Assessment

Moving *into* a New Future

Muhammad Naseer Ud-Din
Norman Reid
Ikram Ali Malik

Following a critical survey of the research literature on assessment, practical ways forward for national certification are outlined, the aim being to offer development and enhancement for the 21st century

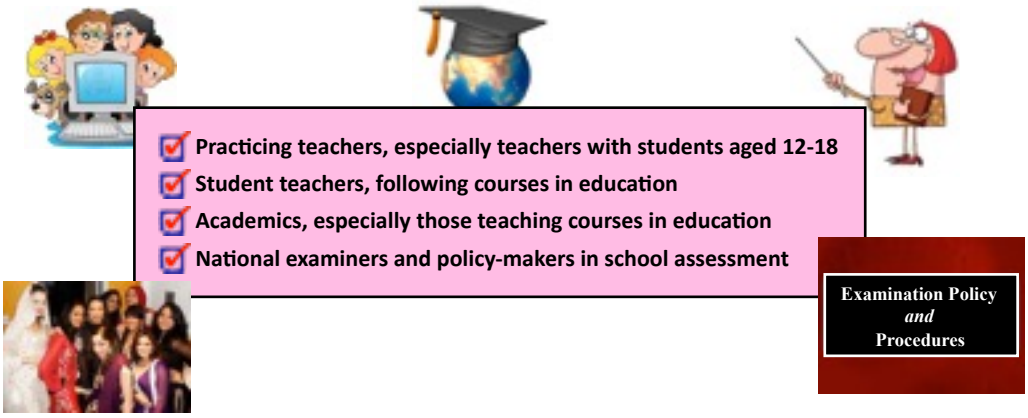
Preface

There have been enormous changes in what is taught in schools and universities throughout the world in recent decades as well as enormous changes in how out students are being taught. The pace of change has been accelerated with the growing impact of modern technologies which make knowledge on almost every conceivable subject available at the press of a computer key.

At the same time, despite considerable research on assessment, changes in assessment have lagged far behind. Inevitably, what is credited in formal tests and examinations determines what students and teacher see as priorities for learning. In this way, curriculum developments are being held back by inadequacies in assessment systems.

This monograph brings together numerous findings from carefully conducted research to provide a picture of ways forward for assessment. On the basis of this, the way assessment can be developed in the context of examinations set by national examination boards is outlined. Throughout, the aim is to base everything on clear research evidence but, at the same time, to bring the findings into the practical arena of national assessments covering thousands of schools.

This monograph has been compiled primarily for the Federal Board of Education, Islamabad, Pakistan, to provide a basis for future training and developments. The monograph is based on an original text developed by Naseer Ud-din and a set of workshops delivered by Norman Reid. However, many other audiences may find the text useful:



The monograph includes only the key references to the research literature and is unashamedly academic. However, the style is designed to be user-friendly, with extensive use of diagrams, pictures and even cartoons. The presentation is based on the simple belief that all learning, even highly academic learning, should be fun and should be challenging to the reader.

We hope you enjoy the monograph and find it instructive.

Naseer Ud-Din
Norman Reid
Ikram Ali Malik

The Authors

Professor Ikram Ali Malik, Professor and Chairman
Federal Board of Intermediate & Secondary Education, Islamabad

Dr Muhammad Naseer Ud Din
Assistant Professor, Instituted of Education and Research - Kohat University of Science and Technology

Professor Norman Reid, Emeritus Professor
University of Glasgow, Scotland

Acknowledgements

Naseer Ud-din expresses his appreciation to the Higher Education Commission, Pakistan, for the funding support to undertake a year's post-doctoral study at the University of Glasgow, Scotland. Gratitude is also expressed to the Vice-Chancellor of Kohat University of Science and Technology for granting a year's leave to undertake these studies. During this time, a monograph collating research evidence on assessment was developed. This took advantage of having access to the library of the University of Glasgow, one of the finest libraries in the world, as well as drawing from the expertise of the Scottish Qualifications Authority.

The monograph was developed in collaboration with Professor Norman Reid, Emeritus Professor of Science Education at Glasgow, and, in 2016, Professor Reid led five days of workshops for Masters Trainers in the Federal Board, Islamabad, under the supervision of Professor Ikram Malik. The present monograph draws together the original monograph and the material used in the workshops. It seeks to apply the principles into the practical situations of national examination and certification.

Grateful thanks are expressed to the following for their valuable support and encouragement:

Mr. Muhammad Sarwar
Secretary
Federal Board of Intermediate and Secondary Education, Islamabad

Mr. Mirza Ali
Deputy Secretary
Federal Board of Intermediate and Secondary Education, Islamabad

Dr. Iqbal Shah
Assistant Professor, Department of Science Education
Allama Iqbal Open University Islamabad

Dr. Faridullah Khan
Assistant Professor, Institute of Education and Research
Kohat University of Science and Technology

Dr. Muhammad Imran Yousaf
Director, Department of Continuing Education, Home Economics and Women Development
Pir Mehr Ali Shah, Arid Agriculture Rawalpindi

Mr Muhammad Ishfaq Ud Din
Advocate High Court
District Bar Association Rawalpindi

September 2016

Contents

Assessment Moving *into* a New Future

Why Assessment?

Aims

This chapter looks to outline the concept of assessment in general terms so that you can gain an understanding of key terms and ideas which shape the way we think about educational assessment

Introduction

Assessment has grown to become a huge 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 every day life in primary, secondary or tertiary education as well as in industrial or commercial training. Assessment, therefore, has been given a powerful role and plays a key part in modern education. It has an impact on career directions and opportunities as well as opening or closing doors in further education and training.

Of course, teachers need to be skilled and informed in this area. Sadly, very often it has not held a high profile in teacher-preparation or trainer-preparation courses. Thus, many teachers and trainers are still left to acquire their assessment expertise through in-services courses, the assistance of colleagues or simply by trial-and-error. Given the power of assessment to determine the futures of our learners, it is vital that the whole area of assessment is explored. There are exciting developments taking place throughout the world and these need to be incorporated into normal practice at all levels in Pakistan.

Educational Assessment

Educational assessment refers to the process of gathering accurate and reliable information on the knowledge, understandings, skills of learners and how well they can apply their understandings in life. *Assessment can focus on the individual learner, a group of learners, a learning institution such as a school, or the educational system as a whole.* Assessment is a form of educational research since it is concerned with obtaining reliable, verifiable information on the learning achievements of those being assessed.

The process is summarised in Figure 1.1.

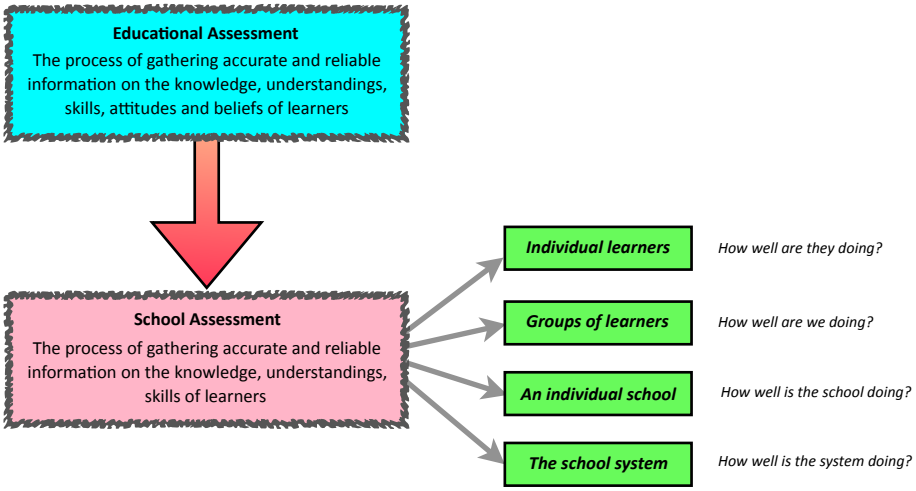


Figure 1.1 The Scope of Educational Assessment

Assessment for You as Teacher

It is helpful for you to be knowledgeable about educational assessments because the results may have significant impact on other people's lives: selection, certification, diagnosis, special instruction or placement. Furthermore tests, exams, quizzes, projects, assignments or portfolios are an integral part of teaching and it is valuable for you to have some knowledge of their development and use. That 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.

It has been estimated that teachers spend as much as one-third of their time in assessment. Indeed, where teachers are trusted in making their own assessment (with a reduction of formal external examinations), the load and responsibility will increase. This is a worldwide trend. Consciously or otherwise, assessments play a large role in your actual teaching. For instance, your assessments may occur informally during evaluation of your instruction and can act as a catalyst guiding you to offer further help for the learner; or, at the other extreme, a formal assessment at the end of a course may determine the career prospects of a learner. Indeed, the way assessment will take place can strongly influence what you emphasise in your teaching and the way you encourage specific aspects of student learning. This means that you need to be informed about the effects of different forms and methods of assessment. Whatever happens, assessment will strongly influence the way you teach and the way your students learn.

Assessment can be used in many ways. For example,

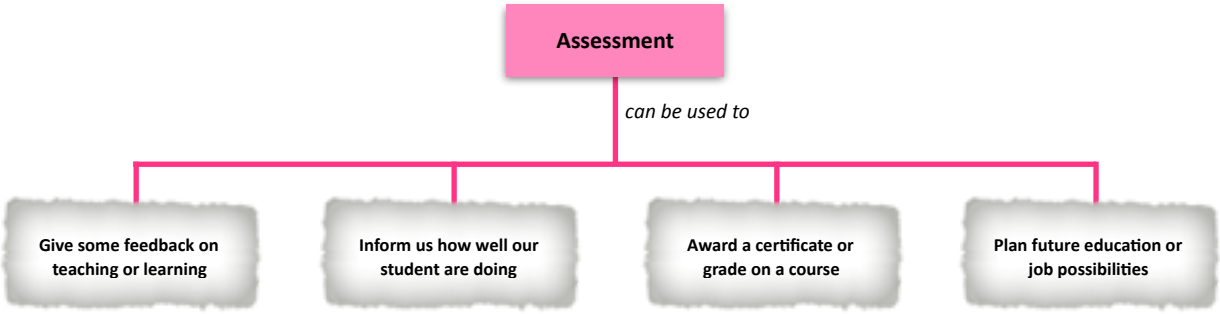


Figure 1.2 Some Use for Assessment Data

There are two key words that are important: quantity and quality:

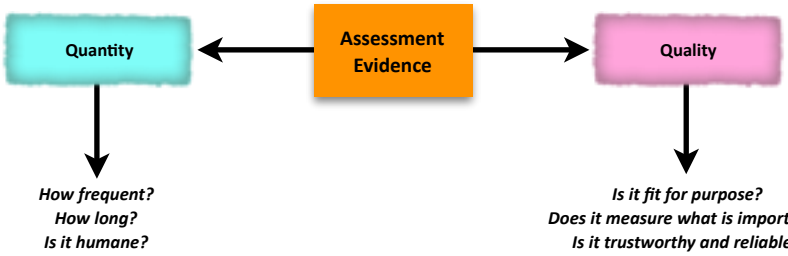


Figure 1.3 Quality and Quantity

We need to think carefully about the quality. Are we measuring what we want to measure? Are we measuring important outcomes from learning? Is our assessment fair for our students? Does it allow them to show their full range of skills and abilities? Questions about how long assessments take and how often we assess can then be considered.

New Developments in Assessment

Recent decades have seen significant developments in the field of educational assessment. New approaches to the assessment of student learning achievement have been accompanied by the increasing prominence of educational assessments as a policy issue. In particular, there has been a growth in modes of assessment that promote, as well as measure, standards and quality. These developments have profound implications for individual learners, institutions and educational systems.

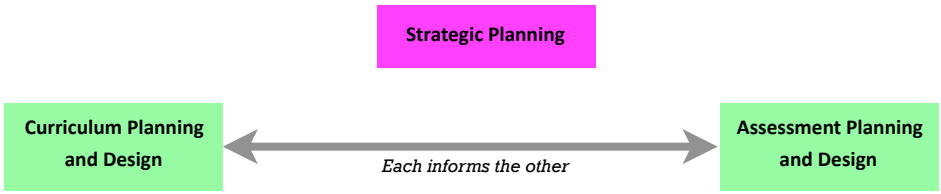
In addition, new ways to assess have been developed. Some have proved to be very powerful and reliable in assessing understanding while others have been useful in measuring skills. Most of the assessment we experienced in our educational careers simply measured our ability to recall information or procedures. It is essential that we move forward from this to measure a much wider range of skills. In the days of instant information at the press of a computer key, the place of recall is now of diminishing importance. In a sense, it is how the learners can use their knowledge and understand that is much more important. That means assessment MUST change to reflect this.

Faculty members, instructional staff, and assessment professionals are often responsible for specifying student outcomes for academic programmes, designing the curricula on which the programmes are based, delivering the curriculum, and determining the quality of these learning experiences. Assessment plays a key role in all of this.

Assessment information can help us to see where our programmes are working well. It can also help us to see how well our teaching strategies are working. However, all of this will only happen if assessment procedures are built into programme planning from the outset. In addition ….

Assessment
Can only assist in enhancing future education provision
if
Those with the power to take the decisions on future planning
understand
The assessment findings
and
Are willing to make the necessary changes

It is, therefore, essential to plan and refine curriculum design and assessment together. Each informs the other. This may need radical changes in thinking. Too often assessment has become detached and is seen as a separate activity, to be conducted at the end of a course but never used to inform those who design the curriculum.



Why Assessment?

Assessment is an integral part of all learning. Ideally, assessment should enhance learning. It should inform the learner and it should inform the teacher. While it can direct and support future learning, assessment is also important for the purposes of marking and grading. Our students need to know how well they have done and they need that information to lead to future learning opportunities or the qualifications to enter the workplace.

Danili and Reid (2005)[1] noted that assessments play an important role in the teaching and learning process, and for specific uses. For individuals, assessments (particularly public examinations) profoundly affect life chances, not just in the first years after leaving school, but many years later. As Boud (1995, p. 35)[2] stated,

> *'The effects of bad practice are far more potent than they are for any aspect of teaching. Students can, with difficulty, escape from the effects of poor teaching, they cannot (by definition, if they want to graduate) escape the effects of poor assessment. Assessment acts as a mechanism to control students that is far more pervasive and insidious than most staff would be prepared to acknowledge.'*

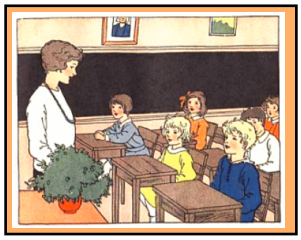
Assessment is an integral part of our daily lives. Every time we have to make a decision, we have to, for example, assess, value, judge, estimate, appraise, the situation first before we can go any further.

Assessment means measurement

[1] Danili E. and Reid N., (2005) Assessment formats: do they make a difference? *Chemistry Education Research and Practice*, 6, 204-212.
[2] Boud, D. (1995). *Assessment and learning: Contradictory or complementary?* In P. Knight (Ed.), Assessment for Learning in Higher Education, London: Kogan Page.35.

What is Assessment ?

Huba and Freed (2000)[3] describe assessment as:



'..... the process of gathering and discussing information from multiple and diverse sources in order to develop a deep understanding of what students know, understand, and can do with their knowledge as result of their educational experiences; the process culminates when assessment results are used to improve subsequent learning.'

Figure 1.4 Assessment means Looking at Learning

One of the most basic and difficult tasks that teachers face in their work is the process of assessment. Classroom assessment includes all the process involved in making decisions about students learning progress. It may involve observation of what learners can do, it may involve listening to what they can tell us, it may involve reading what they have written.

Assessment is the on-going process of gathering, analysing, and interpreting evidence so that we can make judgements about how well our students have learned. In addition, reflection on that evidence may be able to guide us to to make informed and consistent judgments to improve future student learning.

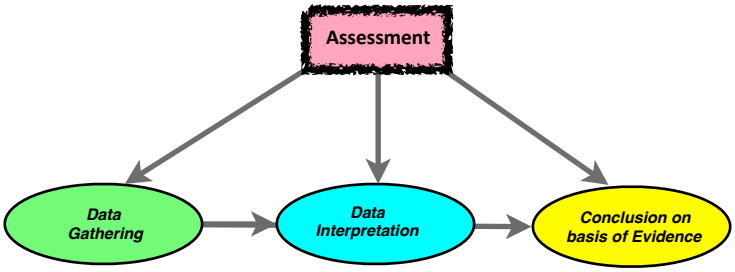
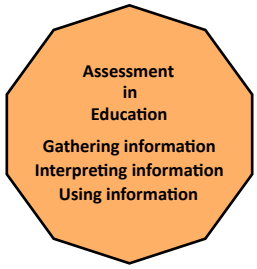


Figure 1.5 Concept of Assessment

Assessment is the means of obtaining information which allows teachers, learners and parents to make professional judgments about the learner's progress. Assessment will improve the quality of learning and teaching if data for information gathered has a clear purpose. The data must be seen as clear evidence on student achievement. The data also has to be collected systematically and used appropriately. The teacher can use a wide variety of approaches to obtain information about learner's success, achievement and progress.



[3] Huba, M. E. and Freed, J. E. (2000) *Learner centred assessment on college campuses – shifting the focus from teaching to learning*. MA: Allyn and Bacon.

Who Uses Assessment Information?

Five main groups may be involved here:

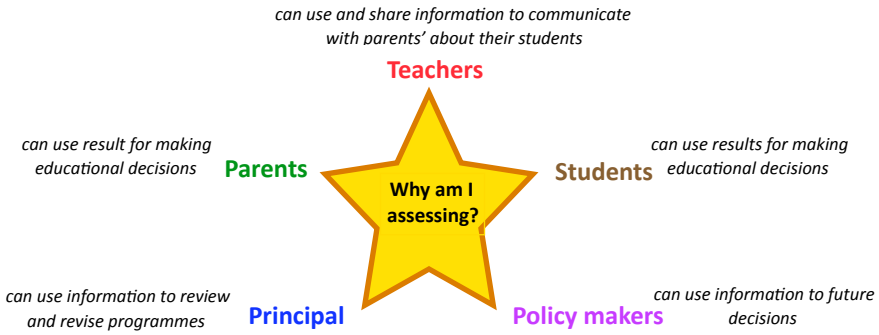


Figure 1.6 Who Uses Assessment Information

There is a very important principle underpinning the figure 1.5. Assessment is designed for a purpose. If assessment designed for one purpose is then used for another, the outcomes may be very misleading. Thus, assessment designed to pinpoint the strengths and weaknesses of learners cannot be used as a means of evaluating a course. Similarly, assessment of student performance in a course CANNOT be used to evaluate the strengths and weakness of the teachers or the school. The outcomes may be highly misleading.

Important Principle
If assessment designed for one purpose is then used for another, the outcomes may be very misleading

As teachers, lecturers and facilitators of learning we tend to think that the most important thing we do for our students is to assess their work. We assess our students in examinations at the end of the course. The outcomes may determine the award of diplomas, degrees and, indeed, the future careers of our students. However, there are major problems if we think this way.

Are we really *confident* that the examinations we set are valid measures of the learning achievements of our students? What if a student simply had a '*bad day*'? Can we assess important outcomes like critical thinking, creative thinking, scientific thinking, ability to solve problems in this kind of '*one-off*' examination? Can examinations test the ability of our students to *apply* what they have learned in practical and useful ways?

The questions are endless and reveal that traditional approaches to assessment are highly flawed.

Here are some important points that we must consider:

- Assessment is very time consuming and potentially very dangerous: we may take hours to set and mark examination papers and yet the outcomes may be a poor reflection of the achievements of our students.
- Assessment means we have to deduce what is happening in the human brain: the important thing is that this kind of deduction is not very reliable.
- Assessment can distort the whole process of learning: we only will teach, and our students will only learn, what is likely to bring credit in examinations.
- Assessment is often the major driving force which gets students down to serious studying: yet they are often simply cramming their heads full of information which will largely forgotten within a few weeks.
- We assess students to find how what they can do, to find out how well we have taught, to allow them to move on, to certificate them, to provide information for employers: no assessment can do all this at the same time.
- We live in a society where people are appointed and employed on the basis of their qualifications: it is essential that we get assessment as accurate as possible.
- Students themselves need feedback to help them to find out how their learning is going: this is perhaps the most important aspect of assessment - students need quality feedback.
- We need feedback on how well student learning is going, so that we can adjust and develop our teaching: we cannot assume that we have '*got it right*' - we need to challenge ourselves continually.
- Changing what we do in the light of new challenges requires thinking clearly about what assessment is for and ensuring that it does not inadvertently undermine our educational priorities.

Assessment is a complicated business !!

Figure 1.7 Untying the knot is not easy

The aim of this monograph is to truth to unravel all the complexities in the assessment process so that you can move forward to develop assessment tasks for your students that will lift education forward into the future.

Why Assess?

Why do we assess learners and their learning outcomes in educational institutions, schools and universities? This may seem like a redundant question, but it is important for all teaching staff to reflect on the purpose of assessment.

Assessment is the process of evaluating an individual's learning. It involves generating and collecting evidence of a learner's attainment of knowledge and skills and judging that evidence against defined standards.

Broadfoot and Black (2004)[4] proposed a model for assessment that encourages learning and provides feedback on learning and teaching to both the learner and teacher.

