing access to technology leading to newer forms of assessment, such as printed test papers originally and most recently the use of computer-based testing; and (c) other developments in fields such as psychometrics (i.e., the field of psychological testing and measurement), statistics and educational research.

PURPOSES OF ASSESSMENT

Assessment can also be linked to the following educational purposes: diagnosis, prediction, placement, evaluation, selection, grading, guidance or administration. In all fields of education, assessment results are used to decide about students (i.e., student progression), to decide about teaching and learning (i.e., curriculum decisions) and increasingly assessments are linked with certification of competence and the validation of performance on job-related tasks.

Leaving aside these policy and administrative considerations, you can make many decisions that are dependent upon some form of assessment. For instance you can use assessments to give answers to questions, such as the following.

- How realistic are my teaching plans for this group?
- Are my students ready for the next unit?
- What learning difficulties are students facing?
- Which students are underachieving?
- How effective was my teaching?
- Which learners are advanced?
- Which learners are gifted or talented?
- Which learners require special assistance?

Assessments are a natural accompaniment to instruction. The assessment process is an integral aspect of education that helps you make judgements about students'

current levels, about the most appropriate method for teaching, or when to conclude teaching a topic. Table 2 shows a few instances where there is scope for educational assessment at many points in the teaching process.

Assessments can be a valuable component of your teaching and few formal courses would be complete without an assessment component. Even the best assessment plan, however, can test only part of an individual's educational achievement.[7] Your own experiences would already have shown you that a person's development and growth, his/her interests and values may not be assessed accurately or consistently. As teachers our aim is to use assessments fairly, to analyse the results carefully and to combine the results with other evidence of progress in order to provide the best possible evaluation of our students' achievement and development.

Table 2. Assessment and the teaching process

BEFORE TEACHING
To determine the level of skills/knowledge prior to instruction
To diagnose learning difficulties or advanced achievement
To plan instruction
DURING TEACHING
To make on-going changes and improve teaching and learning
To focus on a small segment of instruction
To identify learning errors and misconceptions and take remedial action
AFTER TEACHING
To certify the attainment of outcomes at the end of learning
To self-evaluate your teaching effectiveness and improve teaching plans
To assign grades and communicate results

When assessment is conducted appropriately it will help people to learn in a way that is meaningful and encourage their motivation to learn. Assessment can be used to improve the quality of learning and instruction. It provides important feedback on progress and helps people to realise their strengths and weaknesses without being judgemental. From your viewpoint, effective assessments give you the necessary information to decide how a person is progressing with their learning.

Appropriate assessments can encourage what we call deep rather than shallow approaches to learning in a subject or occupation. What does this mean?

– *Deep approaches*: learners focus their attention on the overall meaning or message of teaching. Ideas are processed and interests developed in the topics. Where possible the content is related to experiences to make it meaningful. Modern approaches to assessment such as self-assessment and peer-assessment may help achieve these goals.
– *Surface approaches*: with this type of learning the focus is on acquiring skills, test-taking techniques or knowledge that is necessary to do well on assessments.

There is less focus on understanding and on being able to transfer knowledge in other situations. There is a concern for the details that need to be remembered for assessment purposes. The quality of the learning outcomes is much lower. Assessment approaches that focus too much on past papers and test taking techniques are of this kind.

If you are not careful, then the types of assessments that you use may inadvertently develop only surface or shallow approaches to learning. You need to remember that the type of the assessments you use may be a valuable tool that may encourage learning but if not used wisely, may also deter learning and make your work more difficult.

ETHICAL ISSUES ASSOCIATED WITH ASSESSMENT

We would like to complete this chapter by indicating to you that there are some important ethical issues associated with assessments. Some of these issues relate to ensuring that people are assessed appropriately, that everyone is marked on an equal basis, that they are not disadvantaged by assessment results, that any assessment is for the benefit of the student and that their confidentiality or privacy is respected. In particular, assessment results are the property of the person and should only be released to other colleagues on a need to know basis and only if the person will benefit directly (the next year's teacher may want to know past performance of his/her students in order to plan teaching) or indirectly (the teachers of the school may use the results to evaluate the effectiveness of different teaching approaches in their school).

A fundamental question is: 'Will my pupil, student or trainee benefit from this assessment?' While some form of assessment is helpful in teaching, we may need to reduce the inappropriate use of tests, especially long tests that may cause anxiety and may tire the students. Even a cursory glance at most curricula should show you that most students are being over-assessed or that assessments are being used in a less than optimal way. Assessments have the potential to be very helpful but the application of assessment also needs to be well-organised. Some of the critics of modern-day testing deserve attention because assessment has not always been applied appropriately.

Most of us are far removed from the educational decision-making that affects our everyday work but if you are at the coalface then you have to put these policies into practice. And this raises a large number of ethical issues, many of which centre around assessments.

We can aim for an ideal assessment system, however, and it might be one that encourages meaningful learning, that has fair and equitable procedures and that produces results that are both valid and reliable. Some possible steps for you to consider are to:

- use tasks that encourage learning;
- use tasks that encourage interest;
- ensure that tasks are linked to the learning outcomes;
- offer tasks that foster self-direction;
- use a variety of tasks as part of your assessment;
- give grades that reflect true levels of achievement;

– treat every learner fairly and equitably;
– give timely and appropriate feedback;
– provide alternative assessments for special groups; or
– adhere to a code of ethics for test users.

We have taken the liberty of reproducing for you some relevant sections for educational test users of the Code of Fair Testing Practices in Education[8] (see Appendix A). The Code was not intended to cover the use of classroom tests made by individual teachers but it is considered helpful because it outlines some major obligations that we have to test takers in education. It is directed more towards those assessments used in formal testing programs.

There are also influential professional associations that are involved in testing, such as the International Testing Commission, International Association for Educational Assessment and the National Council on Measurement in Education and at the time of writing, regional professional associations for assessment are being formed in the Asia-Pacific region. As well, there are published Standards for Educational and Psychological Testing jointly prepared by the American Educational Research Association, American Psychological Association and National Council on Measurement in Education. Some links to professional associations, helpful list-serves and assessment resources are provided in the WebResources companion. Appendix B provides references to additional assessment topics.

*Visit **WebResources** for examples of professional assessment bodies*

SUMMARY

In this chapter we have tried to provide you with some idea of the importance of assessment for teaching but also of the importance of assessment for the lives and the future of your students. The fundamental aspect of assessment is not about pieces of paper or exams or marks or grades or intricate scoring systems; assessment is really about the question of how we use certain tasks or events to prove or establish that learning has occurred, that someone is able to do something or knows something, or even to provide information to evaluate our own teaching effectiveness and improve our teaching methods. The process of assessment and its key components address the important issue in this world of how do we know that something is the case – you may regard this as an evidence-based practice of the teaching profession.

Assessments have assumed social importance because of their links with formal qualifications and because more and more people pay serious attention to education and its outcomes. Modern educational systems are now centred on various forms of assessment and it is important for teachers to be familiar with educational assessment. In part, this trend reflects the community's desire for validity, fairness and objectivity. The need to consider the appropriate and ethical use of assessments in teaching has been stressed. The valid concerns or problems that some critics have identified, such as the use of test results to grade pupils, teachers and schools, or the value of testing as an indicator of competence or the role of testing in

teaching and learning may not really be assessment problems. They are mainly issues of educational policy and administration and how assessment results ought to be used rather than misused.

In the next chapter, specific issues associated with the nature of assessment are considered in further detail. References to the sources cited are listed in the Notes. We have also provided a glossary of the terms used throughout the text. Now take a moment to review the key ideas and maybe undertake some of the exercises or activities at the end of this chapter.

-oOo-

REVIEW QUESTIONS

Try these review questions to help you reinforce some of the key ideas in this section. These are all true-false questions to make it easier and quicker for you to complete.

You will find a set of similar questions at the end of each chapter. First, think whether each statement is mainly true or false. Then just circle the T (True) or F (False). If you are not sure, just guess.

T	F	In education, the word ‘assessment’ is used in a special way that is different from its ordinary, everyday meaning
T	F	In education, the term ‘assessment’ has taken over from terms such as ‘testing’
T	F	The everyday use of the term ‘assessment’ refers to a process of collection of information, judgement and comparison
T	F	Educational assessment has developed over a period of some 2000 years
T	F	*Viva voce* refers to an oral exam
T	F	Written formal testing dates from around the 1500s
T	F	Psychometrics refers to the field of psychological testing and measurement
T	F	Assessment results are used to decide about students
T	F	The main scope for assessment is after the teaching process
T	F	The *Code of Fair Testing Practices in Education* outlines obligations that we have to test takers
T	F	Consent of a test taker is required before providing results to any outside person or organisation
T	F	Educational test results are a privileged communication to other teachers
T	F	It is helpful to explain how passing test scores are set

EXERCISES

Here are some review exercises for you to answer or they can be used as the basis for discussion.

1. What is one stated purpose for assessment in the education or training context in which you work?

2. Do you consider that assessment is essential in education or training?

3. Indicate three classroom teaching decisions that can be made by using the results of assessments.

4. How can you use assessments to influence your students' learning?

5. What valid reasons would a teacher have for assessing a group on the first day of instruction?

6. Reflect on and write down the reasons you have for testing in your organisation.

7. Which document sets out the assessment and testing policy for your organisation?

8. Comment on the following approach to assessment in mathematics.

– Having assessment be an integral part of teaching
– Focusing on a broad range or mathematical tasks and taking a holistic view of mathematics
– Developing problem situations that require the application of a number of mathematical ideas
– Using multiple assessment techniques, including written, oral, and demonstration formats
– Using calculators, computers, and manipulatives in assessment.

Source: *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teachers of Mathematics, l989, p. l9l).

9. Read the following excerpt from the *FairTest Principles* produced by the National Center for Fair and Open Testing. Comment critically on these principles and highlight any ethical issues.

Assessment of student learning is undergoing profound change at the same time reforms are taking place in learning goals and content standards, curriculum, instruction, the education of teachers, and the relationships among parents, communities, schools, government, and business. These Principles provide a vision of how to transform assessment systems and practices as part of wider school reform, with a particular focus on improving classroom assessment while ensuring large-scale assessment also supports learning. To best serve learning, assessment must be integrated with curriculum and instruction.

High quality assessment must rest on strong educational foundations. These foundations include organizing schools to meet the learning needs of all their students, understanding how students learn, establishing high standards for student learning, and providing equitable and adequate opportunity to learn.

The Principles reflect an 'ideal'–what the National Forum on Assessment believes is the best that assessment can be and do. We understand that they will not be implemented immediately or with great ease. We do firmly hold, however, that education systems must move toward meeting these principles if assessment is to play a positive role in improving education for all students.

Principle 1: The Primary Purpose of Assessment is to Improve Student Learning
Principle 2: Assessment for Other Purposes Supports Student Learning
Principle 3: Assessment Systems Are Fair to All Students
Principle 4: Professional Collaboration and Development Support Assessment
Principle 5: The Broad Community Participates in Assessment Development
Principle 6: Communication about Assessment is Regular and Clear
Principle 7: Assessment Systems Are Regularly Reviewed and Improved

Source: FairTest, The National Center for Fair & Open Testing, http://www.fairtest.org/princind.htm

10. Read this brief excerpt from the National Council on Measurement and Evaluation (December 2000 NCME Newsletter). What ethical issues are involved?

Opponents of High-Stakes Tests Seek To Breach Exam Security

Some opponents of high-stakes tests have turned into would-be saboteurs. In at least three separate instances over the past six months, people apparently have sought to undermine the purpose and validity of state or district exams by sending copies of them to newspapers in the hope of having the questions published. Two leading daily newspapers, The Atlanta-Journal Constitution and The Los Angeles Times, and Substance, a monthly teacher-run publication in Chicago, have been the recipients. Only the latter actually published questions from a test.

*Note: You can visit the **WebResources** for more Exercises*

CHAPTER 2

THE VARYING ROLE AND NATURE OF ASSESSMENT

> Every lecturer has had this experience at one time or another: You're explaining some especially intricate and fascinating aspect of your discipline when you see a hand shoot up in the back row.
>
> "Yes?" you ask, eager to engage on a favourite topic with a bright, inquisitive mind.
>
> "Um, do we have to know this? Will it be on the test?"
>
> As far as students are concerned, there is nothing more central to the learning experience than assessment. Some learning researchers call this the backwash effect. The type of assessment students know will be coming determines when they "tune in" to a lecture and when they "tune out." Evidence from student diaries indicates that students spend less than 10 percent of their time on non-assessed academic work[1].

No matter whether you are a primary, secondary or higher education teacher, the above extract may remind you of one or more similar events in your own career. It shows how important assessment's role can be and how it may help to shape many aspects of any education system. It shows that assessment can either be the driving force behind improved teaching and learning or a catalyst that can de-motivate learners, destroy the climate of the class and have a negative impact on learning. Whether assessment has a positive or negative impact is determined by its aims and the way it is implemented.

In the previous chapter we mentioned briefly some of the aims of assessment: for diagnosis, motivation, prediction, placement, evaluation, selection, grading, guidance or administration. Each one of those aims of assessment has its own importance for any education system. In this chapter, we will refer to the different roles of assessment a little bit more closely and we will investigate how these different roles may be served more efficiently when different assessment approaches (such as tests, portfolios, performance assessments) are utilised where and when needed.

It is possible to split assessments in two large categories based on their perceived importance: high stakes and low stakes assessments. High stakes assessments are those which are perceived by one or more groups of stakeholders as very important for some reason. Such high stakes assessments can be end-of-year examinations which are used for school graduation or university entrance examinations. High stakes assessments can also be those assessments that certify eligibility to exercise a specific profession or the assessments which evaluate the effectiveness of schools or teachers. In effect, the high stakes nature of an assessment is not determined by any characteristic other than the consequences of the intended use of the assessment results.

High stakes assessments

A high stakes assessment such as an examination may have a huge impact on the life of various stakeholders. To mention just a few examples, students may proceed to the next level of education (e.g., access to tertiary education); graduates may be licensed to practice a profession (e.g., medicine); teachers may get a promotion or lose their jobs and schools may be closed down depending on the performance of their students; students may get a PhD if they submit their thesis and succeed in the oral defence (examination). There are cases, however, where high stakes assessment benefits other people as well. For example, strict and thorough assessments for licensing people to work as pilots or doctors benefit the whole society because they reassure us that only the most competent individuals will practise the most demanding professions. To offer another example, university entrance assessments seek to ensure that only the most academically oriented students will proceed to further education so that scarce public resources are used more efficiently.

As shown above, most often high stakes assessments serve important social purposes which may also have significant financial, political as well as educational benefits. They are used frequently as instruments to monitor the education output of schools as well as countries. International bodies, as well as governments use large scale assessments to compare their education output to that of other countries. Frequently, leading financial journals like *The Economist* take up a lot of their space to discuss results of international assessments. These international comparisons of student achievement involve assessing the knowledge of elementary and secondary school students in a number of subjects. These subjects usually include mathematics, science, reading and writing and technology. The main medium of assessments is written tests because they are very practical and easy to standardise. They use high quality test items that have been agreed upon by participating countries. Complex comparability studies have been carried out since 1959 and have reached a level of maturity.

One typical example of such studies could be the Trends in International Mathematics and Science Study (TIMSS) which included different types of items such as multiple-choice, short answer, extended response and performance items. TIMSS was administered by personnel from the participating countries after they were given extensive training to assure the high quality and comparability of the TIMSS data. Quality-control monitors were also hired from different countries and they were trained to visit the national research centres and to review their procedures. Although it might be theoretically desirable to enrich international comparability studies with other means of assessment such as laboratory tasks, this would lead to huge practical and logistical problems as well as increased costs; therefore, tests remain the main assessment medium.