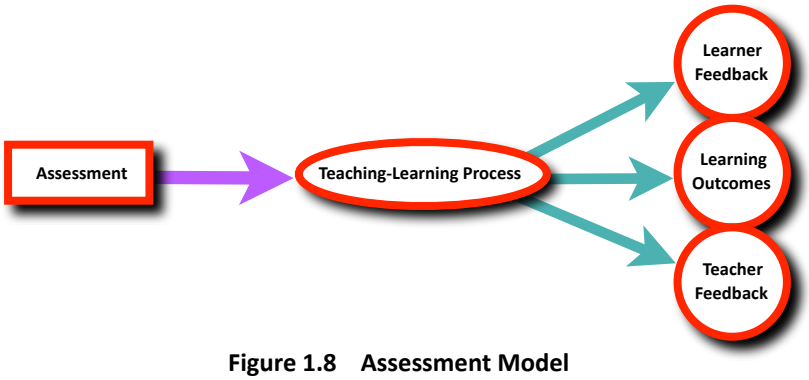


Figure 1.8 Assessment Model
(based Broadfoot and Black 2004)

Assessment tasks determine to a significant extent what learners will learn, and the methods they will employ to retain, reproduce, reconstruct and engage with learn to material (Biggs, 2002)[5]. Learner responses to perceived, or actual, assessment tasks will often dominate motivation to learn. In simple terms, assessment drives what is learned and how it is to be learned.

Assessment for learning is the process of seeking and interpreting evidence of a learner's performance for use by learners and their teachers to identify where the learners are in their learning, where their next learning goals are, and what to do next to achieve them. Assessment is the single most important determinant of student learning. Teachers communicate to students their values, priorities and expectations through assessment.

Thinking of higher education, Ramsden[6] argues that:

> *'Assessment that is the servant rather than the master of the educational process will necessarily be viewed as an integral part of teaching and the practice of improving teaching. A sophisticated theory of teaching leads directly to the proposition that the assessment of students is above all about understanding the processes and outcomes of student learning, and understanding the students who have done the learning. In the application of these understandings, we aim to make both student learning and our teaching better' (p.180).*

Assessment in education is about gathering, interpreting and using information about the processes and outcomes of learning. It can be used in a variety of ways, such as to test and certify achievement (Ikke Junior and Leaving Certificate), to determine the appropriate route for students to take through a curriculum or to identify specific areas of difficulty or strength for a given student. The assessment process is just like a window into the student's mind that gives opportunity for teachers to see what students know and learn (Lorsbach *et al*., 1992)[7]. Moreover, it is a central component and integral part of curriculum (Killington and Mitchell, 1980[8]; Wills, 1993[9]; McGrath and Ingham, 1992[10]).

---

[4] Broadfoot P. and Black P. (2004) Redefining assessment? The first ten years of assessment in education. *Assessment in Education: Principles, Policy and Practice*, 11(1), 7-26.

[5] Biggs J.B. (2002) *Aligning teaching and assessment to curriculum objectives*. LTSN Imaginative Curriculum Guide IC022. [available on line, free, from: jisctechdis.ac.uk].

[6] Ramsden, P. (2003) *Learning to teach in higher education*. London & New York: Routledge Falmer

[7] Lorsbach, A.W., Tobin, K., Brisocove, C. and Ulerick, S. (1992) An interpretation of assessment methods on middle school science, *International Journal of Science Education*, 14(3), 305-317

[8] Killington, S.A. and Mitchell, A.C. (1980) Designing an assessment system Using art activities in the assessment of science in the primary school, *School Science Review*, 73(264), 33-47.

[9] Wills, D. (1993) Learning and assessment: exposing the inconsistencies of theory and practice, *Oxford Review of Education*, 19(3), 383-401.

[10] McGrath, M. and Ingham, A. (1992) Using art activities in the assessment of science in the primary school, *School Science Review*, 73(264), 33-47.

Assessment can be used to certificate. It can be used to diagnose. It can underpin selection. It can lead to gradings. Increasingly, assessments are linked with certification of competence and the validation of performance on job-related tasks.

The common view is that assessment is something done at the end of the course or unit. However, it is integral to the *entire process of teaching and learning* (Burton and Haines, 1997[11]) and involves gathering, recording, interpreting, using, and reporting information about a student's progress and achievement in developing knowledge, skills and attitudes (Lloyd - Jones, 1986[12]; Brooks, 1993[13]).

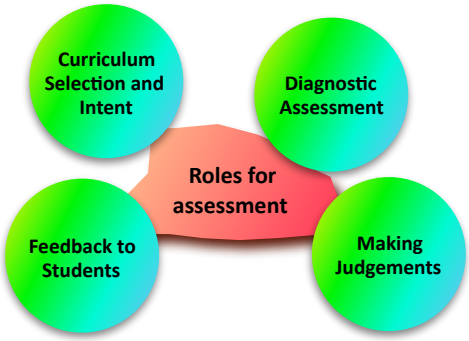


Figure 1.9 Assessment Role

Purposes for assessment

Many groups can be informed be means of data from assessments:

For *students:*

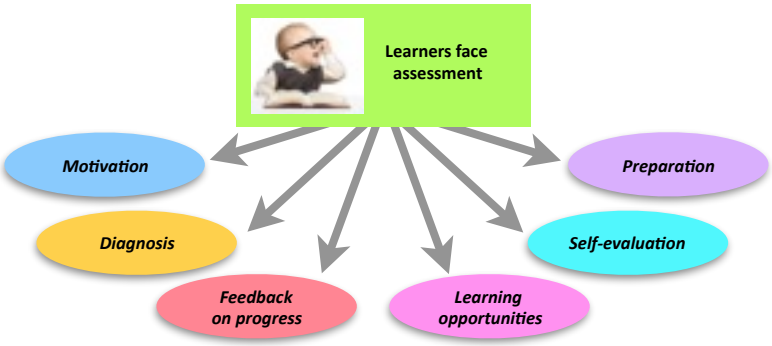


Figure 1.10 The Student and Assessment

[11] Burton, L and Haines, C. (1997) Innovative in teaching and assessing mathematics at university level, *Teaching in Higher Education,* 2(3), 273-293.

[12] Lloyd - Jones, R. (1986) An overview of assessment, in Lloyd - Jones, R., Bray, E., Jonson, G. and Currie, R. (eds.) *Assessment from Principles to Action*, London, Macmilon Education Lts., 1-13.

[13] Brooks, V. (1993) The resurgence of external examine in Britain: a historical review, *British Journal of Educational Studies,* 4(1), 59-72.

This can be expanded:

- ***Diagnostic*** - to enable students to find out their level of competency/knowledge/understanding at the beginning of a course.
- ***Feedback*** **-** for students to ascertain their progress in relation to the learning outcomes of a course.
- ***Learning opportunities*** - to provide students with the opportunities to develop their mastery of ideas and practise skills and competencies through articulating them in writing or /oral work.
- ***Self-evaluation*** - to encourage students to make judgments about the quality of their own work.
- ***Motivation*** **-** assessment tasks can enhance student motivation by providing frameworks for developing, reviewing or extending their understanding (for example, in a piece of research or a collaborative investigation).
- ***Preparation for longer term learning*** - formative assessment can be used to help students develop the capacity to self-evaluate, an important component for any future occupation. However, assessment is infrequently thought of in terms of preparation for lifelong learning.

For *teachers*

- ***Diagnostic***- teachers can use assessment tasks to ascertain what students bring into a course so as to make the teaching and learning responsive to student's needs and build on existing knowledge.
- ***Feedback*** - teachers can gain feedback on student's learning, detect misunderstandings, assess the effectiveness of their teaching and make appropriate modifications and adaptations.
- **Teaching *and* learning** - teachers can use assessment tasks as teaching and learning tools both through the nature of the tasks themselves and through formative feedback.
- ***Promoting*** **self-evaluation** - teachers can give feedback which encourages students to make judgments about the quality of their own work and prepare them for future participation in the workforce.

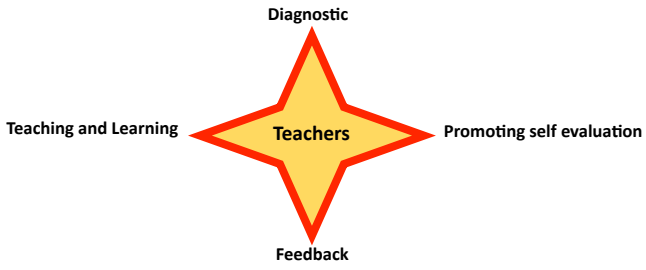


Figure 1.11 Teacher and Assessment

For Institutional and professional purposes

Assessment is also used for the following institutional and professional purposes to:

- Indicate '*pass*' or '*fail*'
- Select for entry
- Select for future courses and programmes
- Award grades
- Demonstrate institutional standards
- Select for employment
- License for practice

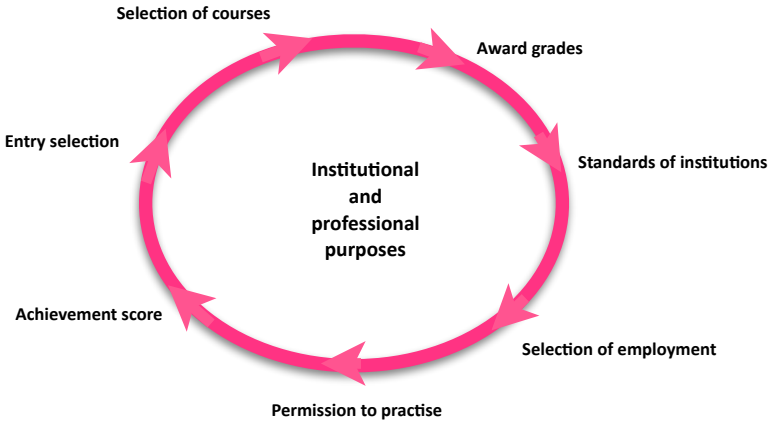
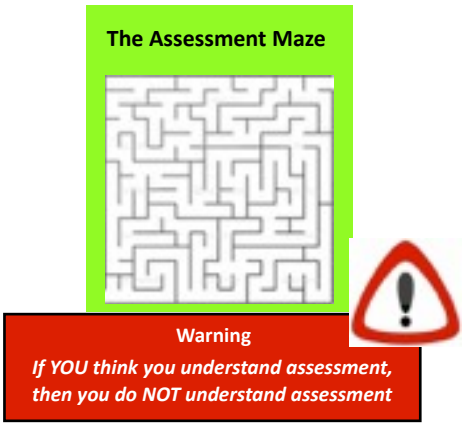


Figure 1.12 Institution and Professional purposes of Assessment

Students, teachers, Institutions

It is important to note that the needs of the learners may not coincide with the needs of the teacher and neither my the same as the needs of the institutions. Assessment is much more complex that at first sight!



Conclusion

Assessment is important for it controls the destiny of our students as they move along on their educational journey. Thus, it has a huge impact on all learners. Assessment is also a complex business because we are try to measure what is going on inside their heads. Assessment is also quite inaccurate for it only offers an approximate indication of success at *one* moment in time following *one* test or examination.

Assessment Moving *into* a New Future

Nature of Assessment

Aims

This chapter seeks to outline the characteristics of assessment so that you can see its real nature and be better equipped to understand what assessment can involve

Nature of Assessment

Assessment is an integral component of any successful teaching effort. Assessment in the context of education involves deciding, collecting and making judgements about evidence relevant to the goals of the learning. Johnstone (2003)[14] stated,

> '*Assessment forms an integral part of the teaching and learning process. It is applied during a course (formative assessment) to help students and teachers to take a realistic view of progress and to catch misunderstandings at a stage when they can be easily rectified. It is also applied at the end of a course (summative assessment) to ensure that the student has not only learned the bits, but can also piece the bits together into a coherent whole.*'

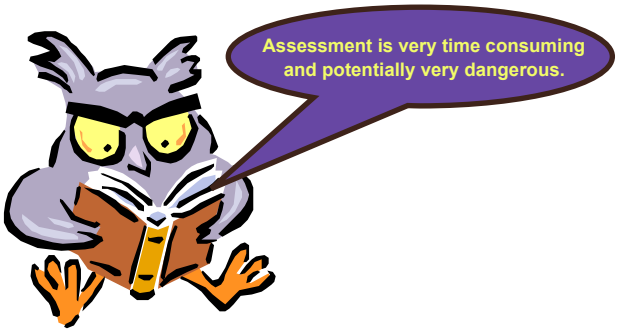


Figure 2.1 An Important Insight

We are well aware how time-consuming the setting of assessment tasks and their marking can be. Indeed, assessment places great time demands on our students as well. However, it is too easy for us not to assess the right outcomes or not to assess accurately. This is what makes assessment dangerous for it controls the future learning of our students and may well open or close opportunities for further study or employment.

[14] Johnstone, A. H. (2003) *LTSN Physical sciences practice guide: Effective practice in objective assessment*. Hull: LTSN. This monograph is online: http://www.heacademy.ac.uk/assets/ps/documents/practice_guides/practice_guides/ps0072_effective_practice_in_objective_assessment_mar_2004.pdf

We always need to remember:

> ***All assessment involves reasoning from evidence, is imprecise, and is only an estimate of what a person can do.***

Rowntree (1977)[15] views assessment as:

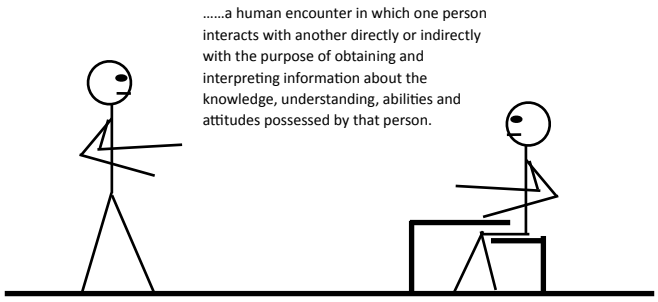


Figure 2.2 Assessment as human encounter

Assessment is an integral component of any successful teaching effort and *'if we wish to discover the truth about an educational system, we must look into its assessment procedures'* (Rowntree 1977). Internationally, academics are seeking better ways to assess students, recognising that diverse methods are available which may solve many of the problems associated with the evaluation of learning. Commonly, assessment is viewed as the process of collecting information with the purpose of making decisions about students. Data is collected using various tests, observations of students and interviews.

> Assessment is the ongoing process of **collecting, documenting, reflecting on, and using** information to develop rich portraits of learners in order to support and enhance their future learning

The overall process can be represented:

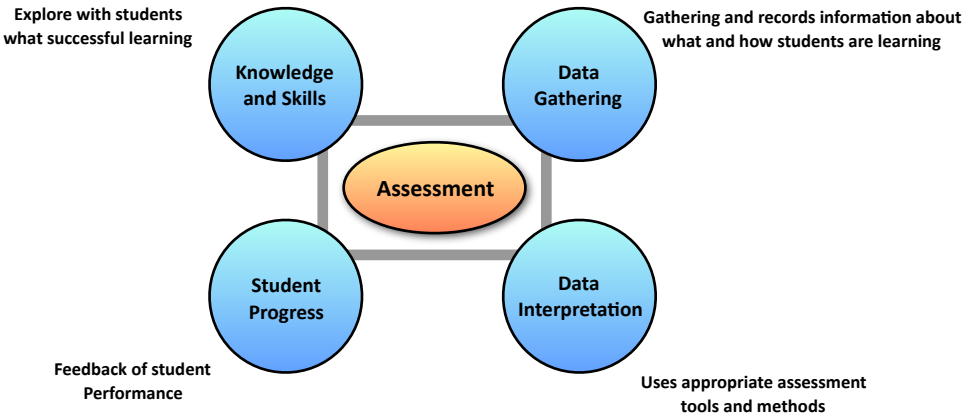


Figure 2.3 Process of Assessment

The figure above shows that assessment is the process of gathering and interpreting and reporting information about a student's progress and achievement in the way developing knowledge, skills and attitudes with the help of uses of appropriate assessment tools and methods.

[15] Rowntree, D. (1977) *Assessing Students: How Shall We Know Them ?* (1) London: Harper and Row.

Assessment and Learning

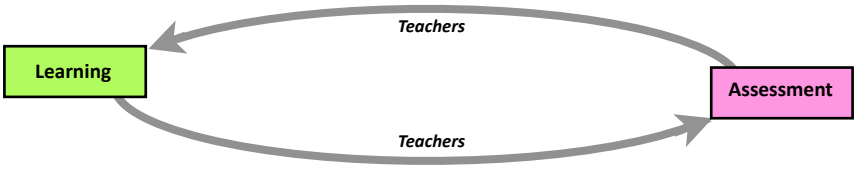
Assessment is one of the most significant areas of an educational system. It defines what students take to be important, how they spend much of their academic time and in many ways how they value themselves.

Assessment is a process of obtaining information that is used to make educational decisions about students, to give feedback to the student about his or her progress, strengths, and weaknesses, to judge instructional effectiveness and curricular adequacy, and to inform policy. Educational assessment is a complex endeavour, involving gathering and analysing data to support decision-making about students and the evaluation of academic programs and policies.

Students study mainly what they think will be assessed, imitating procedures presented by the teachers and expending effort proportional to the effort necessary to produce the amount and quality of the work perceived as acceptable. Thus, the naive view that assessment follows teaching and learning and measures their final outcomes is inadequate, because students see it coming (Bostock, 2004)[16]. '*Assessment backwash*' as defined by Biggs (1999)[17] means that student learning is mostly determined by the assessment tasks and not by the teaching or the lists of topics or objectives of the official syllabus.

Working back from desired learning outcomes, in relation to a unit or subject area, good assessment is designed to assess a broader range of student abilities (eg problem solving, critical thinking, effective communication, working in groups). Assessment shapes learning in positive and negative ways: for example, it can promote rote learning or learning in depth (Soliman, 1999)[18]. Teachers transmit feedback messages to students about strengths and weaknesses in their work assuming that these messages are easily decoded and turned into action.

Teachers may be able to identify where errors have been made, explain ideas or concepts, demonstrate techniques or procedures. They can also engage students in more in-depth thinking, suggest further study or reading, explain and justify marks or grades and suggest how to approach subsequent assignments (Brown *et al*., 2003)[19]. Effective feedback aims to improve student performance but for this to happen instructors must ensure that students know the performances expected of them, the standards against which they will be judged, and have opportunities to learn from the assessment in future assessments.



[16] Bostock, J. (2004) *Motivation and electronic assessment*. In: Irons, A. and Alexander, S. (Eds.) Effective Learning and Teaching in Computing (86-99), Routledge Falmer: London.

[17] Biggs, J. (1999) *Teaching for Quality Learning at University Buckingham UK* : SRHE and Open University Press.

[18] Soliman, I. (1999) *Assessing Student Learning.* Introduction to University Teaching Series, Teaching and Learning Centre, University of New England.

[19] Brown, S., Rust, C. and Gibbs, G. (2003) *Strategies for diversifying assessment.* Oxford: OCSD.

Principles of Assessment

It is possible to consider the key principles under five headings. These follow each other in a sequence:

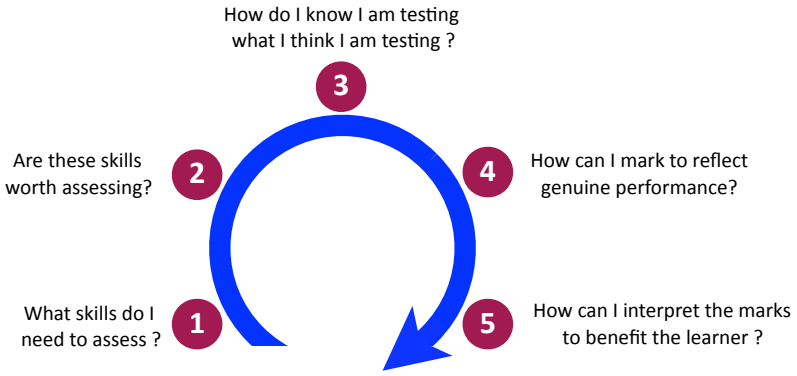


Figure 2.4 Principles of Assessment

Suppose you have taught a course on simple electrical circuits based on the ideas on Ohm's Law. You start by listing the key skills that you expect your students to be able to undertake as a result of completing the course. At this stage, you need to stop to ask if all of these skills are really important.

At this point, you start to develop questions to test the skills which you consider to be important. The difficulty is in being sure the learners gain their answers to your questions using the skill you are seeking to measure. School students are very clever at getting answers to questions by methods which surprise us!

You now mark the assessment but you need to give the right amount of credit for the various skills, reflecting their importance and their difficulty. You may well give the marks back to the students. However, it is possible to use the evidence you have gained form the assessment to show the learners their strengths and weaknesses and to help them to improve.

Key Description

Figure 2.5 shows a useful way to describe assessment.

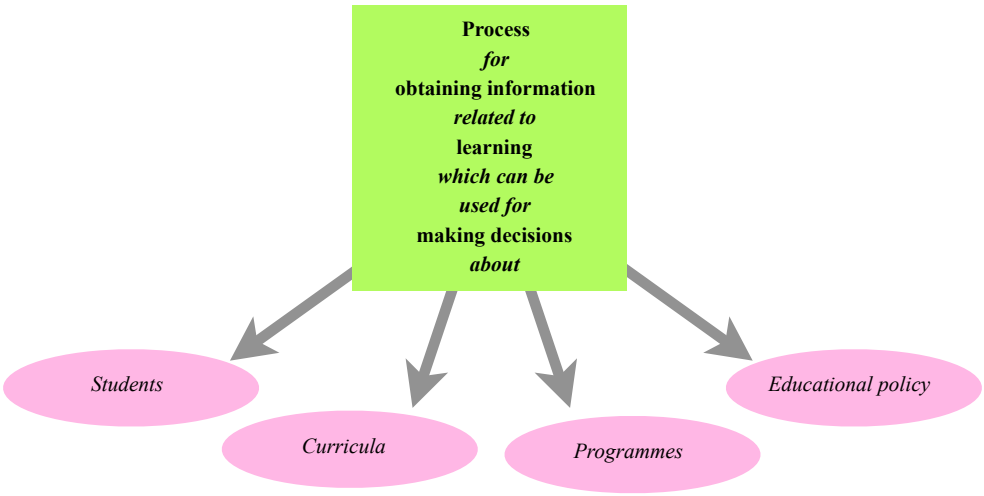


Figure 2.5 Description of Assessment (based on Capper, 1996)[20]

[20] Capper, J. (1996) *Testing to Learn - Learning to Test*, Washington DC, International Reading Association and the academy for Educational Development.

Purposes of Assessment

These can be considered under three headings: learner, institution and society.

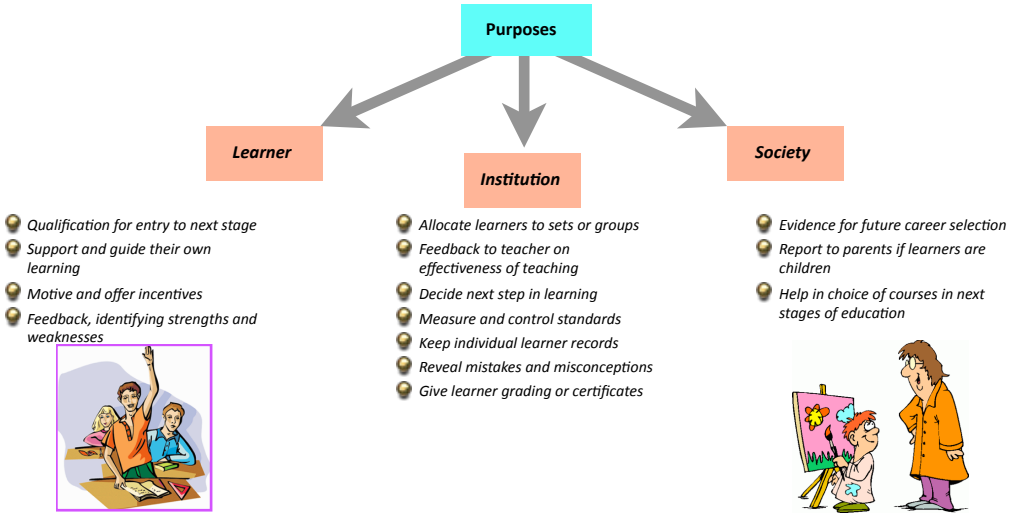


Figure 2.6 Assessment Purposes
(based on Ambu-Saidi, 2000, page 1-2)[21]

The range of purposes is considerable. It is important that we focus clearly on the purpose we need and that we then design the assessment in the light of that purpose. Indeed, it is vital that assessment designed for one purpose is not then used for another. If this is done, the outcomes may be very misleading. The purpose of the assessment will influence how we assess, the way we assess and the frequency of our assessments. We must move away from the idea that assessment means end-of-course examinations. There are numerous other ways to assess and most of them are far more useful than end-of-course assessment.

Key Questions

In thinking of assessment, we can look at it as a series of steps and these are shown in figure 2.7.

We need to start by asking what we are assessing: is it to inform our students of learning progress, to identify areas we need to teach further, or is it to award some kind of grade? Of course, we need to look at the curriculum goals and make sure that we are testing against these goals.

Then we can decide how we are going to carry out the testing: will it be an oral test, a written test and, if written, what formats of questions will we use to gain the greatest insights?

Probably the most difficult area is to be sure we are testing what we think we are testing. Far too often, students gain their answers by recall despite our best attests to measure other skills. Therefore, interpretation may not be as easy as it sounds.

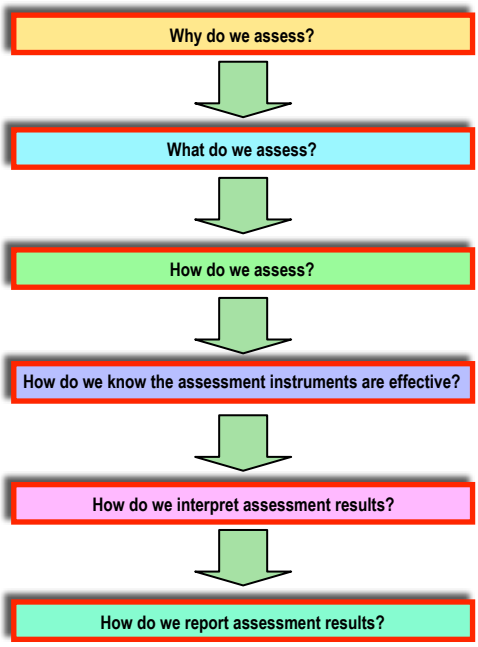


Figure 2.7 Questions about Assessment

[21] Ambusaidi, A.K.A. (2000) *An investigation into fixed response questions in science at secondary and tertiary levels*, PhD Thesis, University of Glasgow, Glasgow.

Key Questions About Assessment

It is possible to ask four questions of all assessment

Figure 2.8 Four Questions about Assessment

Teacher and learner identify need to know *why* we are assessing. Of course, assessment takes considerable time, energy, ingenuity and skill. That time is then not available for teaching. We need to know *what* we are assessing: this means defining the goals for assessment and then turning these into precise descriptions of the skills we want to identify. Many others may use the assessments we make. These users may have agendas and these agendas may or may not be consistent with how we see our assessments. There is, then, the vexed question about *how* we assess: what types of questions and tasks and a knowledge of how reliable these may be.

This can all be expanded further:

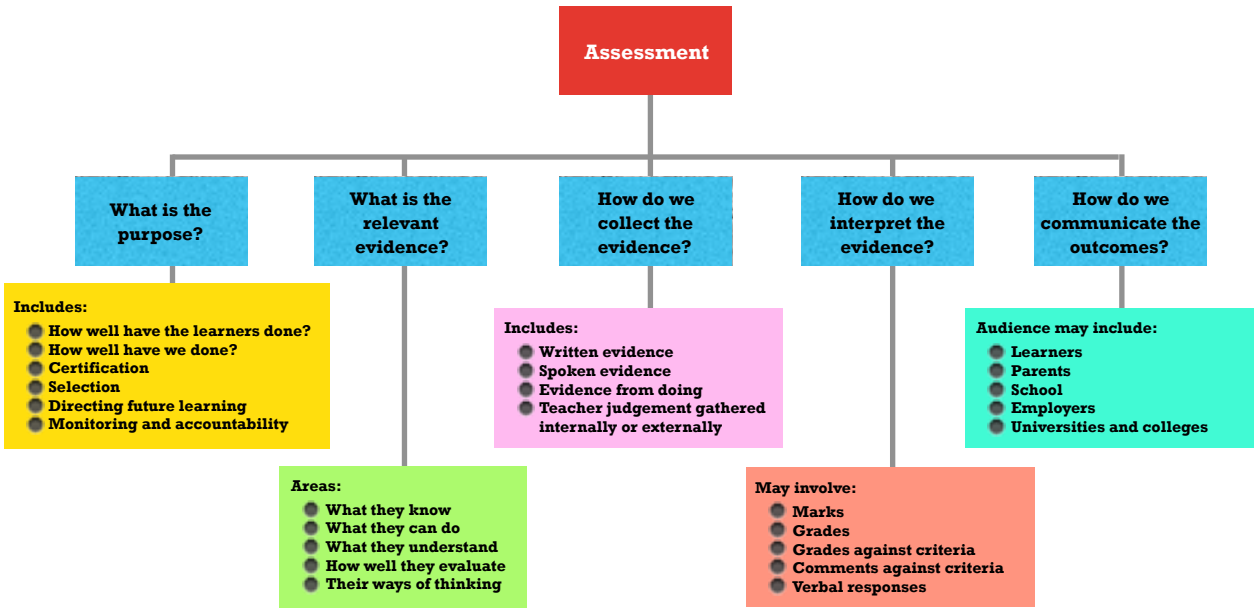


Figure 2.9 Analysis of Assessment

Notice the six suggested purposes for assessment. It is critical to observe that assessment designed for one purpose may give very misleading outcomes when used for another purpose. One glaring example of this is to use assessment data to measure the performance of schools or subject departments when that assessment was originally designed for certification. This is often done and almost always gives meaningless outcomes. Similarly, if we design assessment tasks to guide future learning, pin-pointing strengths and weaknesses in performance, then we must be very cautious if we use that assessment for certification.

Later, five areas for assessment will be suggested. What we as teachers need is precise, unambiguous specifications for what we to assess and then we need the assessment tools to give clear evidence on performance in these defined areas. We can express assessment outcomes in many formats and different audiences tend to prefer different formats. We need to know the strengths and weakness of these formats and how to share assessment findings in the most meaningful way.

It is sometimes argued that it is only worth teaching what can be assessed. This is a false argument. It is easy to list numerous skills which are almost never assessed formally at all: critical thinking, respect for others, good manners. All are highly important but we do not yet have good tools to assess them at an individual level. Assessment must never limit what we seek to achieve.

Assessment must never limit what we seek to achieve
Some of the most important outcomes from education are not easy to assess
Formal assessment must allow credit for such skills
This monograph will suggest ways to achieve this

Indeed, research demonstrates repeatedly that assessment can distort the whole process of teaching and learning in ways which are highly detrimental to good education. Some skills are best left un-assessed in any formal sense. Research often indicates how such skills can be developed, using statistical samples. The completion of specific tasks may then be known to develop the skills and the completion of the task can be recorded not the performance in it. We shall expand this idea further later in this monograph.

Hornby (2003)[22] suggests that assessment has four main functions:

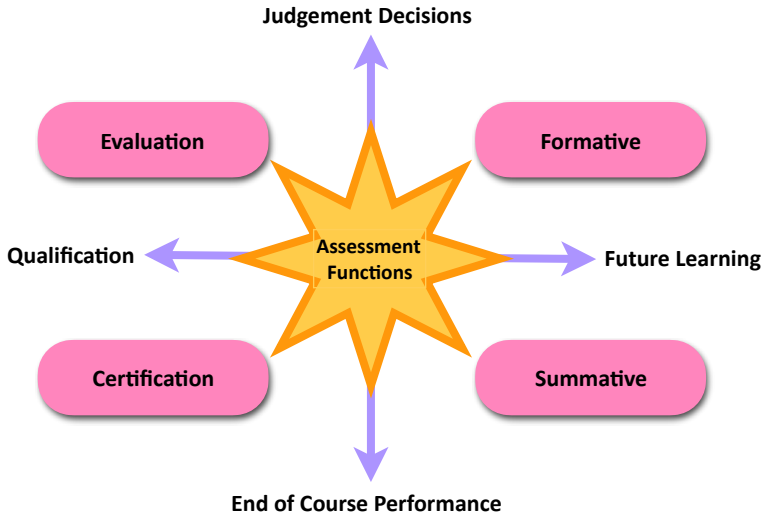


Figure 2.10 Assessment Functions

[22] Hornby, W. (2003) Assessing using grade-related criteria: a single currency for universities? *Assessment and Evaluation in Higher Education*, 28, 435 - 454.

Assessment for learning

Black and Wiliam (1998)[23] have reviewed some research in assessment. They recognise the powerful influence that assessment has in all learning. However, assessment practices need to take into account what is known about how students learn.

Assessment for learning involves assessment being constructed and used to pinpoint strength and weakness and, thus, directed future learning. They are strong advocates for assessment *for* learning. In this, they want to move the conception of assessment away from '*one-off*' tests and examinations at the end of courses or units of work. They want to move assessment so that it is an integral part of all learning. However, such formative assessments cannot really be used to award marks and grades.

The key issues relating to assessment for learning can be summarised:

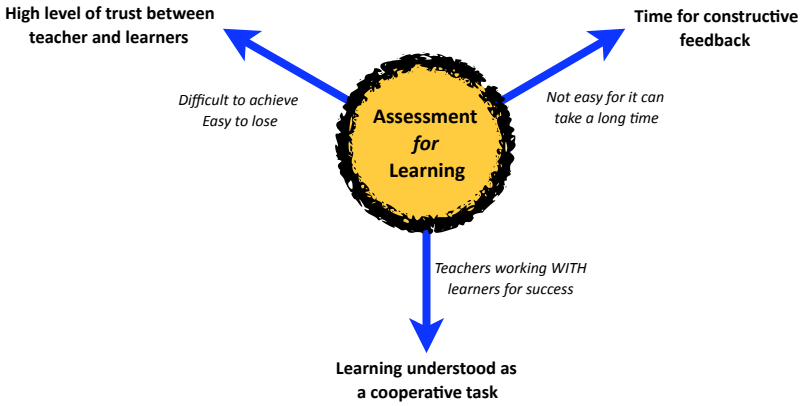


Figure 2.11 Issues with Assessment for Learning

It is difficult to disagree that assessment should enhance learning. However, this use of assessment means a radical re-thinking for assessment. It needs new kinds of assessment formats and an appreciation that assessment is no longer to be seen in terms of end-of-course formal written examinations. However, to use assessment this way requires:

- Imagination and creativity;
- Strong teacher empathy for the learner;
- A clear shared understanding of curriculum goals;
- Much time and careful planning

Assessment for the award of grades need not depend on formal end-of-course written examinations. The most difficult thing is that most of us have been brought up and educated in systems where examinations and tests give the final overall evidence of success or otherwise at the end of a course.

Traditional assessments needs re-thought completely

Figure 2.12 Traditional Assessment - no longer the way forward

[23] Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5 (1), 7–74.

The important thing to grasp is that using assessment to inform future learning by means of diagnosing strength and weaknesses is VERY different from using assessment to award grades. The purposes are radically different and the approaches adopted by an examination board to award grades and certificates are not necessarily going to be appropriate for diagnosing strength and weaknesses

It needs to be stated:

Using assessment to enhance learning DOES NOT imply that using assessment to certificate is removed

The fundamental purpose of examination board assessment is the award of grades and certificates. The evidence for making these awards can be gained from formal end-of-course written examinations. Equally, it can be gained from assessment conducted during the course as well as teacher assessments.

Here are five statements about assessment:[24]

Statement	Comment	Significance
Assessment is the process of gathering, interpreting and using evidence to make judgements about the achievements of students in learning	A description	Three processes: Gathering: Never easy Interpreting: Open to uncertainty Using: Open to misuse or abuse
Assessment is very time consuming and potentially very dangerous	Very true	Assessment drains time and energy. Assessment has power for good or ill
Assessment in education must serve the purpose of supporting learning	Easy to say, difficult to do	Research shows the influence of national examinations dominates our ways of assessment
All assessment involves reasoning from evidence, is imprecise, and is only an estimate of what a person can do	Welcome realism	Caution always needed. However, teachers tend to 'know' the abilities of their students!
Assessment is at the heart of learning. Through assessment, the student is able to measure their own progress and the teacher is able to offer guidance. [Learning Landscapes, 2009][24]	Role of BOTH teachers and pupils (plus parents)	Too often assessment forces only one kind of learning (memorisation) However, assessment can direct, support and encourage

Bringing it Together

Assessment is the process of gathering, interpreting and using evidence to make judgements about the achievements of students in learning

This can be summarised:

Aspect	Some Details
Use	Helping learning, tracking performance, informing student/teacher/parents, selection, certification Formative = assessment *for* learning; Summative = assessment *of* learning
Type of Task	Regular work, embedded tasks, tests/tasks
Agent of assessment	Teacher, student or outside agency (like an examination board)
Basis of Assessment	Norms or criteria, but norms almost always define the criteria.
Report or Feedback	Score, grade, rank order, profile, comment.
Moderation	Ensuring assessment instruments are valid, fair and reliable. Can be done within school, between-schools, or externally.
Verification	Checking actual assessments that they meet standards and are marked consistently. Often carried out internally and externally.

[24] Learning Landscapes (2009) Centre for Open Learning of Mathematics, Science, Computing and Technology. Open University, Miltm Keyes [downloadable: http://www.open.ac.uk/opencetl/resources/colmsct-resources/learning-landscapes-2009}

This chapter has aimed to give you an overview of the nature of assessment so that you are better equips to plan and carry out assessment which enhances the learning of your future students. The key things to remember are:

Nature of Assessment

- All assessment involves reasoning from evidence, is imprecise, and is only an estimate of what a person can do.
- Assessment involves anything form formal written examinations across to informal teacher observations.
- Assessment requires teachers to make judgements about a learner's performance, always set against sou kind of standards of expectations.

Process of Assessment

- Assessment is the ongoing process of collecting, documenting, reflecting on, and using information to develop rich portraits of learners in order to award marks or grade, or to offer support to enhance future learning.

Principles of Assessment

- Assessment needs to be designed to offer clear evidence of achievements related to the objectives of the course being assessed.
- Credit must be given for all the skills outlined in the course objectives, reflecting the balance of importance of these skills in relation to each other.

Purpose of Assessment

- Learner needs: to reward learners, to certificate learners, to direct learners
- Institutions needs: to assist schools and colleges in the way they develop teaching and learning
- Society needs: to guide for the appointment to jobs and future learning opportunities

Function of Assessment

- *Formative Role:* Assessment provides feedback to teachers and students about progress in order to support future learning
- *Summative Role:* Assessment provides information about the level of student's achievements at points during and at the end of of school.
- *Certification Role:* Assessment provides the means for selecting by qualification.

Assessment Moving *into* a New Future

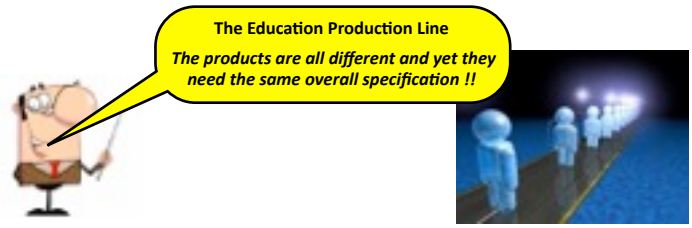
What to Assess

Aims

This chapter discusses the key goals for all learning and shows how assessment must make measurements of achievements against all these targets to move assessment away from pure recall

Introduction

In deciding what we want to assess, we need to think first of how we see the purpose of education. We need to develop a mental picture of the kind of '*product*' that we want our schools and universities to '*produce'*. The students will become the '*product*' of the entire process of education. Just what is our picture of the '*ideal student*' emerging from our schools or from our universities?



Here are some suggestions. We want our students to be able to:

- ☑ Develop all the skills and abilities that lie latent in our students
- ☑ Apply their understandings in ways useful to society
- ☑ Be able to do things with their knowledge with confidence and competence
- ☑ Be able to think critically, creatively and, perhaps, scientifically
- ☑ Play a full part in society, contributing to its welfare and benefit

Now, if that is our list of desirable outcomes, then we need evidence that we have achieved that. We need to assess against these broad aims.

Challenge

It is an interesting - and challenging - exercise to try to write down a description of what we think our ideal school student should be like at the point when the leave school.

Attempts have been made to do this for university graduates and the outcomes are fascinating[25].

[25] Harvey, L., and Green, D. (1994). *Employer satisfaction: summary.* Innovation in Higher Education Unit, University of Central England: Quality in Higher Education Project.

The suggestions given on the previous page can be translated in terms of '*products*':

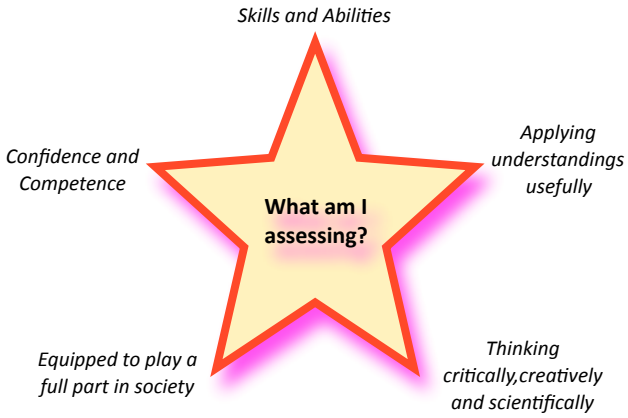


Figure 3.1 What am I Assessing

This picture can be developed further:

Figure 3.2 What am I Assessing

If these are goals, then we now have to develop ways by which we can gain the evidence about how well they are being achieved. This is moving away from the traditional assessment that merely rewards recall of memorised information, memorised explanations or memorised procedures. Traditionally, most forms of formal students assessment have involved reading and writing: the so-called '*paper and pencil tests*.' However, traditional tests and exams cover a very small portion of the potential range of skills, competencies and aptitudes that might may usefully be included.

Hargreaves (1992)[26], in an influential report, suggested that all the following aspects are potential areas for assessment:

- Written expression, knowledge retention, organisation of material, appropriation selection.
- Practical, knowledge application, oral and investigative skills.
- Personal and social skills, communication and relationships, working in groups, initiative responsibility, self-reliance, leadership.
- Motivation and commitment, perseverance, self confidence, constructive acceptance of failure.

[26] Hargreaves, D. (1992) *Report of the Committee on the Curriculum and Organization of ILEA Secondary Schools.* London; ILEA

In his latest book, Hargreaves (2004:12)[27] argues for generic skills like:

- Managing one's own learning
- Problem solving
- Thinking
- Research, enquiry and investigation
- Invention, enterprises and entrepreneurship
- Communication
- Social and interpersonal skills
- Teamwork skills
- Leadership

The 'perfect' student

Most of important of all, Hargreaves suggests, is the role that assessment plays in supporting the development of '*learning to learn*'.

A key point here is that '*what to assess*' is not to be confined to the traditional areas of knowledge and understanding - often referred to as '*the cognitive domain*'. Not only are skills involved but so also are personal qualities. As we shall see later, assessing these vitally important dimensions poses fundamental challenges to the traditional assessment toolkit and therefore requires a very different set of assumptions and techniques.

The assessment of secondary school student learning has long been a feature of secondary school of education in Pakistan. Teachers develop, construct and administer their own tests, administer standardised tests, and report the results of the assessments to parents and to other stakeholders. Teachers also engage in their own informal assessments of students and use their findings to inform on-going teaching and learning activities.

Impossible Tasks ?

Traditionally, in Pakistan, almost all national examinations and, indeed, university examinations, simply test the student's ability to recall information or procedures. What Hargreaves is stressing is that assessment involves ***very much more***, simply because education is very much more than filling the heads of our students with increasing quantities of information. However, the immediate thought in your mind might be that what Hargreaves has suggested in impossible!