*Visit **WebResources** for more information on international comparability studies*

High stakes assessments are also used frequently for selection, placement, grading or evaluation purposes and are becoming more and more important for more countries around the world. Although there are some countries with excellent education systems, like Finland, which still tend to avoid competitive, high stakes

tests, today many countries are shifting towards more formal, paper-and-pencil, high-stakes assessments. For example, Sweden, a country with a very weak high-stakes assessment culture is now reforming its educational system, firming up the assessment process, emphasising the role of teachers as assessors, investing money on training teachers to build their own tests and introducing more external tests. England is now piloting the new concept of single-level tests: they are intended to be short (but focused and more accurate) tests assessing the curriculum at a particular level in order to verify teachers' judgements (students will take a test matching their ability).

Ukraine, Georgia, Lithuania as well as Russia and other former Soviet republics are refreshing their interest in high stakes, formal assessments, emphasising fairness and quality for access and equity issues. The recent huge investment of resources in the Anglo-Saxon world (USA, Australia, New Zealand and U.K.) on issues directly related to high-stakes assessment (and especially tests) is also very impressive. All over the world, formal, large scale and high stakes assessments seem to be perceived as one of the main driving forces not only for selection, reporting and administration but also for evaluation and prediction.

Written tests are the preferred option in almost all countries around the world when it comes to large scale, high stakes assessment. In England, 600.000 pupils at each one of the first three Key Stages of mandatory education took end of year tests in order to assess attainment – in other words, to determine the output of the education system. It would be rather extreme to suggest that other means of assessment such as an oral examination or portfolio evaluation would be more practical than paper-and-pencil tests. For example, it would take a breath-taking amount of resources (money, time, personnel and infrastructure) to assess those students with personal interviews, and it would also take a surprising amount of time to assess them through portfolio or project evaluation. In the context of high stakes assessment, practicality is not, however, the only virtue of written tests.

Of course, an assessment may not be perceived by everybody as high stakes. For example let us consider the situation where university students who wish to skip an introductory level course are given a screening test to determine whether they already know the content and can proceed to the next stage. Succeeding on the test will reduce the time spent at university. Therefore, students who need to graduate and get a job as soon as possible will consider the test to be high stakes but those who do not worry too much about the extra cost of graduating from the university a bit later may consider it to be of low stakes.

Low stakes assessments

Low stakes assessments are also gaining ground all over the world. It is widely recognised today that more resources need to be invested in order to improve informal, classroom based, low stakes assessment. The reason behind this is that educationalists around the world have realised that classroom assessment, if used wisely, can be a powerful tool for diagnostic, placement and motivation purposes,

as well as for grading and prediction. But the most useful purpose of classroom assessment is arguably diagnosis.

For example, try to picture a situation where a primary school teacher needs to teach 11-year-olds some basic principles of probabilistic thinking, which is one of the goals of a mathematics curriculum in many countries around the world. Probability is a topic which is taught at the later grades of primary schooling but all children come to school with some relevant intuitive concepts.

From everyday life while playing with friends, or through observation or through trial and error, many children already know that if they have a bag with two red and four blue pencils, if they close their eyes, put their hand in the bag and randomly pick one pencil they are more likely to pick a blue rather than a red one. In a recent experiment, 426 primary school pupils in a European country were asked a similar question as above and 85% gave a correct response saying that there were more blue than red pencils in the bag so they would expect to get a blue pencil by chance. Then, the same pupils were asked a similar question but in a different context.

They were put in the context of a zoo, where there were two elephants and four monkeys. They were told that the staff of the zoo would like to wash all the animals but they did not mind with which animal they started. So, they decided to pick one animal purely randomly. The children were asked which animal was more likely to be selected by chance alone: an elephant or a monkey. A staggering 20% of the children gave responses based on non-probabilistic processes. Some said that the staff of the zoo would pick an elephant first because an elephant is usually dirty but the monkeys are cleaner. Other children said that the elephant can neither run nor climb on the trees; therefore, an elephant was the animal most likely to be picked. When challenged further with an 'explain why' question, it was obvious that these children would frequently switch to non-mathematical, subjective answers to various questions when they found the context of a question interesting; in other situations they would revert back to more scientific thinking, based on their observations of every-day life. It would obviously be a bit more difficult for a teacher to teach those children their first formal concepts of probability, and special teaching and preparation would probably be required to motivate them to comprehend the meaning of chance.

Thankfully, in any context like the above, a primary school teacher could use ready-made tests with tasks and questions especially designed to pinpoint the cognitive errors and misconceptions of the pupils. Such tests are called diagnostic because their aim is to identify errors and misconceptions early on, before the teacher spends one or two weeks trying to introduce the children to a new concept.

Diagnostic tests are excellent tools because they are easy for the teacher to administer and score; and can also be very easy for the pupils to complete. In some cases, the pupils read a question, pick one out of five multiple-choice options (or provide a short response) and then explain very briefly why and how they reached this answer. The tests are deliberately very short, and they do not aim to measure achievement – they only focus on a very specific and

*Visit **WebResources** for more information on diagnostic tests and performance assessment*

narrow sub-domain trying to find errors and misconceptions. In one or two minutes, a teacher can scan a pupil's responses and prepare for the next class in the best possible way, taking into account the intuitive errors and misconceptions the pupils bring to the class.

The above paragraphs have demonstrated some of the important social, financial and educational roles of assessment. Tests have traditionally dominated the arena of high stakes as well as low stakes assessments, although other methods such as performance assessments and portfolios have their own place, especially when skills are assessed. There are growing numbers of teachers and schools that would like to see a more frequent use of alternative methods of assessment, even for high stakes purposes such as selection for access to tertiary education. Such initiatives, although non-practical and expensive have made their appearance in the last years in various places: one example is the New York based Performance Standards Consortium, which is a group of 28 high schools currently promoting the concept of less high stakes tests and more performance assessments.

In all cases, however, the development of high quality assessment instruments and activities demands not only in-depth subject matter knowledge but also a good deal of experience and theoretical knowledge. Teacher-made assessments can be very efficient and informative if they are tailored to the needs and the characteristics of the students of their class; but to be able to build high quality assessments the teachers need much initial experience and in-service training. This has been widely recognised and more and more countries invest resources to improve the assessment skills of their teachers. For example, the body of school inspectors in Wales (UK) recently published guidelines asking all schools to provide opportunities to their teachers for adequate in-service training to update their assessment skills. Sweden has recently been reforming its education system and investing money on training teachers to build their own tests and introducing more external tests. Across Europe more and more countries spend resources to help their teachers improve the quality of assessments. The next section explains the complexity of building high quality assessments, both for high and low stakes purposes.

VALID AND RELIABLE ASSESSMENTS ARE DIFFICULT TO DEVELOP

Assessment results need to be reliable as well as valid. We will deal with the concepts of validity and reliability in much depth later but it suffices here to say that an assessment is valid to the degree that its results determine or measure accurately what they were supposed to determine or measure. The assessment is also reliable to the degree that the assessment results are consistent. So what does a low-stakes, teacher-made assessment involve?

First, the teachers need to study the curriculum very carefully and comprehend and unpack the standards (i.e., understand the level at which their students should be working). This is not as easy as it may look at first glance and it is not something a teacher should wait to build up by experience – it does not come naturally; it needs training. Tables of specifications need to be created for each separate sub-domain within each subject. This is necessary for the teachers if they are supposed to know exactly what they should assess.

Then, a teacher needs to decide how, when and where assessment will take place and for what purpose: for example, a diagnostic test at the beginning of the week may be shorter and more narrow in scope but the questions need to be well prepared to address specific possible errors and misconceptions. On the other hand, a grading test needs to cover as much of the taught curriculum as possible. The methods of assessment need to be determined (e.g., a project or a test) depending on the available time and the scope of the assessment. The actual assessment needs to be prepared (i.e., the questions of a test need to be written) and a scoring guide (i.e., a rubric) needs to be developed and trialled. Then, the assessment needs to be administered (i.e., conducted), scored and the results need to be analysed to draw conclusions. Finally, grading and reporting may (or may not) happen, depending on the purpose of the assessment exercise.

Among all those issues just mentioned, a teacher needs to keep in mind that the assessment needs to be fair to everyone and should not be biased against groups of students. The teacher should also focus on gathering evidence from different sources, so when useful, different media of assessment should be employed: written tests, computer-based assessments, portfolios, group projects or oral questioning. How to combine evidence from different (and frequently incompatible) sources of information is another difficult task. On the other hand high-stakes, external assessments may even be harder to develop. They have to follow a very strict development process that involves an amazing list of quality control procedures, field trials, evaluation and improvement of the assessment under construction. The development of high stakes assessments is so complex and costly that it requires the collaboration of a whole array of specialists under the umbrella of dedicated organisations.