What this monograph is seeking to do is to show, from research evidence, what is possible *now* so that you can start to widen the way you assess with your students. It is vitally important that we are realistic. We do not have infinite time nor are there endless resources. We must work within what time allows and what will cost no more than present.

Before we look further at the specific goals which are important for education, let us stop for a moment and look at a key word: ***literacy***. You may have met the words '*literacy*' and '*numeracy*'. It is essential that all learners leave school with sufficient skills in both in order to function fully and effectively in modern society and this involves literacy and numeracy. Here is a way to understand these words:

> ***Literacy*** might be described as having the knowledge, understanding and skills that are needed by everyone to be fully functioning in terms of ***communication*** within a society.
>
> ***Numeracy*** might be described as having the knowledge, understanding and skills that are needed by everyone to be fully functioning in terms of ***measurement*** within a society.

[27] Hargreaves, D. (2004) *Learning for Life: The foundations of lifelong learning.* Bristol: The Policy Press

Now let's look at the idea of '*assessment literacy*'.

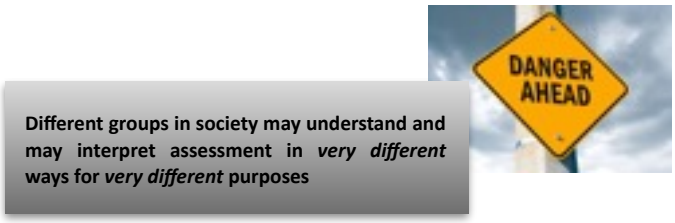
Literacy in any area might be described as having the knowledge, understanding and skills that are needed by everyone to be fully functioning within a society. Thus, literacy itself might be described as having the knowledge, understanding and skills that are needed by everyone to be fully functioning in terms of *communication* within a society. Similarly, numeracy might be described as having the knowledge, understanding and skills that are needed by everyone to be fully functioning in terms of *educational measurement* within a society.

Assessment Literacy

Using the same model, assessment literacy needs to have a purpose: assessment for a goal. Thus, for the teacher, we might see assessment literacy as:

Assessment literacy might be described as having the knowledge, understanding and skills that are needed to be fully functioning in terms of evaluating what the learners have achieved.

Assessment literacy will have different meanings for parents, for the learners and for wider society. This is because each group has a different set of purposes. This has enormous implications in that different groups in society may understand and may interpret assessment in *very different* ways for *very different* purposes. All kinds of inconsistencies may then arise.



For the moment, we shall focus on ourselves as teachers and draw in the problems and inconsistencies when we look at the way assessment is often used in society.

Figure 3.3 Assessment Literacy

What to Assess

If we want our assessment to be the basis by which awards and grades are given, then we need to know very precisely what will give our students credit. Simply rewarding the correct recall of information when it is written down in an examination papers should not be the basis for the award of grades. Recall, today, is not a very important skill at all.

If you were to ask a teacher what should be assessed in the classroom, the immediate response would be '*of course, the facts and concepts taught*'. There are the facts and concepts found in the sciences, history, geography, language arts, religious education and other similar subjects. However, there is *very much more* to these subject areas than simply facts and concepts. Indeed, education is very much more than the mastery of facts and concepts.

Several developments have occurred in the past 50 years or so in school assessment:

(a) There has been a move towards seeking to assess higher order cognitive skills and we shall explore what these are in more detail later

(b) There has been a move to depend less on end-of-course assessment and a move towards using assessment during a course, taking in a wider range of skills

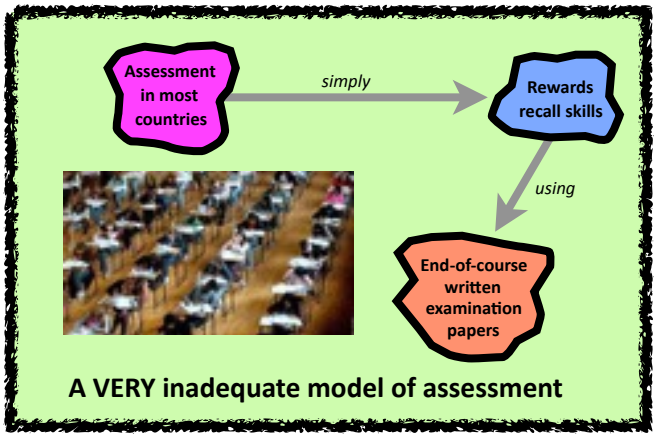
(c) There has been a move away from seeing assessment only as an evaluation of the '*product*' and a move towards assessing the process of learning and using assessment to guide learning.

(d) Finally, there is growing awareness that there is a wide range of ways to assess and a wide range of question styles and formats. The wider the range employed, the more accurate the assessment will be

The problem is two fold:

(a) Education is very much more than the facts and concepts. A student is only educated if there is understanding and the student apply that understanding in novel situations. A student is only educated if they are able to think about what they are learning: thinking creatively, critically and scientifically.

(b) Evidence from several research studies (Pidikiti, 2006; Al-Ahmadi and Reid, 2012; Hindal et al, 2013; Almadani et al 2012)[28] shows that typical school examinations in all subjects areas are simply testing recall. If there are no rewards for other skills, then teachers will not focus on other skills and students will concentrate on simply memorising.



Long ago, this problem was addressed by Benjamin Bloom and his colleagues. They recognised that assessment needs to reward much more than recall and they suggested six broad areas. Let us go back to what these people said then.

---

[28] Pidikiti, N.P. (2006) *Performance of secondary students in India related to working memory with reference to some learning styles*, MSc Thesis, University of Glasgow, Glasgow.
Al-Ahmadi, F. and Reid, N (2012) Scientific Thinking - *Can it be Taught? Journal of Science Education*, 13(1), 18-23.
Hindal, H., Reid, N. and Whitehead, R. (2013) A Fresh look at High Ability, *International Journal of Instruction*. 6(1), 59-76.
Almadani, K., Reid, N. and Rodrigues, S. (2012) What examinations test, *Problems of Education in the 21st century*, 1(1), 6-19.

Bloom's Taxonomy of Learning Domains

Bloom and his colleagues created what they called a Taxonomy of Educational Objectives. First of all, they identified three broad areas of assessment (Bloom, 1956[29]):

- Cognitive: mental skills (what we know)
- Affective: growth in feelings or emotional areas (how we feel)
- Psychomotor: manual or physical skills (what we can do)

These can be seen as domains for learning and domains for assessment.

Figure 3.4 Learning Domains

Domains can be thought of as broad categories and they describe areas where assessment is required. Bloom and his colleagues then developed a set of sub-categories within each.

Cognitive (Knowledge)	Affective (Feelings)	Psychomotor (Skills)
Recall	Receive (awareness)	Imitation (copy)
Understand	Respond (react)	Manipulation (follow)
Apply	Value	Develop Precision
Analyse	Develop personal value system	Articulation (integrate skills)
Synthesize	Internalize value system	Naturalisation (automate)
Evaluate		

Figure 3.5 Learning Domains Model

For the moment, we are going to concentrate on the cognitive domain: Bloom's Taxonomy of Cognitive Learning Outcomes

You will notice six ideas in the cognitive list. Here they are, with knowledge placed at the foot:

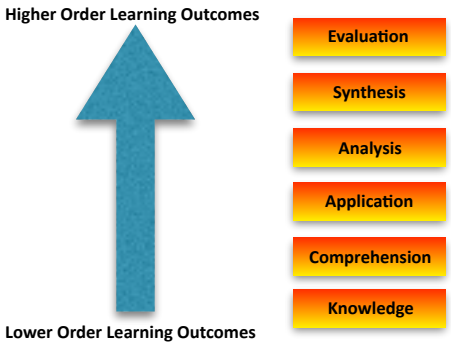


Figure 3.6 Bloom's Taxonomy

[29] Bloom B. S. (1956). *Taxonomy of Educational Objectives,* Handbook I: The Cognitive Domain. New York: David McKay Co Inc.
Bloom B.S. Bertram B. M, and. Krathwoh, D.R. (1964). *Taxonomy of Educational Objectives* (two vols: The Affective Domain & The Cognitive Domain). New York. David McKay.

Sadly in the 1950s, Bloom and his colleagues found that over 95% of test questions students encountered required them to think only at the lowest possible level: the recall of information. Analyses today suggest that things have not improved much, at least in some countries (*Pidikiti, 2006; Al-Ahmadi and Reid, 2012; Hindal et al, 2013; Almadani et al 2012)*[30]. Indeed, a look at examination papers in Pakistan suggests that the pattern Bloom observed is still there today: sadly, we are simply testing recall skills.

Bloom argues that we must test at all six levels. He saw the levels as *hierarchical*, each depending on the ones below it. In developed countries, during the 1960s, there was a rapid movement to look again at examination questions in an attempt to ensure that all national examinations test at all levels. Progress was made but only to a limited extent. Lorin Anderson, a former student of Bloom, revisited the cognitive domain in the learning taxonomy in the mid-nineties and made some changes, with perhaps the two most prominent ones being changing the names in the six categories from noun to verb forms and slightly rearranging them (Pohl, 2000)[31]. This new taxonomy reflects a more active form of thinking and is perhaps more accurate:

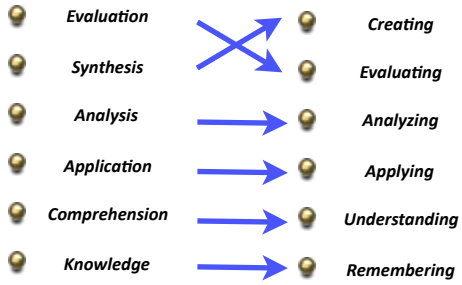


Figure 3.7 Bloom's Revised Taxonomy

Johnstone Comments on Bloom's Taxonomy

In 2003, Johnstone[32] wrote a monograph with the title, '*Effective practice in objective assessment*'. He pointed out that the six levels suggested by Bloom were not all of the same character, but placed different demands upon students. What follows is a free adaptation of his analysis. Objectives are placed on six levels and Bloom suggests that they form a hierarchy of complexity, but not necessarily a parallel hierarchy of difficulty. There are six levels in Bloom's classification with the lowest level termed '*knowledge*'. Let us describe each of these six levels, based on on the insights from Johnstone.

Level of Thinking	*Determine of Questions*
Knowledge	Questions ask students to give back what they have been given by recalling, defining, recognising and identifying.
Comprehension	Questions ask students to use knowledge in a familiar way to describe, compare or illustrate.
Application	Questions expect students to use knowledge in unfamiliar situations by applying, solving, classifying.
Analysis	As its name suggests involves concluding, inferring, deducing and distinguishing.
Synthesis	Requires the students to construct, develop, predict and solve.
Evaluation	Is the weighing-up process in which students have to argue, appraise, judge and decide.

Table 3.1 Six levels as described by Johnstone

[30] Pidikiti, N.P. (2006) *Performance of secondary students in India related to working memory with reference to some learning styles*, MSc Thesis, University of Glasgow, Glasgow.
Al-Ahmadi, F. and Reid, N (2012) Scientific Thinking - *Can it be Taught? Journal of Science Education*, 13(1), 18-23.
Hindal, H., Reid, N. and Whitehead, R. (2013) A Fresh look at High Ability, *International Journal of Instruction*. 6(1), 59-76.
Almadani, K., Reid, N. and Rodrigues, S. (2012) What examinations test, *Problems of Education in the 21st century*, 1(1), 6-19.

[31] Pohl, M. (2000) *Learning to Think, Thinking to Learn: Models and strategies to Develop a Classroom Culture of Thinking*. Cheltenham, Vic: Hawker Brownlow.

[32] Johnstone, A. H. (2003) *LTSN Physical sciences practice guide: Effective practice in objective assessment*. Hull: LTSN. This monograph is online: https://www.heacademy.ac.uk/resource/effective-practice-objective-assessment

This provides us with a useful language for discussing the nature of questions and their demands on students. There is a temptation to confuse the degree of complexity of these levels with difficulty, but it is very possible to have very difficult *knowledge* questions and relatively easy *analysis* questions. The other pitfall is to try to use more than one level in one question. For example, if a student has to recall some knowledge and then use it in some way such as in application, the quality of the answer depends upon both levels. Because the student fails to recall something correctly, it may be impossible to show application skill. At which level the failure has taken place is often hard for the marker to discern. This is particularly so in multiple choice questions which we shall visit later. The basic rule is that if you want to test the skills of evaluation, provide the knowledge for the student to evaluate. In this way, you separate the levels and test a skill with some confidence.

Johnstone suggested that there are several problems with the Bloom Taxonomy:

- The levels appear to be discrete and watertight, but there is no guarantee that what is application for one student is not just comprehension or even knowledge for another student who has read more.
- Many users of Bloom's work lay great store by the Taxonomy being hierarchical. This has the unfortunate effect of relegating knowledge and recall to a lowly status and inflating the value of the others. However, without knowledge, a person cannot understand, apply, analyse, synthesise or evaluate.
- By emphasising knowledge, we assume that what a student knows is key. However, in the days of electronic information, the key is knowing how to find out, not to the actual knowing.
- There is a lack of evidence that all six levels are hierarchical. Thus, is there any evidence that in order to be able to synthesise, the learners must first be able to analyse? These are just two different skills.

Johnstone suggested ways round these criticisms and one of them was later picked up by Yang[33].

An Alternative Model

Yang presented the Johnstone developments in terms of a simple diagram:

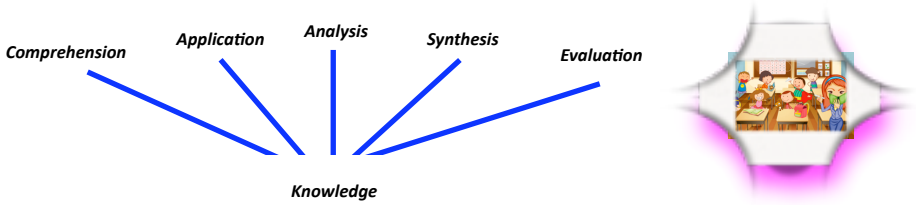


Figure 3.8 Yang model of cognitive skills[34]

The advantage of this model is that it acknowledge the point that what we know *or can access* is important. We do not understand, apply, analyse, synthesis or evaluate in a vacuum. We understand what we know, we apply what we know, we analyse what we know, we synthesise what we know and we evaluate what we know. However, what we know is still relatively unimportant because we can access information so easily today.

[33] Yang, M-Y. (2000) Problems Solving in Chemistry at Secondary School Science, PhD Thesis, Glasgow, University of Glasgow. [http:// theses.gla.ac.uk/2161/]

[34] Yang, M-J. (2000) Problem solving in chemistry at secondary school, PhD Thesis, University of Glasgow, Glasgow. [http://theses.gla.ac.uk/2161/]

Reid[35] presented the Bloom levels of the cognitive domain in this way:

Evaluation	*Make judgments about the value of ideas or materials.*
Synthesis	*Put parts together to form a whole, with emphasis on creating a new meaning or structure.*
Analysis	*Separates material or concepts into component parts so that its organizational structure may be understood.*
Application	*Use a concept in a new situation. Applies what was learned into novel situations in the work place.*
Comprehension	*Understand the meaning, interpretation of instructions and problems. State a problem in one's own words.*
Knowledge	*Recall data or information*

Table 3.2 Six level described by Reid

However, Reid developed the ideas much further. He described five broad areas for all cognitive assessment and translated these aims into something very practical:

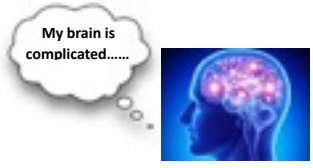
- What they **know**
- What they **understand**
- What they can **do**
- How they can **think**
- How well they can **evaluate**

Let us consider how these five broad areas can be described. This will enable us to consider how to assess against them.

Broad Areas	Working Description
Know	What the student knows (facts, concepts, skills) or can be accessed.
Understand	Described in terms of the extent to which the student can apply their knowledge in novel situations with some prospect of success.
Do	Skills (practical or procedural) which the student can demonstrate successfully.
Think	The extent to which students can think creatively, critically or scientifically in relation to the material being studied.
Evaluate	The extent to which the student can ask the questions *why? what?* and *how?* of new information, its sources and the way it links to which is already known.

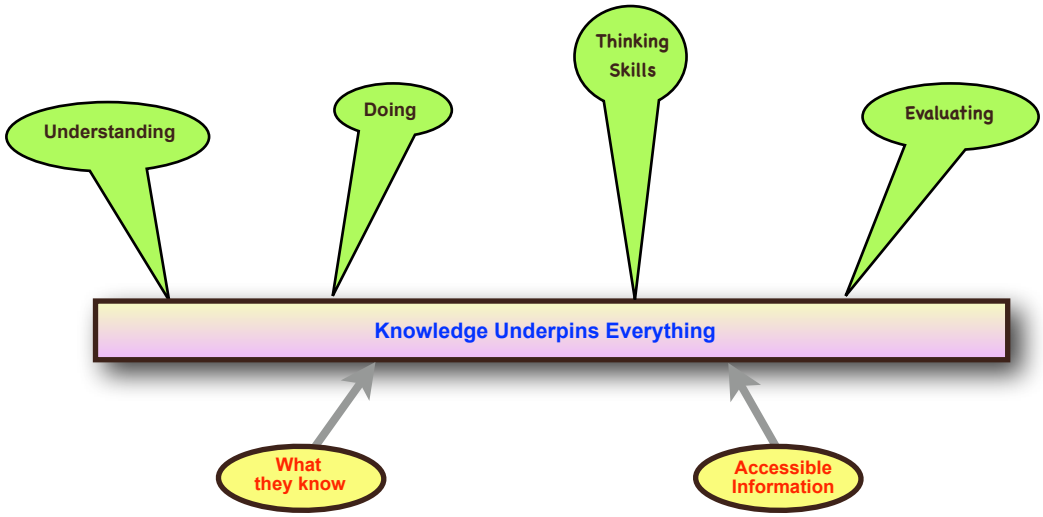
Table 3.1 Five areas as described by Reid

Modern education is seeking to emphasise skills beyond recall. However, it is worth remembering that the other four skills are all *based on knowledge* although much of that knowledge can be provided as needed, especially in the age of the internet. Although the five areas can be seen usefully as separate skills, there may be a little overlap. Thus, although evaluation may involve an attitude of mind in relation to new information, evaluation may use some skills derived from critical thinking. They key thing, however, is that the four skills (understand, do, think, evaluate) ***all depend on knowledge***, either known or accessible.



[35] Reid, N. (2008) *Landmarks in Assessment*, Unpublished monograph, Centre for Science Education, University of Glasgow, Glasgow

Here is a way of seeing how it all fits together:



The extent to which examiners test against this this range of cognitive skills may vary from one discipline to another, but there would seem to be a case for use of all of them in every subject area, perhaps in different proportions.

An analysis in one of the basic sciences at university level revealed that between 80-90% of the questioning was confined to knowledge and comprehension, and scant attention was paid to the others (Bennett, 2004)[36]. Things may not have changed much since then!

We have now set out the vocabulary which will be used throughout this guide. Other terms will be added as required, but the fundamental ideas elaborated here will underpin much of our thinking.

Words of Warning

Typical curriculum documents are replete with encouragement for us to teach towards all kinds of laudable goals. We are told to work towards understanding not memorisation, to aim for higher order thinking skills, to encourage problem solving skills, and so on. Rarely do we find any clear description of what is meant by such aims and almost never is there any practical indication on how we might assess them.

Let us be honest in all of this. In our hearts, we *want* to encourage the development of many of these skills in our learners but, when it comes to the final examinations, the learners gain their grades largely by recall. Pressures are placed upon us to achieve good grades. Therefore, the tendency is to focus on what the learners know and can reproduce or apply in a routine way.

The world of assessment does not need yet another set of platitudes. It needs to specify exactly what is meant by these desirable outcomes and give a clear indication of how assessment might be approached. This is not easy and we need to admit that in the present state of knowledge, assessment of some skills is not yet a viable goal.

[36] Bennett, S. N. (2004) Assessment in Chemistry and the Role of Examinations, *University Chemistry Education*, 8(2), 52-57.

Let us start with a much misunderstood word: '*understanding*'. Here is a simple description:

If you really understand something, then you can use that knowledge in a novel situation with some prospect of success

That means that, to assess understanding, we need to ask questions where the learners are faced with something new and where they can apply their understanding to seek to gain an answer. We shall give examples of this approach later.

Conclusion

Traditional assessment tends simply to measure the recall of what has been memorised. This is a completely unsatisfactory outcome for education involves very much more than memorising information. A model is suggested where there are five broad areas which need to be considered in assessment:

- What students **know**
- What students **understand**
- What students can **do**
- How students can **think**
- How well students can **evaluate**

Later in this monograph, we shall expand these ideas and show how assessment can be developed to measure against them. In addition, we shall expand them slightly to give a wider coverage of the goals for assessment in formal national examinations.

Assessment Moving *into* a New Future

For whom do we Assess

Aims

This chapter seeks to consider the groups in society who will need and will use assessment information. Their purpose varies widely and assessment designed for one purpose may mislead when applied to another.

Introduction

While teachers have, perhaps, three goals for assessment, an Examination board focusses only on one goal:

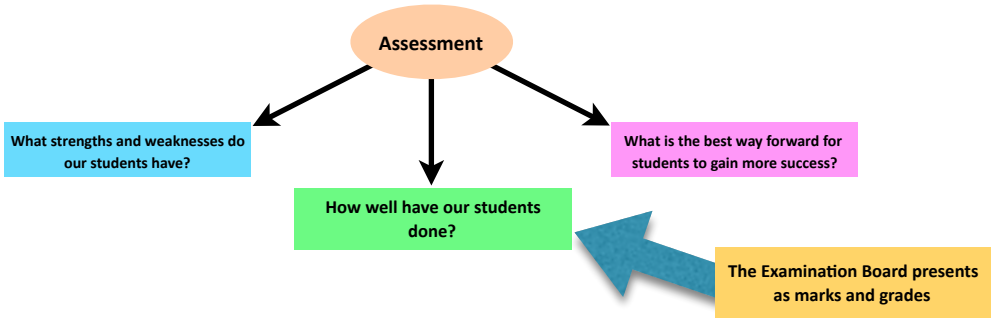


Figure 4.1 Purpose of Assessment

However, we need to remember that we are not merely trying to fill the heads of learners with facts to be recalled. We need to think of understanding, thinking, evaluation and skills. Education involves the cognitive (thinking) and skills (what we can do).

Learning involves changes in learners. The changes may be in what they know. More importantly, the changes may be in what they can understand, do and in how well they can apply their understandings. We are trying to equip the learners of the next stages of their lives. We are seeking to enable them to become well equipped to cope with life ahead. Let us look again at the model with five goals that we introduced before:

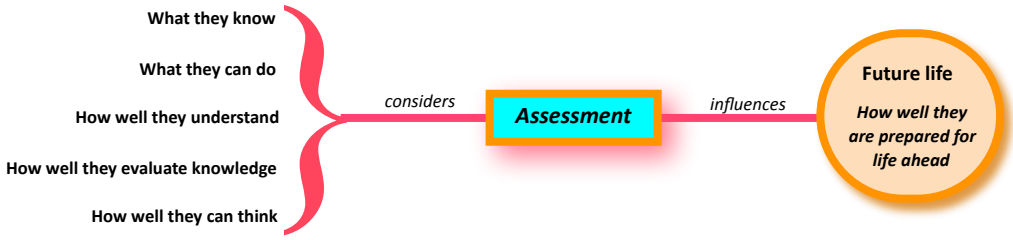


Figure 4.2 The role of assessment for life

Who will use Assessment Information?

There are, perhaps seven main groups:

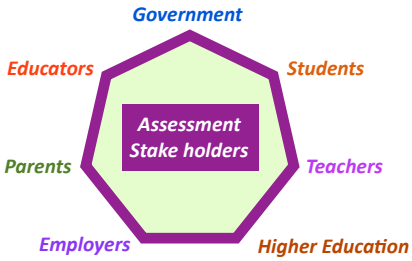


Figure 4.3 Assessment Users

The diagram can be expanded to show that the requirements for the seven different groups are very different:

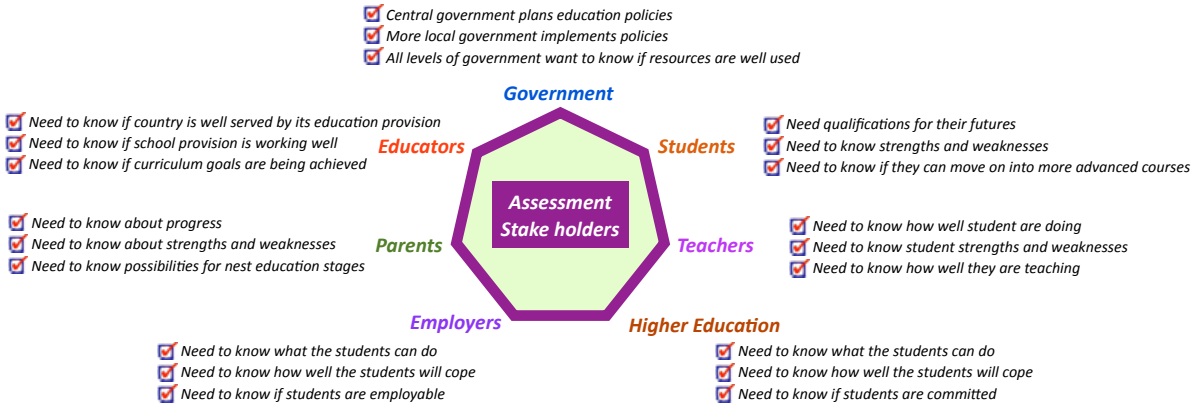


Figure 4.4 Assessment stake holders

The important thing to recognise is that assessment systems must serve many users. Decision makers at all levels need access to a variety of different kinds of information in different forms at different times.

There are two key issues involved here:

(a) Different users require assessment information for different purposes;

(b) Assessment information gathered for one purpose may be very misleading if used for another purpose.

There are diverse expectations and hopes for different user groups and these can be summarised:

Assessment User groups	
Learners	The want to know how well they are performing and some indication of how they can perform better in the future.
Parents	Parents wants information which is accessible to support learning, especially marks or grades. They want know how well their offspring are doing and what future prospects are like.
Teachers	Teachers wants assessment to support learning but often the tools to do this are not available or known. They want success for their students and the evidence of student progress.
Higher Education	Examinations grades are known by higher education to be of a highly dubious reliability but they are still used more than teacher reports.
Employers	Employers want evidence that the students can do the job: high in validity and authenticity.
Governments	Government want to know if standards are improving, if the investment into education is working and if there are economic benefits form education.
Educators	Those who lead education want to know if the curriculum is working and if performance is to an acceptable standard.

Table 4.1 Uses for Assessment Information

It is very obvious from the table above that different user groups have very different agendas. The issue is whether assessment information can be used for such a wide range of purposes. In the context of an Examination Board, the following uses may be relevant:

Award of marks and grades …

- Reflecting accurately the achievements of the learners
- To determine entry to courses in universities and colleges
- To give access to specific jobs and careers

The problems arise when examination data are used by others for other purposes.

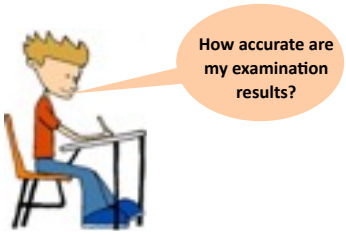
Abusing Examination Information

Suppose that national assessment information is used by school inspectors to indicate which schools are performing well or which teachers are performing well. The outcomes will be completely misleading. This is simply because there are numerous important factors which can influence national assessment results and school performance Teacher performance is only one and, indeed, may not be the most important factor. For example, different schools may draw very different students with very different ability profiles. If a school draw in students with relatively poor innate ability, examination results will never be as good as that obtain from schools with a '*better*' student population. In addition, the resources available to different schools and different teachers many vary widely and that will inevitably influence outcomes in examinations.

Interestingly, the same problem will happen if international tests are used to compare countries. Any test set across countries faces impossible problems. Different centuries have different curricula, different styles of learning, different education cultures, different patterns of examination questions. Students in one country may be at an enormous advantage simply because the international test questions are closer to those used in their own country. Overall, international comparisons are completely worthless. It is simply impossible to control the variables so that comparison are made fairly.

It is also worth downloading the questions from such international surveys and looking at the questions. When this is done, our confidence in international comparisons will rapidly evaporate. It is an impossible task to set questions fairly across very diverse educational variations.

There is another interesting observation and it comes by asking a simple question:



Let's look at possible sources of error. Here are a few.

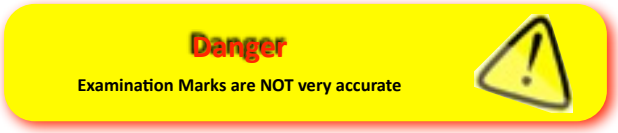
(a) The questions do not test over the areas studied by the students;
(b) The questions are set in ways unfamiliar to the students;
(c) The student did not perform to their maximum ability on the examination day;
(d) The paper marking did not following the marking briefly accurately;
(e) Marks totals contain mistakes.

How can we reduce these possible errors? Here are some ways forward:

	Source of Errors	Ways to Reduce Errors
(a)	The questions do not test over the areas studied by the students	Paper is checked by subject experts and practicing teachers to ensure a balanced coverage of the curriculum and balanced coverage of the goals of the curriculum
(b)	The questions are set in ways unfamiliar to the students	Students are prepared before hand in the kinds of question formats likely to be met in a national examination
(c)	The student did not perform to their maximum ability on the examination day	Never rely on one examination but ensure that there are several assessments taken at different times
(d)	The paper marking did not following the marking briefly accurately	Markers are briefed thoroughly and sample papers re-marked to ensure that all markers are following the same standards
(e)	Marks totals contain mistakes	All totals are checked.

We all have a touching faith in the accuracy of examination outcomes. Education managers and politicians (and the media) are even more confident of their accuracy. However, just how accurate are they. Let us look at a two subjects. The first is mathematics. In terms of marking alone, a mark might be expressed as something like 65 ± 2%. Allowing for other factors, it might be nearer 65 ± 5%. Now let us look at native language examinations. Marking alone can generate errors in a wide range (simply because language is so much more difficult to mark accurately, perhaps as much as ± 20% (if extended writing is being marked) although one study showed a much higher error limit[37]. Most subjects lie somewhere between the error limits shown for mathematics and native language.

We need to be far more sceptical about examination marks. They are simply NOT very accurate. However, if our students sit *several* assessments on *different* occasions, much of the error is greatly reduced.



All this assumes a process marked with integrity. If examination boards do not set and mark examinations with total integrity, then the grades awarded are largely meaningless. In addition, if politicians start to interfere with the examination process, perhaps to control the curriculum or to present an image that their policies are being successful, then the whole process collapses in total disrepute.

This illustrates the major problem. National examinations are set to assess what students have achieved and to give them recognition of that achievement. There are many sources of error. Therefore, we must treat all examination marks and grades with considerable caution. However, if those outside the schools and examination systems start to see that they can '*use*' the examination outcomes either for political purposes or to criticise schools or teachers, then the whole process collapses. National examining systems must be built on total integrity.

High Stakes Testing

High stakes testing is a phrase to describe tests or examinations, the outcomes from which are used to assess teachers, schools, or even nations. The problem is that we are assessing students and then assuming that the student performance reflects on teacher quality or school quality or even the quality of a national education system. The assumption is simply unsustainable.

[37] Spencer. E. (1981) Inter-Marker Unreliability in SCE O Grade English Composition. Is Improvement Possible? *Scottish Educational Review*, 13(1), May 1981.

Considerable research has been conducted on high stakes assessment and numerous problems have been found[38]:

(1) Teachers teach to the test

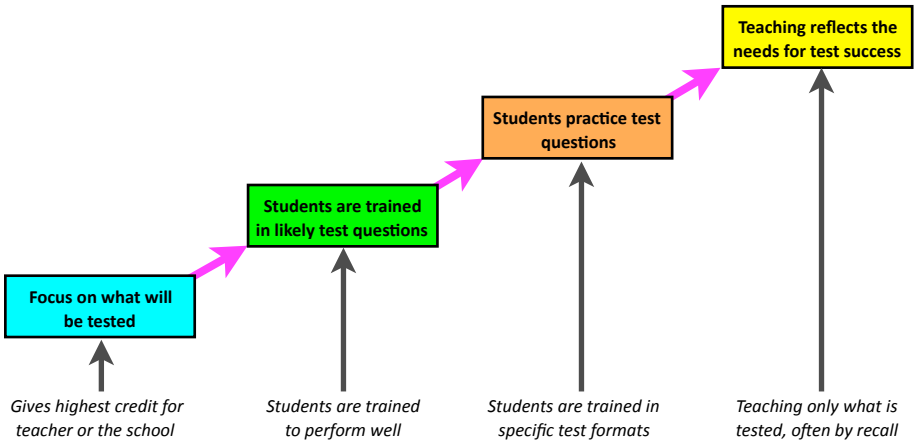
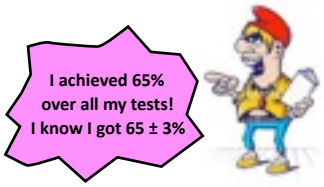


Figure 4.4 Application of test as assessment process

If the rewards come by teaching to the test, then teachers will teach to the test. Their students will perform well and everyone will be happy. In the process, real education may well be destroyed and schools may end up as places where information is poured into the minds of the learners to be recall accurately in examinations which use question formats familiar to the learners. High stakes testing can destroy quality in education.[39]



(2) Test scores rise continually as long as the test style stays the same

It has been observed that, in this kind of high stakes testing, examination results get better and better each year, *as long as the examination format remains constant*. What we are observing here is nothing to do with rising standards. It is simply that teachers get better and better at training their students for the particular tests and examinations being used.

In one school, a teacher used to set aside one lesson towards the end of a two year course where the learners had to face an examination involving a multiple choice paper. As a result of reading the research on how to be successful in multiple choice questions, he knew how to train the students for greater success. From experience, he knew that he could increase the average marks of the class by about 10-15% simply by training the students in the techniques for answering multiple choice questions. This was nothing to do with educating learners. It was all to do with training in examination techniques.

Societies are very foolish (and this is being polite!) to rely on examination data to assess teachers, schools or nations. This will, almost inevitably, lead to wrong conclusions and inaccurate outcomes, and this approach usually damages the quality of education.

Teachers are clever !!

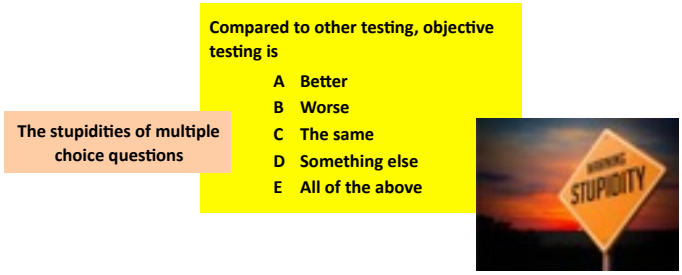
[38] for example, Meyer, H. D., & Benavot, A. (2013). *PISA, Power and Policy: the emergence of global educational governance,* Oxford, Symposium Books.

[39] Robinson, K. (2011) *Out of Our Minds: Learning to be Creative*, Capstone Publishing Ltd, Chichester. (He makes some very interesting observations of the effects in parts of the US).

(3) Tests tend to be as objective as possible in marking and thus rely on recall

Those who want to use tests and examination for high stakes think that they can make the tests more accurate by making them more objective. They do not appreciate that the ***only*** thing that is objective about objective questions is that the marking can be carried by a machine. If the machine is given bad information, we still get inaccurate marks. If the test paper is badly set, the outcomes will be meaningless.

However, there are many problems with objective questions. In most cultures, only multiple choice is used. Setting good multiple choice questions is not easy and there is considerable research to show that multiple choice questions give highly inaccurate measures of ability[40]. Furthermore, multiple choice questions tend to reward rote recall. The whole area of multiple choice will be discussed later and the research findings summarised then.



Thus, the overall picture is of examinations and test which are not very accurate in reflecting real learner abilities and, at the same time rewarding recall skills and ignoring higher order thinking skills. The net effect is an unfair examination and a neglect of important skills like critical thinking, creative and scientific thinking.

(4) Important skills are ignored

One of the problems with high stakes assessment is that it relies on traditional examinations, usually taken at the end of a course. The view is that the only this kind of assessment is reliable. The evidence does NOT support this viewpoint.

First of all, it is easy to show that one-off assessments in the form of terminal examinations are highly unreliable, not matter how well designed[41]. The tendency is to believe that teacher assessments are unreliable. However, evidence shows that teacher assessments, given appropriate criteria, can be highly accurate, probably better than formal end-of-course examinations. Combining teacher assessment and multiple written assessments seems always to give the *most* accurate outcomes, where the grades awarded reflect the learner achievements most precisely[42].

The difficulty is that, if we rely only on end-of-course formal assessment, many important skills cannot be measured in this way. Such skills be ignored simply because there are no rewards given for them. Education is effectively being destroyed and the rewards go to those who can reproduce the information of others most accurately and rapidly.

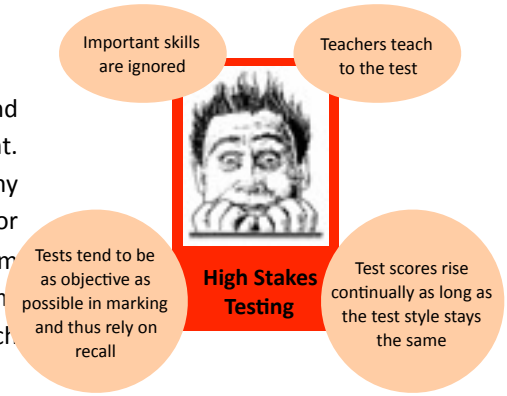
---

[40] Friel, S and Johnstone, A.H. (1978a) A Review of the Theory of Objective Testing, *School Science Review*, 59, 33-8.
Friel, S and Johnstone, A.H. (1978b) Scoring Systems which allow for Partial Knowledge, *Journal of Chemical Education*, 55, 717-9.
Friel, S and Johnstone, A.H. (1979a) Seconds Thoughts in Multiple Choice Tests in Science, *Journal of Chemical Education*, 56, 326.
Friel, S and Johnstone, A.H. (1979b) Does the Position of the Answer in a Multiple Choice Test Matter?, *Education in Chemistry,* 16(6), 175.

[41] Gardner, J., Harlen, W., Hayward, L., Stobart, G., and Montgomery, M. (2010) *Developing Teacher Assessment*, Maidenhead, Open University Press.

[42] Spencer. E., (1981) Inter-Marker Unreliability in SCE O Grade English Composition. Is Improvement Possible? *Scottish Educational Review*, 13(1), 41-55.

A Summary



There are some deep issues that need considered carefully:

(1) Assessment can generate both knowledge reproduction and social reproduction, restricting genuinely creative thought. While the reproduction of *'right'* answers is critical in many subject areas, we need to ensure that there is room for dialogue, debate, challenge and question. This is a real problem in that many outside the education world see education in terms of their past and, very often, wish to preserve a past which served them well.

(2) Testing can be used to control the curriculum and learning. Past examination paper almost define the curriculum and the standards expected. However, in many cultures over the past 30 years, testing has been used by politicians ruthlessly to control curricula and learning. Teachers are largely ignored and yet no one knows the educational needs of young people better than the teacher.

(3) Research shows that testing for high stakes (like assessing teachers or schools or nations) generates the following problems:

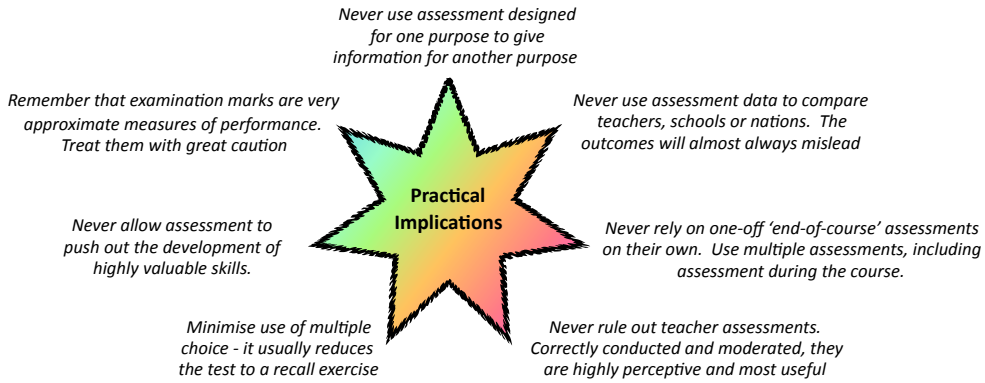
(i) Teachers teach to the test (focus on test content, train students to the tests, practice test, transmission styles of teaching);
(ii) Test scores rise continually as long as the test style stays the same;
(iii) Tests tend to be as objective as possible in marking and thus rely on recall;
(iv) Important skills are ignored (critical thinking, scientific thinking, even thinking!).

(4) There are diverse expectations and hopes from different user groups:

(a) Parents want information which is accessible to support learning, especially marks or grades;
(b) Teachers want assessment to support learning but often the tools to do this are not available or known;
(c) Examination grades are known by Higher Education to be of a highly dubious reliability but they are still used more than teacher reports.
(d) Employers want evidence that the student can do the job: high in validity and authenticity.

(5) Tests are sometimes applied and countries are then compared to other. Such comparisons are meaningless. It is ***impossible*** to control the variables and, therefore, like is not being compared to like: different countries have different curricula, different ways of teaching and learning, different examination structures and numerous other differences.

Practical Implications



Assessment Moving *into* a New Future

Types of Assessment

Aims

This chapter seeks to simplify some of the jargon that is often used to describe testing and to show straightforward ways by which testing can be understood

Introduction

Assessment can be a word word loaded with emotion. In this context, the assessment process can have the following effects:

- ***For schoolchildren***, an impending examination, test or indirect quiz can be a cause for alarm and depression;
- ***For teachers***, there is the administrative burden of setting and grading assessments, with the affiliated concern that they will be judged on their students' results;
- ***For admissions officers*** in academic institutions and employers, assessments are required to provide important information for the purposes of selection;
- **For *government ministers***, assessment results enable them to evaluate the effectiveness of the education system and how it compares with that of other states through the rankings in comparative studies (even though all such comparisons can be shown to be largely invalid!)

The Problem

In the past, test and examination results were predominantly meant to serve as indicators of what a student knew and understood of a subject. Sadly, today, assessment information has become a proxy measure that is supposed to facilitate judgments on the quality of most elements of our education system: teachers and teaching, head teachers, schools, support services, local authorities and even the government itself. This represents a fundamental change from the situation even a few years ago.

Sometimes, where there is more than one use to which assessment data is being put, it is not clear which is meant to take priority. In some countries, school-by-school performance data are published. The data, derived from seeking to measure the progress of learners is then used to make judgements on school quality. This may have negative consequences for the learners: making changes to '*improve*' school performance often goes against the young person's long-term educational needs. For example, teachers might drill students in techniques for earning marks at the expense of teaching for deeper understanding[43].

'Assessment which is explicitly designed to promote learning is the simple, most powerful tool we have for both raising standards and empowering lifelong learners.'
(Beyond the Black Box , Assessment Reform Group, 1999)

Problem Words

Assessment has built up its own jargon. We need to disentangle ideas and clarify meanings. Although the terms *assessment* and *evaluation* are often used interchangeably[44], many writers differentiate between them. Snowman *et al.* (2009) describe assessment as the *gathering or information or evidence,* and *evaluation is the use of that information or evidence to make judgments.*[45] However, that is slightly misleading.