A typical assessment construction cycle usually starts with drafting a very detailed document with the aims of the assessment and the intended audience. Then, expert question writers and teachers come together to build the table of specifications, which is a very detailed document explaining what is to be assessed, in order to fulfil the aims of the assessment. The next stage is to start drafting questions according to the table of specifications. The authors of the questions may generate new questions, or may get ideas from past assessments. The draft instrument is then evaluated again (sometimes by an independent group of experts) to confirm that the questions comply with the table of specifications. The next stage includes the piloting of the assessment and data is collected and analysed to identify whether the targeted population perceives it to be a fair and valid instrument. Statistical analysis is used to build the profile of each question or task of the assessment. Questions or tasks of inferior quality will be deleted or replaced or improved. A new cycle of piloting takes place and the assessment is finalised. During this cyclical procedure, the scoring rubrics are also developed, evaluated and finalised. The assessment is ready when a document is prepared explaining how it should be administered and scored and what the valid uses of its results are.

The development of high stakes assessments usually entails much work from psychometricians who use advanced statistical models to evaluate the quality of the assessment and to link the assessment results to those of other assessments. This shows that the development of assessments – either for low stakes or for high stakes purposes – is a very demanding procedure which requires both resources and expertise. There are many steps involved in either a low-stakes classroom assessment or a high-stakes formal assessment process. The next section delves deeper into the diverse nature of assessment.

*Visit **WebResources** for more information on teacher-made assessments and development of high stakes assessments*

THE DIVERSE NATURE OF ASSESSMENT

Assessment can successfully achieve its complex roles in modern society, because it is flexible and has a very diverse nature. It may be regarded as a process varying in frequency, duration, content, form, formality and intention.

Assessment as a process

In thinking about assessment as a process, we were searching for some way in which we could communicate quite simply the key parameters (i.e., the constituent variables or qualities of assessment). Five parameters are set out in Table 3. These are amongst the most observable parameters and there is no reason why they cannot be supplemented by other factors that are relevant to you. No claim is made that this framework is comprehensive but it may offer you a starting point.

How would you use this framework to describe an assessment process? In thinking about an assessment for a group you could indicate, for instance, that there will be two assessment events, one focusing on knowledge and one focusing on skills; the two assessment events may involve questioning (e.g., a mid-term exam) and a project (e.g., a practical assignment at the end of semester) respectively; the grading of the assessments is criterion-referenced using a rating scale provided to the students; the results for both assessment events contribute to the final grade (i.e., summative); they are internally set and internally marked, and graded pass-fail. You are correct in thinking that this may seem a little academic at present. The main purpose, however, is to provide you with a reasonably descriptive and conceptual framework with which you can meaningfully discuss, describe and plan assessments.

Now, if your assessments have been specified for you in advance by the curriculum, syllabus or training plan, then you can sit back and relax because all the thinking has been done for you. (You may not agree with what has been set out for you but that is not an assessment issue.) Let us describe these parameters in a little more detail for you.

Table 3. Key parameters of an assessment process

Parameter	*Dimension*
Number	frequency of assessment events; number of assessments; duration of assessment.
Content	knowledge, skills and/or attitudinal content.
Form	form(s) of assessment; specific method(s) of assessment.
Intention	formative and/or summative; criterion-referenced assessment.
Formality	standardisation; objective or subjective scoring; internal or external assessment.

Assessments may vary in number, frequency and duration

Assessments may involve one, two or more different components or assessment events. As an example, an engineering class has short weekly tests, a major practical assignment lasting all semester and a final 2-hour exam, while a marketing diploma class has one major assignment, a class presentation and an external exam. Some such as a welfare subject have two assignments (essays) and others, like an outreach program have many informal but no formal assessments. The assessment issue here is the amount and frequency of assessment.

In most formal courses, higher educational institutions prescribe a minimum of two assessment events[2], such as an assignment plus an exam or maybe a combination of theory and practical assessments. Moreover, it has been recognised that few courses should allocate more than 10% of the total teaching hours to formal assessments.[3] So if you are teaching a module that involves 36 hours you should allow for around 3-4 hours of formal, summative and face-to-face (i.e., direct) assessment time. This assessment time would not include formative activities such as class exercises that do not contribute to a final grade. The assessment time would not include the time spent by students outside class (but it would include any time spent in class introducing the topic, describing the assignment, dealing with enquiries or explaining procedures). These guidelines may assist you in not allocating too much time for assessment and thereby reducing the time available for instruction, which after all, is your principal function.

As a general rule, the greater the number of assessments you conduct then the greater will be the reliability of your final decision about a student's competence. As mentioned before, reliability refers to the reproducibility and consistency of your results. And a final practical point – as a general rule, the greater the number of assessments you conduct then the greater will be your assessment workload!

Assessments may vary in content

Assessments can be distinguished in terms of their content, especially the extent to which they are assessing knowledge, skills or attitudinal areas of learning. The main focus of assessment and testing in formal education settings is knowledge and skills but there are many courses where attitudinal factors are also assessed (e.g., standard of service in hospitality; depth of rapport in child care; quality of patient contact in nurse aides; sensitivity of client services in welfare; or appropriateness of customer service in sales).

Remember that this is an artificial classification. It is doubtful that human behaviour can really be grouped into three such neat categories but these categories do provide us with a common terminology.

The issue that is at stake here is the relevance of the assessment process for the learning outcomes. This affects the validity or accuracy of your results. If the learning outcomes for your subject or group are mainly attitudinal then clearly the assessment must focus on attitudes and values. In the same way, the content of the assessment methods that you use (e.g., a role play, an observation of someone's performance) also need to reflect the attitudinal content area. (An exception to this rule may be those cases where an indirect assessment might be more economical in the first instance, say as a means of screening out those who really have no chance of passing on a more expensive and time consuming assessment.)

Holistic assessment. One direction in which you may wish to proceed with your assessments is towards holistic assessments of performance. The term 'holistic' derives from holism, a philosophic view that the important factors in nature are entities that cannot be reduced to the sum of their parts. Holistic assessments are ways of integrating the assessment of knowledge, skills and attitudes into one assessment event. In other words, you are asking yourself: 'Is there some way in which I can assess everything I need through the one assessment task or event?' This approach merges the intellectual demands rather than assessing them separately or in a piecemeal fashion. For instance, in vocational education a single project may be used to assess different aspects of a technician's ability to design cooling systems for a furnace. In primary education, a project on reptiles may form part of a portfolio assessment of both language and science.

Holistic approaches to assessment have become popular because they promise economy of effort from the assessment side but they are not easy to design or to grade. (Note that holistic assessments are different from holistic scoring – this is a process in which a result comes from an overall impression of a finished product that is compared to a standard for a task). One form of holistic assessment that may merit discussion at this point is authentic assessment.

Authentic assessment. In authentic assessment learners complete particular assessment tasks as part of their instruction. The tasks are designed to be meaningful, real-life, adult-like and of long-lasting value. They involve multiple skills, diverse knowledge and attitudinal components. Although they tend to involve individual assessment, students may also be involved in completing a product over a period of time in a group. While this might seem oriented to vocational or professional education, there is increasing reference to the use of authentic assessment in primary and

secondary education. An advantage of authentic assessment is that it allows observation of problem solving and can be used to indicate a student's level of functioning across a range of situations.

Assessments may vary in form

How can you ascertain whether a student knows some facts or can perform a skill or how they feel about an issue? In education we are restricted largely to what a student says or does. As mentioned in a previous section, your assessment processes can basically comprise a number of information sources such as observations, skill tests, simulations, oral questions etc.

We have listed the five major forms of assessment that you are likely to encounter in education and training settings. It is the evidence from these forms of assessment that will be used by you to make your judgements. Each form of assessment could comprise a number of methods.