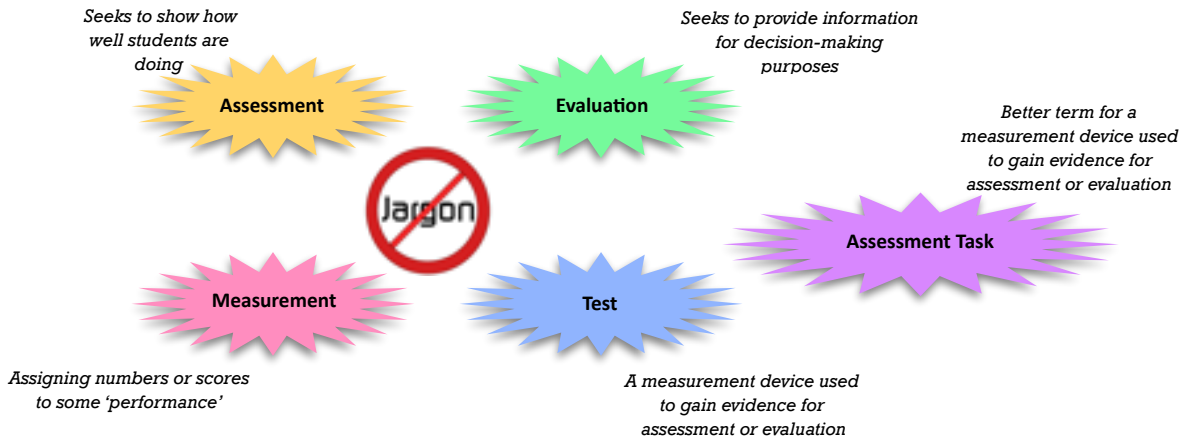
It is more useful to see ***assessment*** as obtaining information about the extent to which learners have achieved learning goals or are competent at some skills. In simple terms, assessment will be answering questions about *how well my students are doing*.

Now let's look at ***evaluation***: This usually describes some activity designed to see if an education programme is achieving its purposes. Educational evaluation is the process of delineating, obtaining, and providing useful information for judging decision alternatives. Furthermore, we can see that the essence of evaluation is to provide information for decision-making purposes. Thus, in education, we can evaluate the new curriculum, an education policy, specific learning resources, teacher or work ethic….

There is another word in common use: ***measurement***. This means assigning numbers or scores to an *'attribute or characteristic of a person in such a way that the numbers describe the degree to which the person possesses the attribute'*.[46] We can see this in examination marks or examination grades like A, B. C etc.

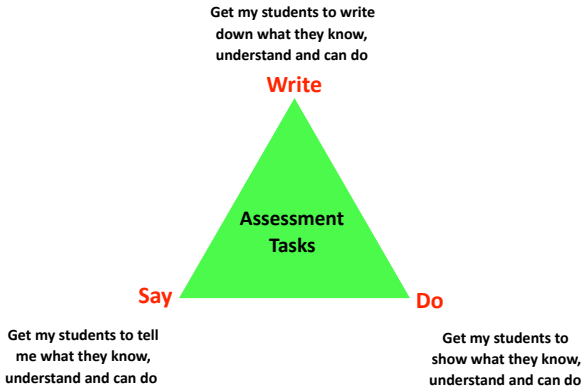
Finally let's look at the word '***test***': This describes the measurement device we use to gain evidence for assessment or evaluation. It includes tests and examinations of the traditional sort where students write answers. Equally, it can apply to tests of skills, like the skills of being able to pilot an airplane safely or the skills of conducting surgery in a hospital. It can also apply to oral examinations like those used for awarding a PhD degree. To include all the possibilities, a new phrase is one often used today: ***assessment tasks***.

[43] Assessment Reform Group, (1999) *Assessment for learning*: Beyond the Black Box. University of Cambridge.
[44] Cooper, J. M. (1999) *The teacher as a decision-maker. In Classroom Teaching Skills.* (6th Ed.) James M. Cooper (editor) pp. 1-19.. Boston: Houghton-Mifflin.
[45] Snowman, Macon and Biehler, (2009) *Psychology applied to teaching* (13th Edition). Belmont, CA: Wadsworth, Centage Learning.
[46] Nitco, A. and Brookhart, S. (2011) *Educational assessment of student.* (Sixth Edition). Boston: Pearson Education, Inc., p. 507

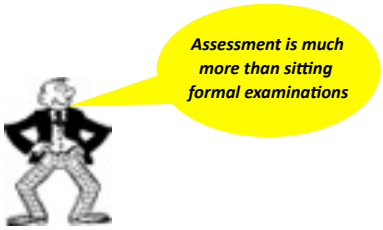


A Way Forward

Assessment tasks can include written examination or test papers, practical examinations or oral examinations.



In particular, in the context of learning in the classroom, the assessment is to determine the progress and outcomes of students, diagnosing learning difficulties, provide feedback, suggest ways for improvement of teaching and learning process, and determining progress in the class.



Data obtained from assessment of our students can be used to tell us how well they are doing, what they can achieve, where they have difficulties. Data from evaluation can be used to indicate how appropriate the curriculum is, how well the education processes are going and where there are weaknesses in provision that need adjusted.

More Jargon

Here is a quotation from an educational book:

> *Assessment data is only useful if it contributes to sound educational decision-making. In order to use this data effectively, beginning teachers must have knowledge and skills in the areas of measurement fundamentals, standardized tests and their interpretation, validity and reliability, constructing formal and informal assessments, and appropriately utilizing the results of both formative and summative assessments.*
>
> James Cooper (1999)

This quotation[47] uses many phrases:

- ☑ Measurement fundamentals
- ☑ Standardised tests
- ☑ Validity and Reliability
- ☑ Formal and informal assessments
- ☑ Formative and summative assessments

We could add also:

- ☑ External and internal testing
- ☑ Continuous assessment
- ☑ Criterion-referenced and norm referenced assessment
- ☑ Objective testing

Before we try to make some sense of the jargon, here are some fundamental principles behind all assessment:

We tend to think in terms only of written formal examinations, given at the end of some piece of work or at the end of a course. We tend to think in terms of marks as percentages and we see the mark as an accurate measure of the learner's success. We also tend to think that all we have to do in setting questions is to find out if our students know important pieces of information or procedures.

Assessment can involve much more than formal written examination papers. We can convert any mark out of any total into a percentage by simple arithmetic but, in practice it is much easier to use a spreadsheet for this. No examination mark is very accurate. The mark merely gives an indication of what our student can do, on a specific day, under specific conditions. The examination paper may not be a fair test of their skills while the marking may contain errors. An examination mark is, therefore, not very precise at all!! Finally, student learning is very much more than what they can recall.

Measurement fundamentals asks the following questions:

(a) Can we write course specifications against which we can set the test or task?

(b) Are we aware of the enormous number of test question formats that are available to assess various skills?

(c) Do we know how to set test questions that are unambiguous and test against the skills that are important?

(d) Do we know how the develop a marking brief that enables marking to be as precise as possible?

(e) Do we know how to handle test data to gain percentages or grades or whatever we need?

(f) Can we harness the enormous power of spreadsheets to make our lives easier?

(g) Are we aware of the accuracy limitations of any examination we set?

We shall explore answers to all these practical questions in a later chapter.

---

[47] Cooper, J. M. (1999) *The teacher as a decision-maker,* in Classroom Teaching Skills. (6th Ed.) James M. Cooper (editor) pp. 1-19, Boston: Houghton-Mifflin.

Standardised tests: Test and examinations can come in all forms and styles. They can be marked on all kinds of scales. Some tests are much easier than others. A mark gained on a very difficult test is of much greater value than a mark gained on an easy test.

The real problem is when we come to want to compare or combine two or more examinations or tests. Imagine your students have sat three tests on your subject area and you need to create an overall grade. You can only add the results of the three together if a mark on one test carries *the same value* as a mark on the other two. The tests will be of different difficulty levels and, therefore, a high mark in the most difficult test carries much more weight than a high mark on the easiest test.

The process of making all the marks carry equal weight is known as *standardisation*. It is very easy to do, with the computer doing all the work for you. Now we can add the marks together fairly because we have all three measurements working on the same scale: a mark in test 1 carries the same value as a mark in tests 2 and 3. The way to do this is shown in chapter 13 and any spreadsheet will do it all for you.

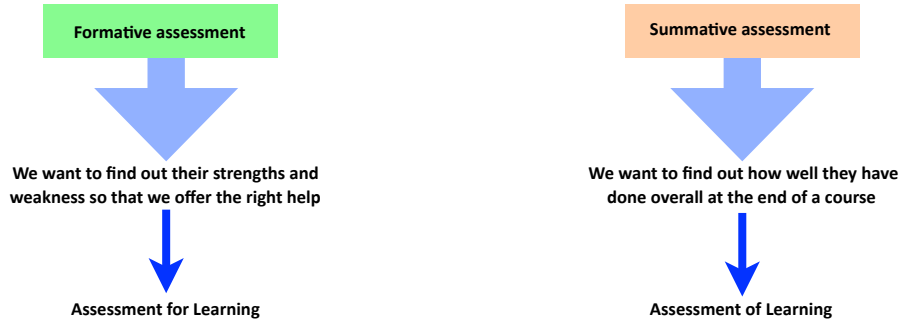
Validity and Reliability: These are just fancy words that mean something very simple! If you set a valid test or examination, it means that your test or examination is measuring what you *think it is measuring*. For example, if you want to test the learner skills in solving quadratic equations, then we do not set a test asking them to solve simultaneous equations. If you want to test learning understanding then we need to check that the students cannot get their marks simply by good recall. Achieving high validity is not as easy as it sounds. The reason is that our students are able to gain right answers by all kinds of methods that we often cannot predict or expect! Therefore, be very cautious when making claims about what you tests and examinations measure. They may be measuring things that are very surprising and not really what you intended.

*Reliabilit*y simply means that the test you used one day will give similar results with your students if you had used on another day. Think of measuring your height. You want to be sure that the measurement that is made one day will be the same as that on another day. Has the measuring tape stretched. Are you the same hight at different times of the day? Were you standing vertically? Did the person making the measurement read the scale correctly?

In a sense, it is a question of accuracy. However, you usually cannot repeat your test or examination on two different days - your students would protest! Nonetheless, we have to be cautious in now we interpret the results for our test. They not be as precise or accurate as we think.

Formal and Informal Assessments: We tend to think of formal examinations and tests. Our students have been prepared, they have a set time, there is a set examination paper. However, informal assessments may be more accurate and full of insight. We can watch our students carrying out tasks in class and we are left with a very clear impression of which students are doing really well and which are struggling. Indeed, we may gain useful insights about *why* there are the struggles. Never under-rate informal assessment. It can be very useful.

Formative and Summative Assessment: Again, the idea here is simple and it relates to the purpose of assessment:



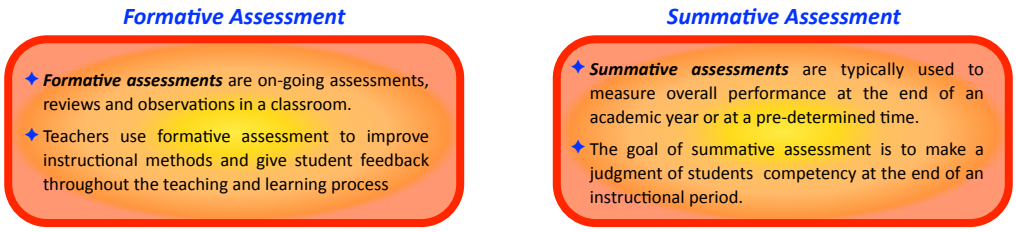
The literature is full of foolish statements that suggests that one kind is better than the other. In fact we need both! Universities and employers need some kind of evidence of student success or otherwise. That is summative assessment. Teachers need to know the strengths of weakness of their learners so that we can provide helpful support for future leaning. That is formative assessment.

Thus, AifL (1999) has noted,

> '*Good formative assessment will support good judgements by teachers about student progress and levels of attainment.*'
>
> ***and***
>
> '*Good summative assessment will provide feedback that can be used to help learning.*'[48]

Maxwell describes progressive assessment, which we consider below, as blurring the boundary between formative and summative assessment[49]. However, it remains the case that formative and summative reflect different purposes of assessment.

Here are examples of the way the two types of testing can be used:



For the purposes of Examination Board assessments, we are talking about summative assessment. This does not imply that the data are always to be gained for end-of-course formal examinations. Summative data can be obtained in numerous aways and at various points during a course. Summative assessment is important for certification, whereas formative assessment is important for checking the student"s development and potential[50].

Formative assessment is often done at various stages throughout a course, even at the start, thus providing the opportunity for immediate evidence for student learning in a particular subject area or at a particular point in a programme. The purpose of this technique is to improve the quality of student learning and should not be evaluative or involve grading students.

[48] AfiL, (2004) *Assessment is for Learning Information Sheet*. http://www.scotland.gov.uk/Resource/Doc/69582/0017827.pdf

[49] Maxwell, G. S. (2004) *Progressive assessment for learning and certification:* some lessons from school-based assessment in Queensland. Paper presented at the third Conference of the Association of Commonwealth Examination and Assessment Boards, March 2004, Nadi, Fiji.

[50] Biggs J.B. (2002) *Aligning teaching and assessment to curriculum objectives*. LTSN Imaginative Curriculum Guide IC022.

The goal of summative assessment is to make a judgment of student competency after an instructional phase is complete. For example, in Pakistan, national examinations are administered at the end of 6 years of primary school; at the end of lower secondary school and at the end of upper secondary school. It is a summative assessment to determine each student's acquisition of several subject areas of several years coverage of content. Summative evaluations are used to determine if students have mastered specific competencies and letter grades can be assigned to assess learner achievement.

At the moment, there is a lack of formative assessment in the Pakistani education and that is a gap that needs addressed urgently. However, discussing formative assessment is beyond the scope of this text.

External and internal testing: This idea is straightforward. It relates to the control of testing and examinations. If teachers arrange the tests and examinations and carry them out and mark them, then it is all internal in the school. Where there are national examinations then the control rests outside schools: setting, conditions of conduct, and marking.

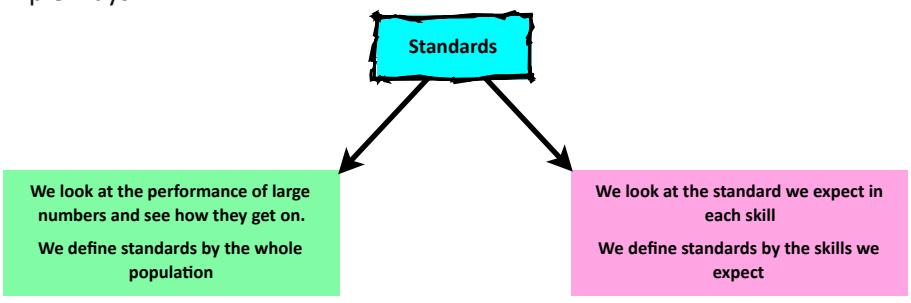
There is a strong view in most societies that external testing is good because it is more accurate and fair. The evidence does **NOT** support this view. Teachers know their own students better than anyone else and, provided that teachers act with fairness, internal assessments can often be *more* valid and accurate than external assessments. If you are teaching, never under-rate your skill in assessments!

Continuous Assessment: The common practice through the earlier part of the 20th century in developed countries was to rely on big examinations at the end of a course. In the latter half of the century, there was a move to have several smaller assessments at various stages throughout the course and combine the marks in some way to gain the overall grade. This was sometimes describe as *continuous* assessment. Continuous assessment is ***not*** *continual* assessment!! We assess periodically, not all the time.

Because all assessment has an inbuilt lack of accuracy and reliability, it makes sense to assess several times and combine the marks in some way to give the final grade. This is now the normal practice in most universities world wide and has been adopted in several countries at school level. It needs to be developed for Pakistan and this will need careful thought as well as tight clear structures.

Criterion-referenced and norm referenced assessment: This was a fashionable way to look at assessment several decades ago. It considers the standard for tests and examinations: how do we judge the standards?

There are two simple ways:

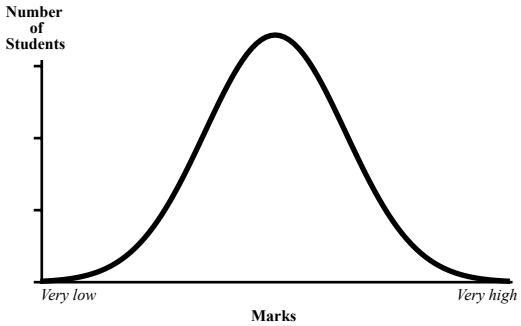


The green box describes norm-referenced assessment while the purple box describes criterion-referenced assessment.

Let's look at each in a little more detail.

A Detailed Look

Imagine a national examination in, say mathematics. Suppose 50000 candidates sit the examination. The marks obtained will follow a normal distribution. Here the majority will obtain marks near the middle with smaller numbers gaining very high or very low marks.



Marks mean very little for they reflect the demand level of the specific examination paper. National examinations involve large numbers of candidates and it is possible to use the marks to place the students in a rough order of ability set against the criteria being measured in the particular examination.

> *'The point on the curve which corresponds to the pass mark can only be determined by human judgement. Thus, the percentage gaining a pass is entirely dependent on the professional judgement of the examiners. Given a large number of candidates, then it is reasonable to expect approximately the same proportion to pass in successive years in any national examination. Thus, if 70% of the candidates passed last year in a specific subject, it is likely that a very similar proportion will have the similar abilities the following year and, therefore, deserve to pass'*.[51]

It is worth looking at the paper by Almadani *et al*. for it gives a clear picture of how national examinations are often applied.

This is all typical of a norm-referenced system. It is widely used throughout the world and it gives an excellent way to maintain standards.

Now let's look at criterion-referenced assessment in more detail.

Anastasi[52] has noted that

> '*Criterion-referenced tests determine what students can or cannot do, and not how they compare to others.'*

Criterion-referenced tests report how well students are doing relative to some pre-determined performance level on a specified set of educational goals or outcomes included in the curriculum. Criterion-referenced tests are used when teachers wish to know how well students have learned the knowledge and skills which they are expected to have mastered. This information may be used as one piece of information to determine how well the student is learning the desired curriculum and how well the school is teaching that curriculum. Criterion-referenced tests give detailed information about how well a student has performed on each of the educational goals or learning outcomes included on that test.

---

[51] Almadani, K., Reid, N. and Rodrigues, S. (2012) *What examinations test, Problems of Education in the 21st century*, 1(1), 6-19.

[52] Anastasi, A. (1988). *Psychological Testing.* New York, New York: MacMillan Publishing Company.

Let us look at an example in mathematics. Suppose we want to set a test in solving quadratic equations. We could say something like this. The standard for a pass is that the student can solve successfully four out of five quadratic equations where the coefficients never exceed 20 and they are capable of factorisation. That sets the criteria. It sets the standard. We can apply that standard to one individual student, to a whole class or to a whole country of students.

However, there are problems in all of this. Here is a list:

Problems, problems …. !

(1) There is no evidence that one form of assessment is better than the other. Criterion-referenced assessment is ***neither better nor worse*** than norm-referenced assessment.

(2) Norm-referenced assessment works beautifully when there are large numbers (probably more than 1000) but is less good with smaller numbers.

(3) Setting the criteria for assessment is not easy. Indeed, it is very difficult. In some subject areas, it is much more difficult than others.

(4) Setting the criteria depends on our experience, as teachers, of what learners are likely to be able to do. In other words, criteria are determined by norms!

The simple answer is to accept that we need both and that each can enrich the other. However, in Examination Board examinations where there are large numbers, norm referencing may an easier place to start.

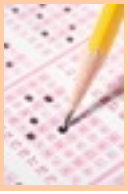
Here is a picture of the differences between the type types of assessment:

Dimensions	Norm - Referenced Assessment	Criterion - Referenced Assessment
Aim	***Compare:*** ✓ Student's performance with other students ✓ Selected students for certification	***Compare:*** ✓ Student's performance against some criteria (e.g.learning outcomes) ✓ Extend to which students has acquired the knowledge or skill ✓ Improve teaching and learning
Types of Questions	***Questions:*** ✓ From simple to difficult	***Questions:*** ✓ From nearly similar difficulty relating to the criteria.
Reporting of Results	***Grades:*** ✓ Grades are assigned	***Grades:*** ✓ No Grades are assigned (whether skill or knowledge achieved or not)
Content Coverage	***Subject matter:*** ✓ Wide content coverage	***Subject matter:*** ✓ Specific aspects of the content
Example	***Formats:*** ✓ National examination ✓ End of semester Examination ✓ End of year Examination	***Formats:*** ✓ Class tests ✓ Exercises ✓ Assignments

Objective testing

It is often suggested that we should move to objective testing and it is implied that, being objective, it is more accurate. Nothing could be further from the truth! The people who suggest this usually mean multiple choice questions. Firstly,

The only objective thing about objective testing is that they can be marked by a machine



Secondly, multiple choice test questions are designed by human beings. They are difficult to set. Because the answers are '*closed*', there is far greater scope for ambiguity and language confusion. Indeed, working memory overload is common. It is also difficult find four or five plausible answers. It is also very difficult to set good questions which test anything other than recall. We often *think* our questions are measuring something else but, if we talk to the candidates, very often we find that they obtain their answers only by recall.

Research has shown that multiple choice tests are not very reliable and they are not really measuring what we expect. Indeed, there are ways to *'train*' students to improve their performance without any increase in knowledge of the subject area being tested! There are some far more useful types of objective tests and we shall return to look at these later. For the moment, there is a definite warning over multiple choice questions. They are not a good way to assess in general.