Table 4. Assessment methods and forms of assessment

Method of assessment	*Form of assessment*
assignments	questioning
attendance/participation measures	observation
class questioning	questioning
drawing	skills tests
essays	questioning
exams	questioning
learning contracts	questioning
practical tests	observation, simulations, skills tests
presentations	observation, questioning
projects	questioning, skills tests
reports	questioning
short quizzes	questioning
speed tests	skills tests
take-home exams	questioning

You will find that there are many different methods or types of assessment tasks or events which are used in education and training. In Table 4 we have listed the most common methods of assessment mentioned by teachers in our discussions with them and we have related these to the five major forms of assessment that we outlined earlier.

The most popular methods include (in order): practical work, classroom questioning, class exercises, class tests, end of topic quizzes, exams, projects and attendance/participation measures. The fact that attendance or participation is used for assessment shows the breadth of assessment approaches but it should be reflected in the learning outcomes for the course. Do not be too alarmed about the dominance of questioning in assessment – it reflects the fact that most of our learning outcomes in education and training are cognitive in nature.

Whether or not they should be is not an assessment issue; it is an educational policy and curriculum issue! Also, keep in mind that much of this questioning can today be carried out with very user-friendly and efficient software, so that teachers do not have to engage in questioning all the time or marking written responses to open-ended questions.

The assessment issue that is at stake here with the forms and methods of assessment is the principle that wherever possible the form of assessment should be consistent with the modality of the learning outcomes for the course being taught. If the learning outcomes involve performance then observation, skills tests or even simulation are preferable to questioning. If the learning outcomes are cognitive or knowledge-based, then questioning is the obvious first choice. The aim is to use assessments that make the process as authentic (i.e., real-life) as possible. This enhances the meaningfulness of your assessment tasks.

*Visit **WebResources** for an example of authentic assessment criteria*

We have not said much about the prior evidence form of assessment. This can take many forms including exemption on the basis of previous achievements or the recognition of prior learning. The aim is to avoid unnecessary assessment. You should use assessments to tell you things that you do not already know about a learner (or could not reasonably be expected to know). For instance, if you already knew that people who completed a particular course of training invariably passed an external assessment then there is little point in continuing with the assessment. It does not add anything to your existing knowledge about the learners. If everyone in your class or course passes and is promoted then what was the point of the assessment? Is it just a rite of passage? Wouldn't more formative assessment or additional time for instruction be preferable? Two additional topics that merit attention are those of portfolios and performance-based assessment.

Portfolio. We are not certain whether we should classify a portfolio as a sub-category of prior evidence but it certainly involves a multi-faceted collection of evidence collected during and following instruction. Portfolios have become popular methods of assessment, originating in the creative areas where there has been a tradition of developing a portfolio of products for evaluation.

The use of portfolios has now spread to other areas of educational assessment (e.g., reading). It is viewed as a means of assessment that encourages learning and student participation. It is also viewed as a contrast to formal, written, summative assessments.

A portfolio is a deliberate collection of materials (e.g., student work, journals, teacher notes, audiotapes, videotapes, other evidence) that relate to major learning outcomes. The portfolio is developed actively by the learner and serves as a record of achievement and development. They offer students the freedom to construct their own assessment evidence and enhance critical analysis of what is relevant, appropriate and acceptable. Portfolio development may accompany instruction and offer multiple evidence of achievement.

Portfolios may cover various forms and methods of assessment. They may include a wide range of work that is criterion-referenced but not standardised in the

strict sense of the word (see the later sections in this chapter for a formal description of these terms).

Of course, there are limits on what one might include in a portfolio, so the parameters of standardisation are much broader. Portfolios still serve as form of summative assessment, even though the component parts might have been developed during the course of instruction or had the benefit of instructor input. They may offer a welcome break from other types of assessments and can be constructed in interesting and diverse ways but some student guidance may be needed in the first instance.

Some limitations of portfolios are concerns about how time consuming they are for the student and how well they sample the learning outcomes of a curriculum but this problem can be overcome with guidance. One should also evaluate their merits and worth for a particular individual, subject area and cohort of students.

A substantial argument in support of portfolios centres upon the beneficial consequences for the learner. It supports an approach to teaching that promotes interest and involvement. The portfolio is seen as a way of showing that the learner has been able to achieve across a variety of contexts and is given the freedom to show the extensive nature of his/her thinking and responding.

Furthermore, it is argued that it does not disadvantage students who may not do as well under the controlled conditions of standardised written or other formal assessments. In addition, the contents of the portfolio are meaningful to the learner and considered worthwhile exhibiting and retaining. While it may be thought that portfolios are less efficient in terms of time and cost than alternative assessments we do not think that this is a major disadvantage given their other benefits and the fact many other methods of assessment are onerous for the teacher. Once again, you are reminded that the promised benefits of portfolios need to be evaluated in a fair comparison with other methods so that you can determine whether they are valuable for you and useful in your teaching context.

One of the recent achievements of technology is the evolution of e-Portfolios which are the digital equivalents of the traditional portfolios. You may store examples of your students' achievements (documents, photos, graphics, spreadsheets, web pages, speeches, music, video, three-dimensional models of physical objects). This type of portfolio brings the merits of technology into the equation: teachers and students save space and they do not have to carry around physical objects; nobody needs to worry that the content of the portfolio may be physically damaged, one can recall and compare the portfolios of different students quickly.

Scoring the quality of the portfolio and storing the assessment results on a file for further use is also straightforward. It is also possible to make the work of the students anonymous and it is easy for teachers to organise common projects for their students. Of course, digital portfolios need the availability of infrastructure such as computers, networks and specialised software but these are becoming more and more widely available.

*Visit **WebResources** for more information on Portfolio and e-Portfolio*

The next aspect of assessment that is related to forms and methods of assessment is the area of performance-based assessment. We were unclear about whether to

include it under this category (i.e., forms and methods) or whether to include it under the category of the content of assessment, especially skills assessment. Either way we do not think that much is lost by including it at this juncture.

Performance-based Assessment. The use of the term 'performance-based assessment' in education and training has increased in recent years throughout the world. As a broad concept it is closely linked to the assessment of a practical activity. There are many definitions of performance-based assessment that you will encounter. Performance-based assessment has been used to refer to:

- a general term for an assessment activity in which students construct responses, create products, or perform demonstrations to provide evidence of their knowledge and skills;[4]
- assessment tasks that require students to perform an activity (e.g., laboratory experiments in science) or construct a response. Extended periods of time, ranging from several minutes to several weeks, may be needed to perform a task. Often the tasks are simulations of, or representations of criterion activities valued in their own right. Evaluations of performance depend heavily on professional judgement;[5]
- requiring students to perform hands-on tasks, such as writing an essay or conducting a science experiment. Such assessments are becoming increasingly common as alternatives to multiple-choice, machine-scored tests. Performance-based assessments are also known as authentic assessments;[6]
- a type of testing that calls for demonstration of understanding and skill in applied, procedural, or open-ended settings.[7]

The key features of performance-based assessments typically centre on authentic activities that involve mixing knowledge, skills and attitudes in a realistic context. The advantage of a performance assessment is its realism and in some fields it is essential to be able to demonstrate both knowledge and skills through performance. Some examples demonstrating performance-based assessment are provided in Figure 5.

There are many reasons for advocating performance-based assessments. The most important of these is their relevance to the curriculum. A second reason is their attraction for learners and teachers. They can add interest to instruction. They can be used for both formative and summative assessments. They can be criterion- or standards-referenced. They can test the underlying knowledge as well as the skills involved in a field.

Performance-based assessments, however, are not a panacea for all assessment problems. Most of what has been written about them is largely theoretical and there have been relatively few evaluations of performance-based assessments in practical settings.

Some people claim that performance assessments may suffer from problems of content or predictive validity. The problem of content validity is related to the fact that it is difficult to sample representatively in a single task all the activities required in a curriculum or for workplace performance. Because of time limitations, no curriculum can cover everything and something has to be omitted. Similarly, teachers may need to settle for limited samples of activities to assess performance because assessment is usually a time-consuming activity. It is also difficult to be

certain that correct performance on one task will be sufficient for the correct performance on related tasks or even on the same task in other contexts. Again, you can only hope that the task you have designed is an adequate indicator. In these cases you can only do your best learning from experience and also learning from what has worked for other teachers.

(Upper Level) Middle or High School

(Provide the students with a copy of a speeding ticket that shows how the fine is determined.) Say to students: "How is the fine for speeding in our state determined? Make a graph that shows teenagers in our town how much it will cost them if they are caught speeding. Excellent graphs will be displayed in the Driver's Education classroom."

Secondary School

(At several specified times during the school day, students observe and count, for a set length of time, the number of cars and other vehicles going through an intersection near the school.) Say to students: "The police department is considering a traffic light or a crossing guard at the intersection near your school. Your help is needed to make graphs that show how many vehicles go through that intersection at certain times of the day. Excellent graphs will be sent to the Chief of Police."

Primary School

(In view of the class, place 10 caterpillars in a box. Place a flashlight at one end, while darkening the other by folding over the box top.) Say to students: "Do caterpillars move more to the light or more to the dark? Make a graph that shows how many caterpillars move to the light and how many move to the dark part of the box. Your graphs will be displayed at Open House.

Source: A Teacher's Guide to Performance-Based Learning and Assessment by Educators in Connecticut's Pomperaug Regional School District 15. To reach the material: http://www.ascd.org/readingroom/books/hibbard96book.html#chapter1

Figure 5. Examples of performance-based assessments in school settings.

Although it is claimed that performance-based assessments are less discriminatory, there is the clear possibility that these tasks may even further discriminate against minority and disadvantaged groups. This series of arguments against performance-based assessments are important not only for this method of assessment but also for any forms of assessment.

On the basis of the information available it would appear that there is a good case for using performance-based assessments but that they should be evaluated against a range of criteria.

Assessments may vary in intention: Formative and summative assessment

Two basic purposes of assessment – formative and summative – are mentioned frequently. The importance of these two approaches to assessment is how they affect

your teaching and the way students in your class identify what is important to learn.[8] It may be helpful to define these terms for you because they are used so widely in education. It was Scriven[9], an educational philosopher and evaluation expert, who made the distinction between what he called 'formative' and 'summative' evaluation.

Let us look at formative assessments first of all. Formative assessments are conducted during teaching or instruction with the intention to improve the learning process. The information that you gain from formative assessments may force you to re-think your teaching plans for your group. The power of this assessment intention is that it is done with a view to making on-going changes or to improve learning before it is too late. This is a classic instance of using assessment information for the benefit of the learner.

Tasks that are used for a summative assessment contribute to the ultimate grade. Typically, they occur at the end of instruction and may provide information to the students but also to someone else (e.g., parents, administrators). Mid-semester assessments that contribute to a final mark might be considered as summative in nature (e.g., the assessment tasks which contribute to the school estimate for the Higher School Certificate in the final year). However, if the teacher uses them to evaluate the effectiveness of his/her teaching and to improve his/her teaching plan for the second half of the term then the assessment results are used formatively. Scriven[10] described the distinction between formative and summative evaluation like this: 'When the cook tastes the soup that's formative; when the guests taste the soup, that's summative'.

One final point of clarification – it is not the assessment that is formative or summative but how you intend to use the results that make it so. As mentioned in the example above, mid-term assessment results may be used formatively as well as summatively. Furthermore the same assessment results in other cases (e.g., weekly tests, projects or homework) may be used both in a formative and summative fashion. That is, a result can contribute to the final grade as well as changing your teaching and instruction. An example of formative assessment in science and technology learning is provided in Figure 6.

Objective	*Activities to achieve this*	*Assessment procedure*
Knowledge about…	Inquire, library search, reading	Short answer test about…
Investigate factors...	Follow instructions, devise experiments	Report investigation of...
Awareness of...	Group discussion, debate, role play	Comment on...
Communicate about...	Design handout, poster, drama	Effectiveness of output…

Source: UNESCO[11]

Figure 6. Formative assessment in science and technology learning.

Our observation is that most of our assessment in education and training is summative (that is, we use it for grading) although this should not be the case. Teachers and trainers are hard put to find enough time for formative assessments

(that is, we tend to use the assessment results mainly in a summative way). Without high quality teaching, there is also the issue of the extent to which some students might take such formative assessments seriously.

Probably the best type of teaching that we have observed is when the teaching and the assessment were intertwined. For example in a carpentry class, students were given a free choice of construction as their assigned task (i.e., summative assessment) and worked on it during the entire course. This was the basis for their assessment but also the source of the teaching and instruction that occurred. The teacher moved around the group and she offered suggestions, guidance and mentoring whenever it was required.

Norm-referenced and criterion-referenced assessment. Whenever we compare an individual's performance with a pre-determined standard or we search for results that are directly interpretable in terms of specified performance, we are focusing on a criterion-referenced assessment.[12] This approach to assessment is consistent with mastery learning, competency-based training and outcomes-based assessment because of its practical emphasis on performance.

It contrasts with a norm-referenced approach, which is used to find out how a person performs compared to others. An example of norm-referenced assessment would be when we compare a student to the overall performance of the class. Another common example of norm-referencing is the Universities Admissions Index (a type of tertiary entrance rank or percentile ranking) following the Higher School Certificate in Year 12 and used for university admissions purposes.

Normative results are useful when performance is age-related or when a student's results should be compared with a specific group (e.g., assessment of performance with students versus experienced workers; linguistic performance of non-English speaking students). Any statement indicating above or below average performance is normative and class rankings are also normative in nature.

It is not always possible to tell whether an assessment is criterion-referenced or norm-referenced from the assessment itself. The difference lies in the way the results are used. The criterion-referenced assessment describes performance, whereas the norm-referenced assessment distinguishes amongst individuals. The same results may be used in both a criterion- and norm-referenced fashion.

It was as recently as 1963, that Glaser, an educational psychologist, put forward the notion of criterion-referenced testing. Criterion-referenced assessment described performance in terms of the nature and order of the tasks performed. It represented a fundamental shift in perspective for educational assessment.

The emphasis is on what the person can do rather than on comparisons with others. In criterion-referenced assessment there is a clearly defined domain of learning tasks. 'Standards-referenced testing' or 'standards based assessment' are terms that are now related to criterion-referenced assessment. Do not think that criterion-referenced means that there is a cut-off point or criterion for passing; it means that the assessment is referenced to a criterion (i.e., a specific content area).

Criterion-referenced tests grew out of mastery learning and approaches to learning that were meant to be specific and observable. The criterion that was set for mastery learning varied but it was defined as a success-rate of around 80% or more. Criterion-referenced assessments are used for various purposes, including:

– classification (e.g., placement, screening, certification, selection, recognition of prior learning);
– diagnosis (e.g., to identify the education and training needs);
– instruction or training (e.g., to provide feedback on the learner's current performance and progress, that is formative assessment; or summative assessment which records performance up to or at a point in time);
– self-knowledge;
– program evaluation; and
– research.

Criterion-referenced types of assessments are useful when we want to check whether people have gained sufficient knowledge or skills to go on to the next stage. This is because the content of a criterion-referenced assessment should be designed to match curriculum learning outcomes closely. Questions in these assessments are directly related to what has been taught. Furthermore, the difficulty level of the questions is linked to learning, so that easy tasks are not necessarily omitted. Criterion-referenced assessments are therefore useful for mastery or competency testing not only because they focus on a special domain but also because they describe exactly what a person can do (e.g., can type at 60 words a minute for three minutes with 98% accuracy). Criterion-referenced tests are suited for workplace assessment; for determining the mastery of basic skills; and when grouping learners for instruction.

Assessments may vary in formality

Another way in which your assessments can vary is in terms of how they are conducted. We have used a simple dimension of formal versus informal to describe assessment. Formal assessment is aimed at obtaining information in a public, structured and prescribed manner. Frequently, formal assessments are also of a high stakes nature. Informal assessment, on the other hand, can be unobtrusive, less structured and private; its results tend to be used formatively and diagnostically. Most usually, the results of informal assessment are of low stakes.