Multiple Choice
A Throw of the dice?

Conclusions

Some key assessment ideas include:

- We assess students; we evaluate courses, learning and teaching.
- The most important thing is be sure that we are testing what we think we are testing: validity. This is not as easy it sounds.
- In formative assessment, our aim is to use assessment to guide future learning.
- In summative assessment, we want to obtain an overall picture of performance.
- Formative assessment (assessment for learning) sounds very good but there are major problems in doing it well.
- In norm-referenced assessment, we award grades on the basis of the overall performance of the population.
- In criterion-referenced assessment, we award grades on the basis of performance set against agreed standards.
- In both norm-referenced and criterion-referenced assessment, the standards are decided by the value judgement of experienced teachers.
- The only advantage of objective assessment is that machine marking saves time after the assessment is over.

Assessment Moving *into* a New Future

Difficulties in Assessment

Aims

This chapter seeks to look at how good assessment is difficult to achieve and it suggests ways by which we may make our assessment more robust and fair, for the benefit of all

Introduction

Let us think about assessment for a moment. We are trying to see inside the brain of learners and find out what they have achieved. Of course, we have to describe carefully what we mean by '*achieve*'. Education is far more than simply memorising information and then recalling it later in an examination. '*Achievement*' for learners involves understanding, applying and using what is understood, and evaluating what has been learned as well as developing skills in thinking.

Assessment as a Measurement

The problem is how can we look inside the head of learners? Normally, we do this by asking our students to do something to show what they have achieved: they talk, write or demonstrate a skill. This is assessment of learning and it means making some kind of measurement. Johnstone[53] makes an important point like this:

> *'It has to be admitted that assessment, no matter how we disguise it, is a measuring device to provide information about progress and attainment.'*

He[54] then goes on say that any assessment we use must be:

☑ Valid	Are we measuring what we think we are measuring?
☑ Reliable	Are we measuring consistently and accurately?
☑ Humane	Is the measurement procedure civilised, fair and kind?
☑ Economical	Can the measurement be carried out in reasonable time and at reasonable cost?

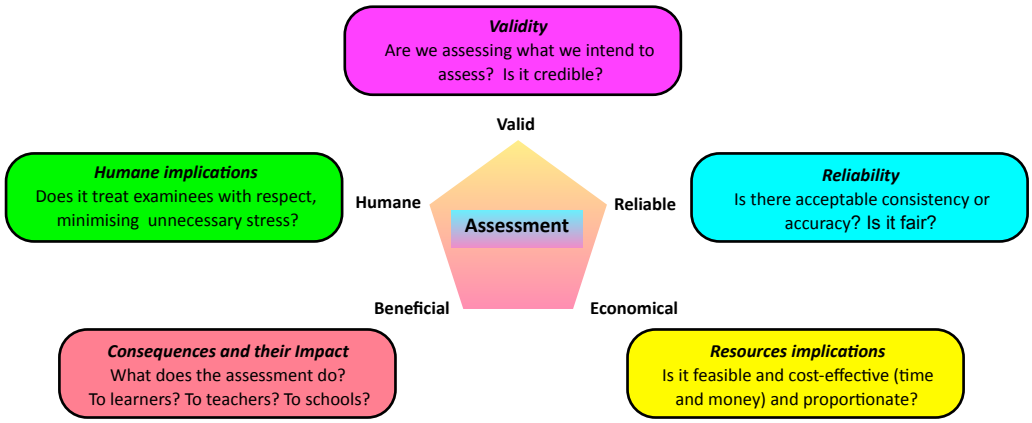
Perhaps there is another issue to take into consideration. Any assessment must be

☑ Beneficial	Does the assessment bring benefit to learners and also to others?

[53] Johnstone, A. H. (2003) *LTSN Physical sciences practice guide: Effective practice in objective assessment*. Hull: LTSN. This monograph is online: http://www.heacademy.ac.uk/assets/ps/documents/practice_guides/practice_guides/ps0072_effective_practice_in_objective_assessment_mar_2004.pdf

[54] Johnstone, A. H. (2003) ibid.

Let us look at each of these in turn:



Let us be honest.

- Too often we assess what is easy to assess and that usually means testing the ability of the candidates to recall information. In an internet age, that skill is no longer important.
- Very often we rely on one end-of-course examination and fail to recognise that this will ***not*** be a reliable way to assess our students. We need several tests it make things fair and consistent.
- Do we prepare our students so that they can see the examination as an opportunity for them to show their achievements in a positive and unthreatening way?
- Have we ever thought through the time we spend in assessment and then look for ways to make it more cost-effective?
- Examinations can destroy learner confidence. Using them to compare teachers, schools or countries can destroy an education system and those who are the key to its success - the teachers themselves.

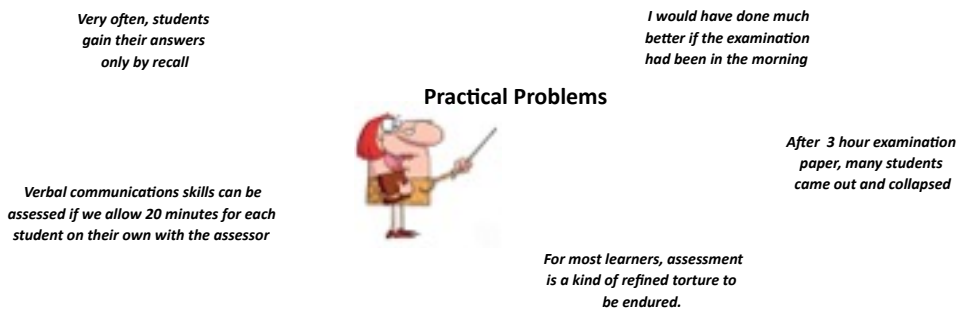


Figure 6.1 Practical Problems

More About Validity

The problem with the word '*validity*' is that we use it with many meanings. In the context of assessment VALID means that the '*test instrument measures what it sets out to measure*' (Johnstone, 2003)[55].

This seems to be utterly obvious and naive and yet many questions are set which are invalid. A good check for an invalid question is to ask if the question could be answered by someone who had ***not*** done the course. An example is given below which was supposed to test biological knowledge but could be answered by anyone using common sense.

[55] Johnstone, A. H. (2003) *LTSN Physical sciences practice guide: Effective practice in objective assessment*. Hull: LTSN. This monograph is online: http://www.heacademy.ac.uk/assets/ps/documents/practice_guides/practice_guides/ps0072_effective_practice_in_objective_assessment_mar_2004.pdf

Look at this question, again drawn for Johnstone (2003):

James and Iqbal are in the same class at school.
In which one of the following respects are the boys most likely to differ?

A Height
B Age
C Rate of heart beat
D Number of toes.

Now think about it: What does this question test?

The two boys are clearly of similar height, and, being the same class, they are likely to be of similar age. Humans usually have five toes on each foot. That leaves answer C. We do not need to know any biology to obtain that answer - just common sense.

Now go back and look at some of the questions you have set in the past. Can students get answers simply by common sense?? Do not panic! We all have done this. However, let us look at our assessments again.

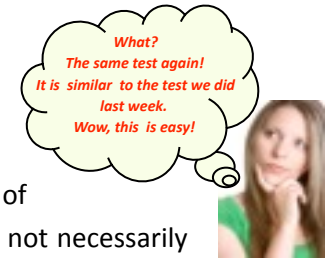
In the assessment of practical skills, it is possible to set very invalid questions without realising it. If a car mechanic is taught that a certain fault will give certain symptoms and if he is always tested in the '*fault to symptoms*' direction, there is no guarantee that he will be competent in operating in the real world where the direction is '*symptom to fault*'. He could be a prize-winning student but a useless practitioner.

And so valid means that the test measures what the setter intended it to measure in terms of knowledge, understanding and skills. If a test is invalid, everything which follows from it is useless. No amount of skilled marking or statistical manipulation will make the test of any value.

However, assuming that our assessment tools are valid, they have a second hurdle to surmount. Are they reliable?

More about Reliability

Imagine you are measuring my height. You find a measuring tape, you put me next to a wall or a door, make sure that I am not wearing shoes, and then measure my height. The question is how accurate is the result? If you measured my height tomorrow at the same time, would the result be the same? Is the measuring tape accurate or does it stretch over time? Do you read it correctly? With what level of accuracy?



It is exactly the same with any test you set. It may be valid: it is measuring what is intended. However, will your test produce similar results with the same students on two different occasions or with a similar group of students on the same course. Strictly speaking, the same test should give the same score for an individual on two occasions close together, but this is a state of perfection rarely achieved. A student prepared for a test on one occasion will not necessarily perform as well on another occasion. Intervening experience or common forgetfulness will play a part in altering performance.

Reliability can be gauged when the same test is applied to two large groups of similar students at the same stage in the same course. The average scores of the two groups could be similar and the spread of scores could be comparable. It is even possible that the average score on individual questions on the two occasions might be almost the same. This is particularly obvious when national examination boards, working with samples in the thousands, use the same questions in successive years. If the assessment is seen as a measuring instrument, it ought to be reliable. No one would use a measuring device, such as a thermometer, which gave different readings for the boiling point of water on different occasions.

However, a test or a question may be reliable and yet not be valid. We can test the wrong things but do it reliably. One can find written tests of laboratory skills which are reliable but which are invalid because those skills may be tested only by observing them at the bench.

Strictly speaking, it is not really the test that is reliable. Reliability relates to the *results* that arise from the test. The literature is full of all kinds of ways of measuring reliability using statistical methods. These sound very good but most tell us almost nothing and some are simply wrong! There are simple ways to achieve reasonable reliability and some are listed in Reid (2003)[56]. Most are practical common sense.

A test will tend to give reliable results:

- ☑ With a large number of candidates;
- ☑ When applied under good conditions (candidates are properly prepared, the time is adequate, the test environment is satisfactory);
- ☑ Cheating is minimised;
- ☑ Marking follows agreed marking briefs.

More about Humane Assessment

Sometimes, the length of tests and examinations is so great that they are tests of stamina and not educational measurements. Indeed, there is a skill in handling tests and examinations that has little connection with ability at subjects. This particularly true of '*final*' examinations. Statistically, it is true that the more questions we ask, the more likely is the assessment to be a true reflection of the student's overall ability, but we have to stop some time! Flesh, blood and brain have limits. Exhaustion is not the best physical or mental state to encourage assessment success.

In fact, using periodic assessment (sometimes called continuous assessment) is not only much more humane but it enhances the reliability of the results. Using several short tests instead of one long examination gives a far more accurate picture of student successes. It is also much more humane on the students. Inevitably students have good days and bad days. Inevitably, the end-of course, big examination creates student stress. This is not humane and may seriously affect performance.

It must be borne in mind that no course is an island. The assessment in your course may be well constructed and humane, but students are simultaneously doing several courses and the cumulative assessment burden may be overpowering. This is particularly so in the new trends to modularise courses which have formative tests throughout and a summative examination at the end. As many as six courses may be running in parallel and teachers seldom confer with their colleagues to harmonise the assessment procedures.

---

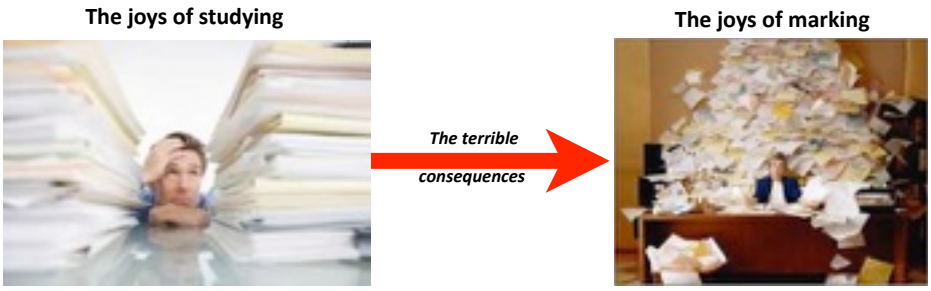
[56] Reid, N. (2003) *Getting Started in Pedagogical Research in the Physical Sciences,* LTSN Physical Science, Higher Education Academy, Hull, ISBN 1-903815-07-X, [online: http://www-new2.heacademy.ac.uk/assets/ps/documents/practice_guides/practice_guides/ps0076_getting_sarted_in_pedagogic_research_in_the_physical_sciences_aug_2004.pdf

More about Assessment Economy

There has to be a balance between the time and resources spent on teaching and on assessment. There is no simple formula to determine what this should be, but there are some facts which help in making a decision about the types of assessment to be used. The form of the assessment is related to the cost in administering it.

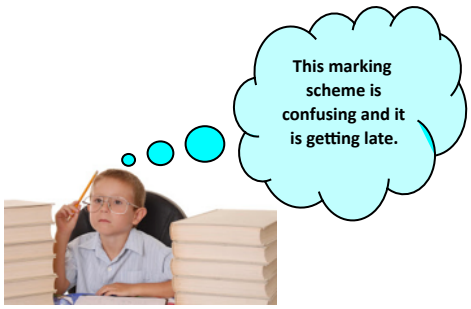
If we consider the extremes of assessment typified by essays or long answer questions on the one hand, and multiple-choice or other fixed-response-questions on the other, a comparison of costs in terms of time and money can be made.

The setting of essay-type questions can be fairly easy, consisting of a sentence or two of a quotation and some instruction. It is almost possible to write such a question on the back of the ticket while travelling by bus! They may not be good questions by any standards, but they are common enough and they are cheap to set.



As far as responding to such questions is concerned, they take a lot of time. Examinations typically last one or two hours and students have to write furiously to cope. However, the cost in teacher effort comes at the marking or grading stage. Hours of tedium, pints of sweat and gallons of midnight oil are expended in this exercise and, if costed realistically, this is frightening. Administrators, of course, hide this cost behind the good-will of the long-suffering teacher, but a price has to be paid in terms of neglect of teaching by jaded teachers.

At the other extreme, we have the fixed-response questions, typified by multiple choice. Let us treat them in the same way as we have done for the extended-answer questions. Contrary to popular belief, they are extremely difficult to set. As we shall see in later sections, they need a great deal of skill and time to prepare. It is not just a matter of one correct answer and three other absurd options. To make these questions work, all the options must be plausible and it is not easy to find them.



In terms of administration, the tests can be fairly brief: typically about one hour. The grading can be carried out by machine and the statistics handled automatically. However, someone has paid for the machine and someone has spent a long time programming the computer to obtain the data.

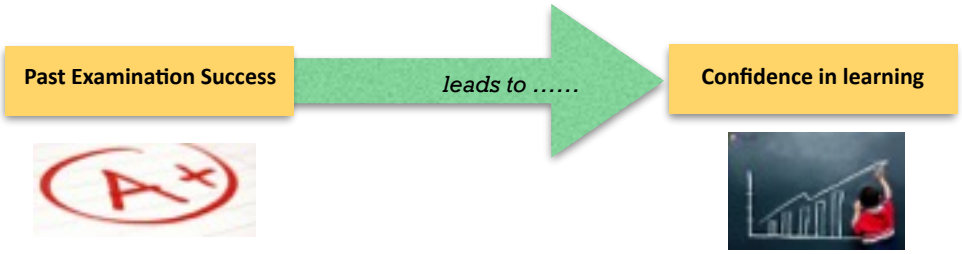
More about Assessment Benefits

It is easy to ***say*** that assessment is carried out for the benefit of the learners. It is extremely hard work for the teachers in both setting, supervising and marking. However, do the learners actually gain any benefits and what problems can assessment data bring for teachers, schools, or universities?

Let us be honest. Assessment can do enormous damage as well as bringing benefits. Here are some outcomes that can occur.

Group	Benefits	Disbenefits
Students	Shows how well they are doing	Can be a 'certificate of failure'
	Can encourage	Can dishearten
	Can show strengths and weaknesses	Lack of constructive feedback leaves uncertainty on how to improve
	Can give entry for jobs or universities	Can deny entry for jobs or universities
Teachers	Student successes can encourage	Student lack of success can disillusion
	Teachers share student and parent happiness	Teachers feel guilty, usually without true reasons
	Hard work and assessment effort seems worthwhile	It all seems a waste of time and effort
		Comparison between schools can destroy teacher morale
		School inspections can use examination data to blame teachers
Schools	Successes bring positive sense of achievement	Lack of student successes can dishearten
	Schools can publish their successes	Comparison between schools can destroy teacher morale
		School inspections can use examination data to blame schools

Research has shown that academic confidence (confidence in learning) is almost completely controlled by past successes in tests and examination. Thus, poor results may leave a student with a deep sense of failure and a lack of ability to try in the future. Confidence has been sapped. In an ideal world, we want to devise an assessment where no one fails! This means that we just have varying degrees of success. There have been some attempts at this, with a measure of success. However, it is not easy. One study revealed that academic confidence was dependent on academic success or perceived academic success[57].



The real problem with assessment comes in the modern trend to use school examination results to compare schools or to compare teachers. Indeed, there are many assessments now being used to compare countries. At the outset, it has to be stressed that ALL such comparison are TOTALLY INVALID. National examinations were designed to give a measure of student successes. Using them for another purpose is almost inevitably totally misleading.

[57] Oraif, F. (2007) *An exploration of confidence related to formal learning in Saudi Arabia*. PhD thesis, University of Glasgow, Glasgow.

Looking at comparisons between schools, here are some of the problems:

- The natural genetic abilities of two school populations will NEVER be the same and, therefore examination outcomes will reflect this, and not reflect qualities of the schools or their teachers.
- Different schools develop different educational cultures, often developed to meet the needs of their own students in their own communities. This can affect examination outcomes but bring enormous educational benefits for students.
- The buildings, equipment and general resource levels vary from school to school and can affect examination outcomes. School and their teachers have no control over this.
- It is easy to show that, in any school, examination outcomes will vary widely from year to year simply by the random nature of populations. No school or teacher can ever be assessed fairly on the examination results for one year on its own.
- It is also easy to show that schools which '*spoonfeed*' and support their students most in order to gain the best examination outcomes generate students who have greater difficulty coping in the freedom of universities and colleges where such supports cannot be offered.
- Examination outcomes are only ONE measure of student success. Can we measure lifelong learning skills, thinking skills, responsibility, courtesy, preparation for the workplace, diligence and determination …. any many more wonderful skills, using national examinations. Yet in the long run, these may be much more important.

Using examination data to compare schools and teachers is
COMPLETELY INVALID

These five areas can be expanded and illustrated:

Assessment Moving *into* a New Future

Process-Product in Assessment

Aims

This chapter explores how assessment can be seen in terms of the '*educational product*' but also in terms of the processes that learners go through in achieving their goals

Introduction

One of the purposes of education is the promotion of the intellectual, personal and moral development of young people. To a very large extent, this depends on the curriculum (which is determined outside the schools), the assessment system (which is also determined outside the schools) and the way the material is presented (this depends on the teachers and the resources made available to them). Thus, the key aim of promoting the intellectual, personal and moral development of people may be achieved in several subjects but goes well beyond the individual boundaries of these subjects.

The practice of assessment clearly serves multiple purposes in education settings including *certification and regulation, selection, direction, motivation, accountability and learning*[58]. Clarifying the purpose of an assessment task, episode or programme establishes the intent of the practice as well as the expected outcomes or consequences of the practice[59].

One outcome of assessment is the awarding of grades at the end of courses, a measure of student achievement. Another intended outcome of assessment that has gained increasing prominence in general educational discussions is the promotion of learning itself[60]. That is the view that teachers (in this context this term applies to the school, college or university instructor or teacher) would engage students in assessment tasks that promote the acquisition of understanding and skills that they may not already possess, or would require students to successfully adapt and utilise previously acquired understandings in new contexts.

This has led to a wider paradigm in thinking about assessment: processes and products.

---

[58] Broadfoot, P., and Black, P. (2004) Redefining Assessment? The First 10 Years of Assessment in Education, *Assessment in Education*, 11(1), 7–27.
Hargreaves, A., Earl, L. and Schmidt, M. (2002) Perspectives on Alternative Assessment Reform, *American Educational Research Journal,* 39(1), 69–95

[59] Boud, D. (2007) *Reframing Assessment as if Learning Were Important*, In Rethinking Assessment in Higher Education: Learning for the Longer Term, edited by David Boud and Nancy Falchikov, 14–26, Abingdon: Routledge.
Hayes, D., Mills, M., Christie, P. and Lingered, B. (2006) Teachers and Schooling Making a Difference, Sydney: Allen & Unwin.

[60] Boud, D. (2007), ibid.
Hargreaves, A., Earl, L. and Schmidt, M. (2002), ibid.