The assessments that involve public end-of year examinations are highly structured and clearly formal in nature; so are the national basic skills tests used in primary schools. Other assessments, like a centrally set but locally marked class project in surveying would be formal, public, summative and quasi-structured. Similarly, a welfare teacher might be visiting a student on a locally set and locally marked work-experience placement in order to make some observations that contribute to his/her final grading of this student's ethical behaviour. Do not be misled by the informality of the situation, as this is still a public albeit formal assessment. In another context, a teacher may use questions directed to every member of the class in order to obtain a general idea about the level of knowledge in the group. This is an informal evaluation rather than a formal assessment. When these questions are carefully directed and varied in difficulty, they can provide a barometer of classroom achievement.

The assessment issue addressed by the formal versus informal distinction relates mainly to the conduct of the assessment. Certain assessments (e.g., high stakes exams) will demand a degree of formality and openness to scrutiny. Public assessments are accountable and have substantial consequences for stakeholders (students, parents, teachers, schools, education systems). Without wishing to tire you with all this background detail we would like to mention some further parameters of formal assessment. The first of these relates to standardisation.

Standardisation. A key feature of an assessment task or event is the extent to which it is standardised. Standardised means the extent to which the procedure is uniform in its administration, answering and scoring. Failure to use standardisation means that results are not comparable across individuals or situations.

We have always considered that standardisation could well be the chief attraction of assessment for the community. Rightly or wrongly, the standardised assessment is perceived by the layperson as intrinsically fair and equitable because it is meant to be uniform from one teacher and situation to another.

An example of a standardised assessment is one of the trade recognition tests in building that uses the same task to determine the value of an overseas qualification. A diagnostic reading test administered by an infant's teacher who is required to read instructions word for word from a manual would also be a standardised assessment. On the other hand, the use of a uniform assessment conditions may not be as important where a teacher initially suspects a learning difficulty or thinks that a student may be gifted in a particular way. An ad hoc assessment might be developed in the first instance before making a referral to a specialist service.

Standardisation can vary. A multiple-choice examination is uniform in the questions it asks and in its scoring; a take-home examination is uniform in questioning, varied in completion but may have an explicit marking guide; a design brief (i.e., an assignment or project for a class in design) is uniform in the task it sets for the students but the completion of the assignment can occur under varying conditions and the marking criteria may vary from one panel member to another.

The degree of standardisation in your assessment may be dictated to you by the syllabus or by the need to make comparisons between students on the same basis. The less emphasis that is placed on examinations and the more emphasis that you give to learning, then the greater will be the scope for varied and more personally designed situations. While standardisation is important in psychological testing, it may be sacrificed in educational settings for the benefit of the student. For instance, assignments, projects and classroom tests might be varied to suit the interests or rate of learning of students. Usually, however, standardised tests undergo thorough statistical scrutiny and various measures regarding the test (such as the average score or the spread of scores for a specific population) are published. Using these statistics, one might use the results from standardised assessments for norm-referenced purposes.

Individual versus group assessment. A second way in which your assessment can vary is whether it is an individual or group assessment. There are some forms of assessment that are restricted largely to personal testing (e.g., clinical skills, trade skills). Reading tests taken on a one-to-one basis and oral examinations are examples of individual assessment. The advantage of individual assessment is that

it gives you an opportunity to observe the person in action and to come into contact with the student as a person; whereas performance in a group assessment is less direct and may be limited to responses on a sheet of paper. The obvious disadvantage of individual assessment is that of additional time and effort, so teachers may need to strive for some balance between these two approaches.

Timed versus un-timed tests. A third way in which your assessment can vary is in terms of time limits. Some tasks are timed or speeded, whereas others are un-timed or power tests. Power tests have ample time conditions. In some tasks (e.g., keyboarding) speed of response is important and it is legitimate to measure the time taken. Furthermore, speed of response is a powerful indicator of the extent to which skilled performance has become completely automatic and time taken to learn is also considered a useful indicator of aptitude for a subject.

Time limits are recommended where speed is a factor in the occupational performance; otherwise generous time allowances should be provided for students in order to ensure that you are assessing only ability in your subject and not a mixture of ability and time-pressured performance. While some of your students are able to cope with the demands of time limits others may find it difficult to demonstrate their best ability when they are rushed and for others performance under fixed conditions with strict time limits can still be anxiety-provoking.

If assessments have prescribed time limits then you should help students to prepare through frequent practice. If you are required to set a time limit, then design your assessments so that around 95% of people can complete them in the time assigned.

Practical implications

Well now that you know more about the variable roles and nature of assessment, you might well ask, ‘So what?’ The real advantage of this framework is that you can now describe the assessment for your group or for the curriculum that you teach. The framework can be used as a mental checklist for you to plan and record the outline of your assessment at the commencement of every semester or when you wish to revise the assessment for a program.

In particular, if you are a college teacher, you should ensure that each person receives a copy of the course outline or a written copy of subject outlines or assessment documents. In some educational systems, a college student is required to sign these documents. It is not clear to us that students always understand what they are signing so it is important that you explain the details to students. Ensure that your assessment is in conformity with the subject or course outline. Do not hesitate to explain to your students as much information as necessary in order to understand all the practicalities regarding assessment, such as details like critical deadlines.

Appeals against assessments are now commonplace – especially in high stakes education – and it is becoming increasingly difficult for teachers to defend themselves. There are fewer problems, however, when assessment processes and details are documented. This is a consequence of the modern overemphasis on

competition and using education to gain vocational or academic qualifications. Secondly, this level of formality is a by-product of the bureaucracy associated with commercial training, mass schooling and further education systems. Nevertheless, within the scope of the assessment requirements for your subject, there should always be considerable leeway for teacher sensitivity to cases of genuine hardship, disadvantage, serious misadventure and difficulty. In particular, special considerations should be given to persons with disabilities.

Remember that the subject outlines are essentially a guideline within which you operate. They indicate the parameters and are there for the benefit of the learner. Very often when a course is actually taught it may bear only minor resemblance to the content specified because: (a) it has never been taught before, or (b) it is out-of-date, or (c) the person who wrote the course is not teaching it, or (d) the teacher is experienced enough to provide additional material, or (e) the nature of the learners and their needs are not satisfied by the subject as it is written. So, provided your learners are not disadvantaged you should feel free to be flexible and to lighten assessment loads when you deem it appropriate – for a start it does wonders for your popularity amongst learners.

You need to make these professional assessment decisions in the light of what is going on around you (in other locations, with other colleagues, in other organisations, in your profession). This needs to be balanced against your responsibility to the learners and the need to keep faith with the essence of the outline. If in doubt, stick to the subject outline. Certainly, you cannot increase assessment loads once an outline of the assessment has been distributed. In some high stake courses there will be little scope for deviation from the curriculum and the expert teacher has an obligation to deliver the content and assessment in accordance with the prescribed requirements.

A FINAL WORD OF CAUTION: THE ‘ASSESSMENT FOR LEARNING’ VERSUS ‘ASSESSMENT OF LEARNING’ DEBATE

There is a number of teachers and academics who believe that tests cause increased anxiety to the students and may have negative effects on the learning process. They also claim that tests make students infer that the sole aim of the teaching is simply to perform well on the assessment. They avoid tests altogether and they often promote ‘assessment for learning’ in the sense that alternative (other than tests) assessments may be a vehicle and a motivation to promote learning, not only to measure it (assessment of learning). It is possible that they may miss the point: assessment (and especially testing) does not have any inherently evil properties. It is rather how people use – and frequently misuse – assessments that may sometimes cause a negative feeling to students and parents.

Having said that, it is also true that different applications of assessment could enrich the assessment toolbox of teachers. For example, self-assessment has long been known to have beneficial effects on learning and on the internal motivation of students. Provided it is well focused and designed, self-assessment can involve students in the teaching-assessment-teaching cycle in a very constructive way.

Particularly when connected/attached to projects and portfolios, self-assessment can have especially beneficial and positive results.

The idea of self-assessment is not new but it is not frequently used for several reasons. Firstly, it is usually claimed that self-assessment is a time-consuming process which shrinks the actual teaching time. It is also claimed that it is a rather complex process which demands cooperation (and maturity) on behalf of the students. The advocates of self-assessment, on the other hand, suggest that self-assessment is not much more time-consuming when it is designed properly. They also claim that the students will need some time in the beginning to get used to the new method of assessment but the positive results will be more valuable than the initial investment in time. It is usually better, they claim, to reduce the time of paper-and-pencil testing and increase the time of this type of alternative assessment.

One of the secrets of the success of self-assessment is that the students are continuously exposed to a flow of feedback about their work when they have the scoring criteria handy. Feedback is known to have very positive results on learning when it is focused, detailed, clear and on-time. The same could also happen with peer-assessment. As a method, peer-assessment means that other students may use scoring rubrics to express their opinion (i.e., assess and evaluate) the work of their classmates. Peer-assessment has been tried in various contexts with very positive results, especially in tertiary education.

Although self-assessment and peer-assessment may have very good results in one context, they may have very undesirable consequences in other settings. At the first stages of your career you may prefer to use more traditional assessment methods until you get more experience and become more confident. If, however, you feel that you would like to try self-assessment and peer-assessment methods in your class, be warned that you will need to be very well organised to succeed. In any case, do not feel that you need to be involved in the assessment for learning versus assessment of learning debate. You now know that you need to design your assessment approach according to the needs of your students. Over-reliance on one single approach is not wise and this chapter has given you a broad selection of approaches from which to choose.

SUMMARY

We have covered a considerable amount of detail in this chapter and you can be reassured that you do not need to recall it all. If we had to say what was most important about the role and nature of assessment then we would focus upon (a) the extent to which the results of assessment are used in a formative or summative way; (b) whether assessment is criterion-referenced in its interpretation; and (c) the extent to which it is standardised.

A case has been made for using formative, holistic and authentic assessments that are situated in reality. They foster the application of knowledge and skills and bring about important changes in your teaching. Studying and learning in your area will be affected by the forms and methods of assessment that are used. More importantly, studying, learning and assessment in your area will be affected by the

approaches to instruction that are commonly used. These may depend upon the traditions in your field.

A minimum of two assessments per subject and no more than 10% of teaching time for assessments has been suggested. It has been recommended that students should be given at least two week's notice of assessments and that teachers should provide learners with course outlines and assessment details early in the semester.

This chapter has shown you that there are many approaches to assessment. Assessment is not a unitary concept but is composed of many dimensions so it is important to be clear about the type of assessment we are discussing.

Sometimes one form or method of assessment is contrasted with another but these comparisons are not always valid. You need to compare two assessment methods that are equivalent in terms of time, effort, teaching implications, predictive validity, consistency of results, learning consequences, economy and so forth. Sometimes the limitations of a particular assessment may be a function of the developer and not an inherent fault of the method itself. (For example, one may develop an excellent authentic assessment or one may develop a less than satisfactory authentic assessment; in the same way one might have an excellent assessment approach that is implemented in a less than optimal manner or one can have a poorly developed assessment that is interpreted and implemented in a positive manner. There are so many permutations and combinations that comparisons are fraught with difficulty.)

There is considerable hype concerning the benefits of some types of assessment; our view is that all the forms and methods of assessment have their place and utility. The issues here can be evaluated for your context and circumstances.

Maybe the framework and the parameters we have outlined will ensure that we can be clearer in our discussions about assessment. In the next chapter we shall focus specifically on planning the assessment. Now it is time to take a break. When you are ready you might look at the true-false questions to review your reading of the chapter and maybe undertake some of the exercises individually or as a group.

-oOo-

REVIEW QUESTIONS

Try these review questions to help you reinforce some of the key ideas in this section. These are all true-false questions to make it easier and quicker for you to complete. Think whether each statement is mainly true or false. Then just circle the T (True) or F (False). If you are not sure, just guess.

T	F	Assessment is a comprehensive generic term
T	F	Assessments may vary in number, frequency and duration
T	F	At least three assessment events are recommended for each subject, unit or module
T	F	As a general rule, the greater the number of assessments you conduct then the higher will be the reliability of your results

T	F	Human behaviour can be grouped into the categories of knowledge, skills or assessments
T	F	Holistic assessment integrates the assessment of knowledge, skills and attitudes
T	F	Holistic assessment involves holistic scoring
T	F	The five major forms of assessment are: observation, simulations, skills tests, questioning and the use of prior evidence
T	F	Summative assessments seek to improve the learning process
T	F	It is how the results will be used that makes an assessment summative or formative
T	F	Classroom questioning is a formative public assessment
T	F	Teachers should provide students with course outlines and assessment details early in the semester
T	F	Around 15% of a subject's teaching time should be given over to assessments
T	F	At least two week's notice should be given to students for a class test
T	F	Teachers can increase the assessment load slightly even though an outline of the assessment has been distributed
T	F	Teachers have an obligation to deliver content and assessment in accordance with the prescribed requirements
T	F	A standardised assessment has instructions for administration and scoring
T	F	A group test can be used to observe individual performance
T	F	A criterion-referenced test can be distinguished by its format and questions
T	F	A keyboarding test is likely to be a power test
T	F	A formal essay examination is likely to be an objective test.
T	F	Competency-based assessments are criterion-referenced
T	F	Norm-referenced assessments are designed to give descriptions of performance

EXERCISES

Here are some review exercises for you to answer or they can be used as the basis for discussion.

1. Give three examples of formative and summative assessment from your teaching context.

2. What are the assessment requirements for students/trainees in the subject or course that you teach or plan to teach?

3. Describe a form of holistic assessment that could be used in your teaching area.

4. Identify three distinctions between (a) criterion-referenced and (b) norm-referenced approaches to testing and assessment.

5. Read the following definitions of criterion-referenced assessment and comment critically on these views of criterion-referenced testing.

> A. Criterion-referenced tests: These tests measure what a student knows and can do in relation to specific objectives or criteria that all students are expected to meet. The tests are designed to reflect the knowledge and skills that a state or community has identified as important for an academic subject. The focus is on whether the student can meet the criteria and not on how he or she performs relative to others. (Source: Education Week, http://www.edweek.org/sreports/qc97/misc/#cr)
>
> B. Criterion-referenced test: A test in which scores are evaluated, not in terms of comparative rankings, but rather, in terms of the percentage of mastery of a predetermined standard. Examples include behind-the-wheel driving tests, tests of typing speed and accuracy, tests in the military for strength, and tests measuring the effects of alcohol on muscular coordination. Most manual skills tests are criterion-referenced. Criterion-referenced tests tend to focus on minimum thresholds, such as the threshold needed to pass a driving test, or to pass for secretarial or military service — or, in the new education system, to graduate from high school. The test focuses — not on the best, the median, or the average students — but on the worst students, those near the minimum threshold. The new education system mandates the use of criterion-referenced tests (and the de-emphasis or elimination of the traditional norm-referenced tests such as ACT, SAT, and Iowa Basic tests), thereby redefining 'success'. The new system intends to 'hold schools and teachers accountable' (by various threats and punishments from the government) for failure to meet its peculiar measure of success. By this means the system compels teachers to forsake students who are average or better, and focus instead on those students near the minimum threshold, for that is how teachers and schools are to be judged. This furthers the twin goals of: (1) educating mostly just for minimum competencies in specific job skills, and (2) 'equalizing' educational outcomes (not educational opportunities) — while turning a blind eye to the development and recognition of academic excellence and the broad-based knowledge needed to keep people free. (Source: Maple River Education Coalition, http://www.mredcopac.org/glossary.htm)

6. Indicate the characteristics of assessments in the subject area in which you teach or plan to teach, in terms of the characteristics used to describe assessments.

7. What is your opinion of the assessment processes used in your area of education or training? How could they be improved?

8. Read this brief excerpt from the American Psychological Association's Monitor (October 1999. pp. 10). In your opinion, does frequent assessment lead to greater achievement?

FREQUENT TESTING MEANS BETTER GRADES, STUDIES FIND

College students who were given a quiz on reading material every week outperformed students who were given comparable homework or who had neither … students taking "spot-quizzes" were compared to students of comparable aptitude who were assigned homework on the same material, and to those who neither took spot-quizzes nor completed homework assignments. On final achievement tests, the spot-quiz group outperformed the homework group by 16 percent and the control group by 24 percent.

S. Kass