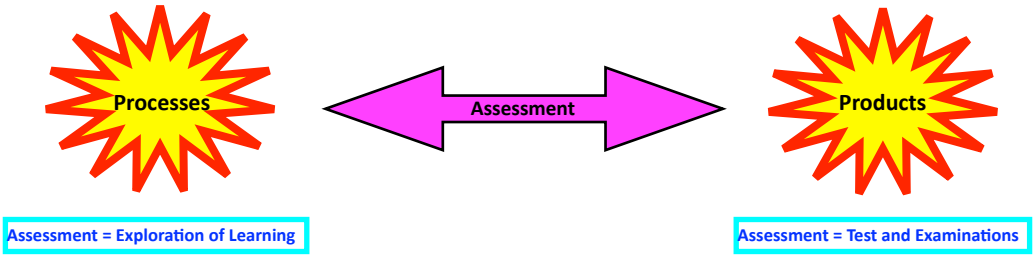


Figure 7.1 Assessing Process and Product

This can be illustrated further:

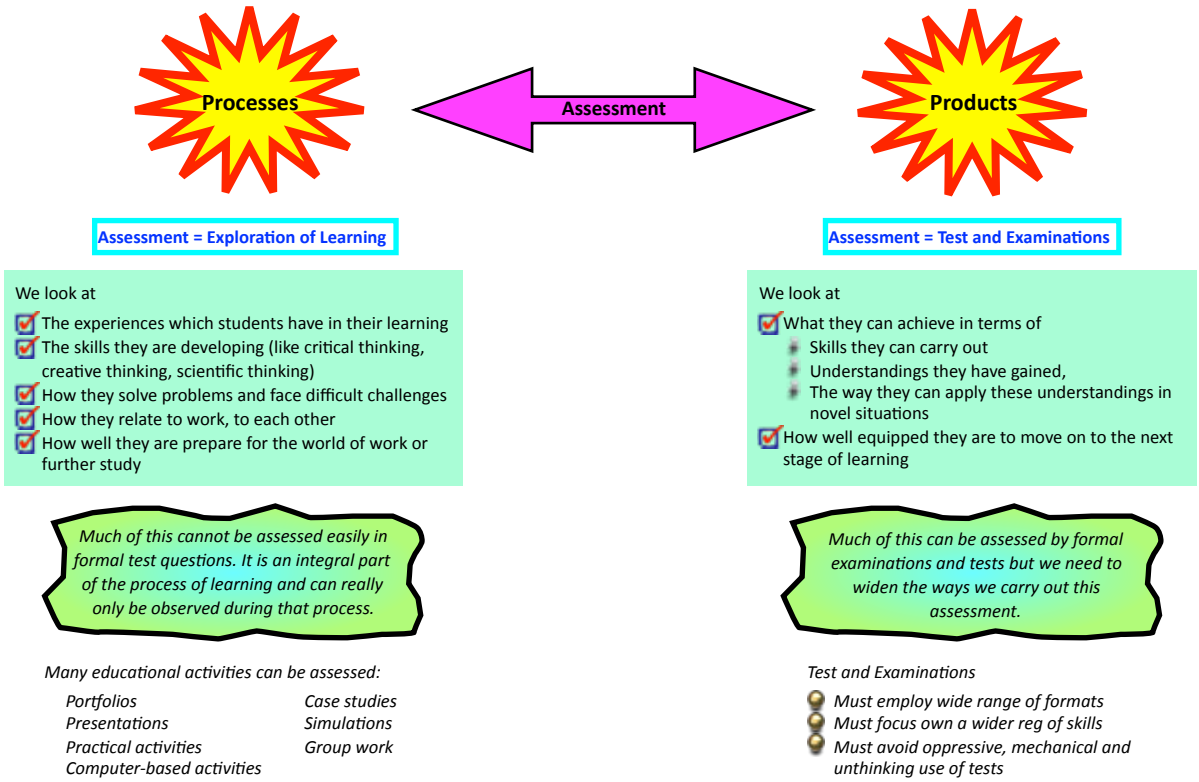
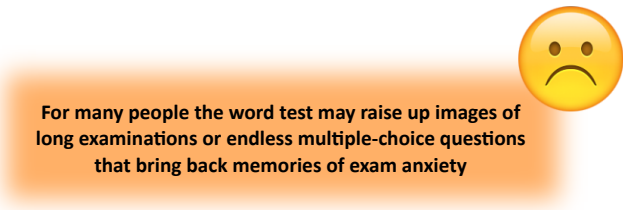
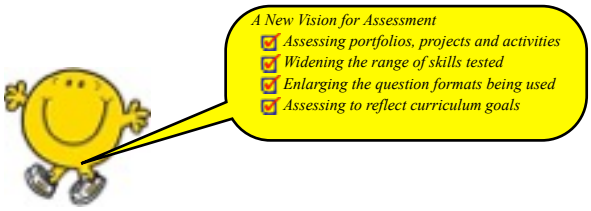


Figure 7.2 More about Process-Product Assessments

The Sad Picture



Examination Boards need not depend only on end-of-course formal examinations. There is a place for the assessment of projects, portfolios and practical activities along with a widening of the scope for the final examination papers to test a much wider range of skill using a much wider range of assessment formats. All of this has been carried out successfully in various places in the world. There is a better future .....



The practical ways by which this can be achieved will be outlined later.

The overall picture can be seen as:

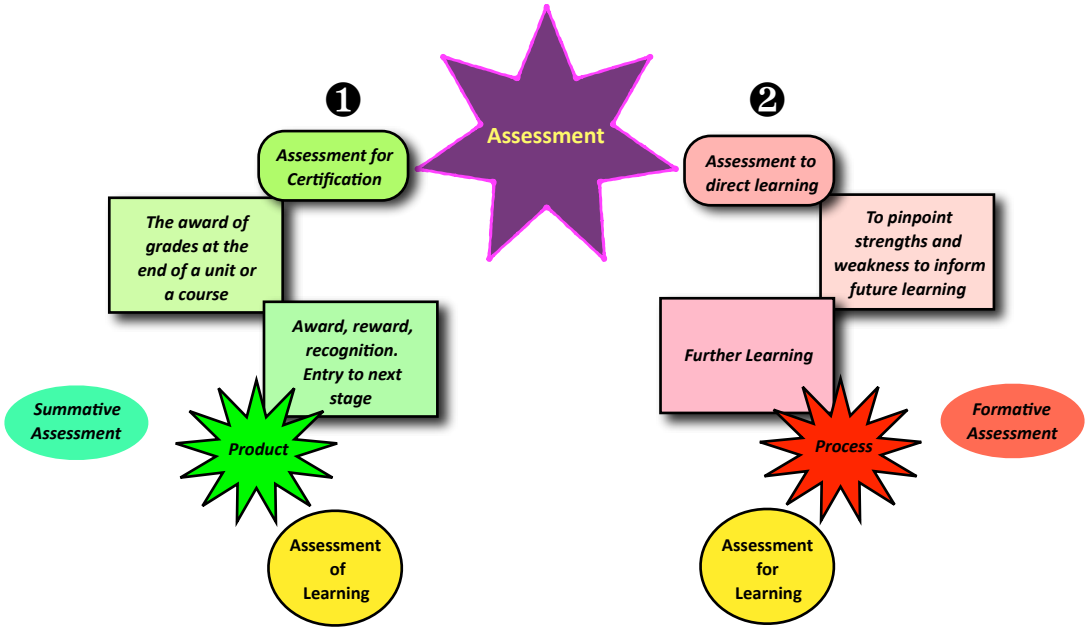


Figure 7.3 Assessment for learning

In the end-of-unit or end-of-course assessments, the assessment seeks to find out the extent to which the learner has achieved the outcomes for the unit or course. This may give the learner a reward or even an award certificate. It may give access to more advanced courses. This type of assessment is often called *summative* and it tends to focus on the *final product*.

Confusion

Summative assessment is used to award marks or grades.
Formative Assessment is to inform learning

but

Summative assessment can take place throughout a course:

(a) Essays, projects or portfolios to give course credit
(b) Tests, say, half way through a course

Summative is defined by how the test data are used: **to give credit**

Formative is defined by how the test data are used: **to inform learning**

Assessment can also be used to pinpoint strengths and weakness in the learning process. It can direct future learning and suggest what the learner needs to do to make progress. It does not lead to any certification. Indeed, it may give no mark or grade. It is often called *formative* and it tends to focus more on the *processes of learning* and not just the final product.

It has to be stressed: both types of assessment are important. However, most assessment tends to be related to certification. Indeed, research shows clearly the dominance of this and the way we set assessment questions is strongly influenced by the needs for certification. In other words, we tend to copy the question style used in national certification and use that in our in-school assessments.

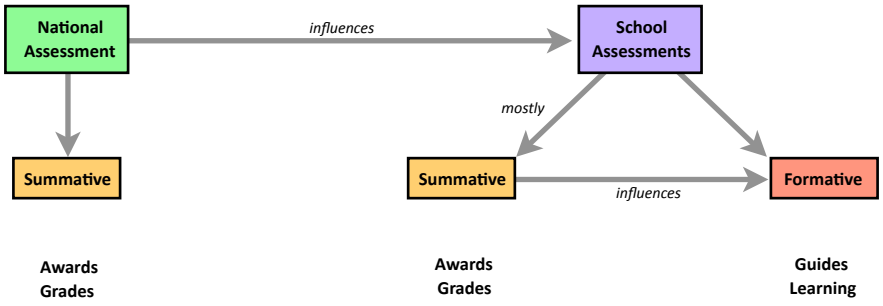
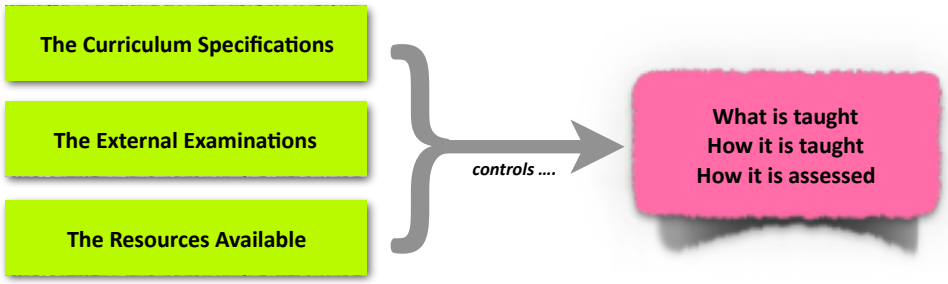


Figure 7.4 The way national examinations control everything

The tragedy is that national examinations determine the way teachers assess in schools simply because teachers have to prepare their students for national examinations. The key to changing school assessment lies outwith the schools: reformation of the national examination system.

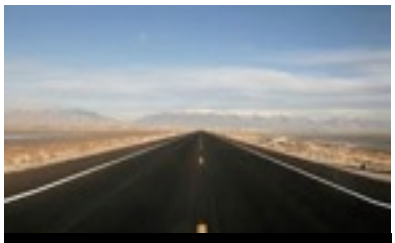
Re-thinking Assessment

Assessment needs to be seen as an integral part of the entire learning process. However, what goes on in schools is controlled almost entirely outside schools !



If things are to develop in schools, then there have to be changed curriculum specifications, changed external examinations and adequate resources for the task.

Teaching is ***no longer*** seen as lecturing. Teaching does ***not*** involve the transfer of information from the head of the teacher to the heads of the students. Teaching is seen as an exercise where, as teacher and learners together, we develop understandings of the world around. Assessment goes on all the time - it is integral to the whole business of learning. We need to know what and how our students are understanding. We need to be able to engage with them as their minds battle to understand difficult concepts.



We are all on an educational journey together

The only difference between students and their teachers is that the teachers have lived longer

We need to:

- ☑ Start from a learner's existing understanding.
- ☑ Involve the learner actively in their learning process;
- ☑ Help the learner to understand the learning aims and the criteria of quality, so enabling self and peer assessment.
- ☑ Support and guide social learning through discussion.

Conclusion

Assessment needs to be seen by the students as something that helps them. It is not a matter of them sitting an examination and feeling rewarded or rejected when the marks are awarded later. It is not a matter of teachers sitting in judgement, awarded '*passes*' or '*fails*'[61]. Many classroom assessment events occur in typical classrooms, all intertwined. The overall sense of expectations that these events build up, the meaning or sense that students make out of this aspect of their classroom, composes the classroom assessment environment[62]. In other words, the students feel that the assessment is not a matter of judgement or condemnation. When this is achieved, at least in part, then then the end-of-course examination board examinations become much less daunting.

Another important issue is the way examination board examinations are taken entirely at the end of a course. There is a place for in-school assessments which are conducted by the teachers and monitored by the examination board. Such assessments can give credit for a wide range of skills, skills that are often very difficult to assess in formal written '*end-of-course*' examination papers. The practical ways to do this will be outlined later but this kind of practice is common in many countries and works well. Indeed, in one country, the entire process of national accreditation is entirely in the hands of the schools and their teachers: standards are high and the system is widely applauded.

[61] Stodolsky, S.S. (1988) *The subject matters: Classroom activity in math and social studies*, Chicago: University of Chicago Press.

[62] Stiggins, R. J., and Conklin, N. F. (1992) *In teachers' hands: Investigating the practices of classroom assessment,* Albany: State University of New York (SUNY) Press.

Assessment Moving *into* a New Future

Criteria for Success

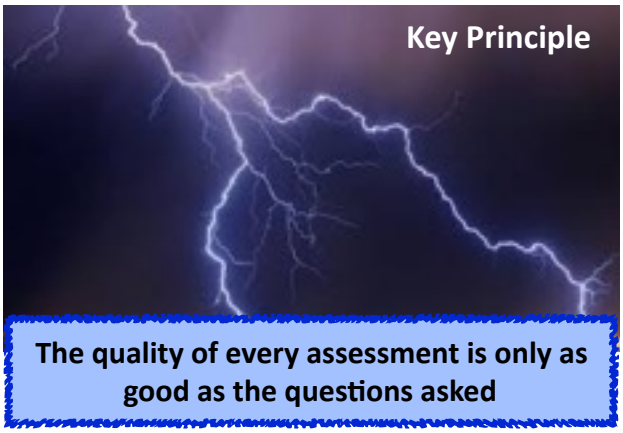
Aims

This chapter outlines how to use assessment successfully so that learners are assessed robustly and fairly in such a way that the right achievements are rewarded and learning is supported.

Introduction

It is easy to see that learning success must be rewarded by good assessment grades. However, it is even more important that the right skills are rewarded. Far too often, the rewards simply are given for correct recall of information and procedures. There is little or no recognition of skills of understanding, thinking or evaluating. If a high grade is to mean anything in the wider world, it must indicate that the successful learner is equipped to make a contribution in future life, in the workplace, and in the creation of new developments for the benefit of wider society. For all of this, the reward of recall is largely irrelevant.

Every examination or test is made up of assessment tasks. Often, we call these questions. The quality of every assessment is only as good as the questions asked. It takes very great skill to develop assessment tasks which are really good. Such tasks are designed to allow the very able to perform well and to show their abilities in understanding, thinking, evaluating and, indeed, employing their understandings in constructive ways.



In simple terms, assessment is ***not*** easy and good assessment will be incredibly demanding for you as a teacher. The aim of this chapter is to give a set of seven key questions which will help you develop really good assessment tasks.

The whole process of assessment can be approached by asking a set of seven questions:

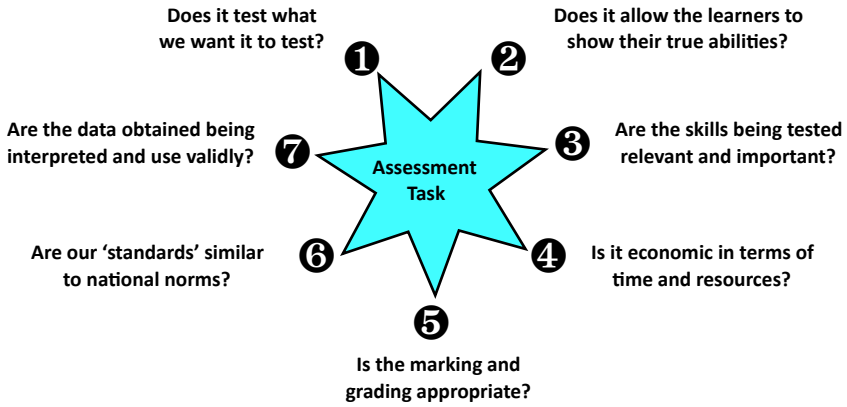


Figure 8.1 Assessment Task

As assessment is central to the recognition of achievement, the quality of the assessment is therefore important to provide credible information. Credibility in assessment is assured through assessment procedures and practices being governed by certain principles. Let us now look at these five key questions in more detail.

What are we testing?

The fundamental question is whether our tests and examinations test what we want. Another way of looking at this is to ask if the students are gaining their rewards for showing the skills that are important. In chapter 3, we discussed this in some detail. We suggested five broad areas that are the key for assessment:

- What they **know**
- What they **understand**
- What they can **do**
- How they can **think**
- How well they can **evaluate**

Let us extend the list by adding two more important skills, both vital for modern living:

- How well they **communicate**
- How well they can **relate their studies to society**

The descriptions can be extended as well:

Broad Areas	Working Description
Know	What the student knows (facts, concepts, skills) or can be accessed.
Understand	Described in terms of the extent to which the student can apply their knowledge in novel situations with some prospect of success.
Do	Skills (practical or procedural) which the student can demonstrate successfully.
Think	The extent to which students can think creatively, critically or scientifically in relation to the material being studied.
Evaluate	The extent to which the student can ask the questions *why? what?* and *how?* of new information, its sources and the way it links to which is already known.
Communicate	The extent to which they can demonstrate that they can communicate their ideas and understandings coherently and logically
Relate	The extent to which they can see how what they have bene stymieing relates to the major practical issues of modern living

Now stop and think about the last test you set. Be honest. Look at the questions you asked. Probably, your students gained most of their marks simply by recalling memorised information, memorised '*interpretations*' or memorised procedures. What about genuine understanding where your students can use their knowledge in a novel situation? Do your students ever have to ***think*** in gaining their answers or was it simply a measure of recall? Did they have look at what they know and understand, and then evaluate it in terms of relevance, importance or meaning?

The best way to find out how your students gained their answers is to talk to them and ask them how they found the answers they did. Sadly, you will almost certainly find that recall is the dominant feature.

In the next chapters of this monograph, we shall look at specific examples of questions which can explore some of these more important skills. However, at this stage, it is worth noting that multiple choice questions rarely measure anything beyond recall. Fill-in questions almost certainly measure nothing more than recall while questions that ask for facts or information are measuring nothing more than recall. We need to break out of this straight-jacket and look at some of the exciting new ways that have developed over the past few decades.

Does it allow the learners to show their true abilities?

Have you ever set an examination or test which was too easy? Or too demanding? A test that is too easy does not allow the very able to show their true abilities and skills. A test that is too hard may cause many to panic and not perform as well as they should and some will even give up.

The problem is to know how to set questions at the right standard but there are two things that are critical in setting the right standard:

- **Your knowledge of your students - you just know what they can do.**
- **Your experience - we all gain immeasurably from our experiences in examination setting.**

This is why examinations set by outsiders (those who have never been teachers or those who have been away from the classroom for some time) are often inappropriate. Such examination setters do not really understand the way young people learn and the way they understand. They frequently do not understand the practical educational needs of young people!!

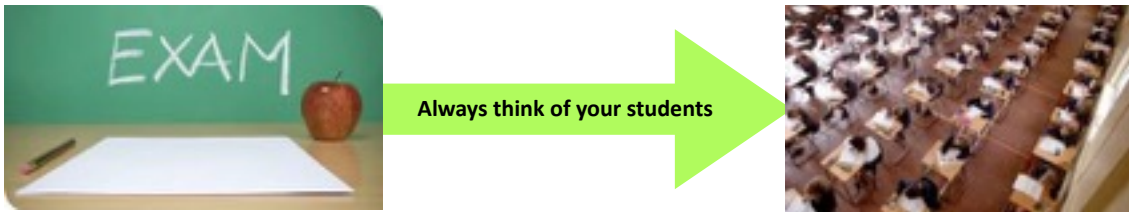
Setting test and examination questions and tasks is NOT easy. Indeed, setting GOOD test and examination questions and tasks is VERY DIFFICULT. Because of this, we tend to set questions which are easy to set but the skills being tested may be not so important.

We start by looking at the curriculum we are to teach. We look at its stated goals and aims. We use our own judgement on these, for curriculum writers often state very unrealistic goals. The tragedy is that curriculum writers and planners are rarely practising teachers and they lack the reality of knowledge that only practising teachers possess. It is interesting in one recent study in Pakistan that the curriculum developed *within* the school in one subject was found to be much better than that imposed by those *outside* the school[63]. This is frequently the case.

At this stage, you need to sit down and write down your understanding of the key skills, understandings and insights that *you* think are the most important ones for your students. You then develop test questions to test these skills. This is what is known as setting a ***test specification***. It is important that you base it on skills (including practical skills, understandings, applications, evaluations) and do not base it on content and subject matter. Remember that most of the subject matter you teach your school students will be forgotten within a few years. However, the skills they gain may well last a lifetime.

Fundamentally, you want to reward a balance of the most important skills related to your subject area. You do not want to reward recall. You do not want to reward things simply because they are easy to measure in formal examinations. We aim to educate, not to fill the heads of our students with endless catalogues of information. We aim to educate and reward what is important for our students for their future lives.

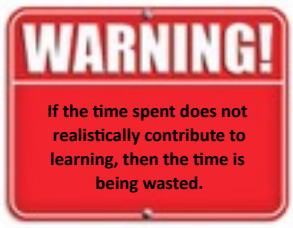
Indeed, it all comes down to placing our own students at the fore-front of all our planning, especially our planning of assessment.



Within the vast industry of national examinations in almost every country throughout the world, there is a real danger that the student gets lost in a process. Let us put the student back at the centre. When assessment does not bring benefit to the student in terms of their futures and their ways of learning, then the assessment system has gone wrong. When assessment leaves the students dissatisfied, feeling that they have been forced into endless memorisation and recall, with no opportunity show their real skills, then assessment has failed.

[63] Ali, A. A. (2008) *Perceptions, Difficulties and Working Memory Capacity Related to Mathematics Performance,* MSc Thesis, University of Glasgow, Glasgow. [http://theses.gla.ac.uk/441/]

Is it economic in terms of time and resources?



Tests and examinations cost time. Time costs money. The time you spend setting and marking examinations is time not available for you to prepare, to teach, and to develop your skills in other areas. The time spent by our students preparing for examinations may be very large while they spend much time sitting them.

It is always much better to set short tests and examinations at various points throughout a course rather than put all the effort into one end-of-course examination. Firstly, multiple assessments give greater accuracy overall. Secondly, it reduces student panic. However, the main point here is that spreads out the workload for yourself and makes your life somewhat easier.

Always ask yourself: Is the assessment I am planning going to bring benefit to my students? Will it enhance their learning? Is it cost-effective? Would my time be better spent on something else?

Is the marking and grading appropriate?

Setting and sitting examinations and tests is one thing. For teachers, marking is perhaps the worst task you have to undertake. Sometimes, you will come away from the task feeling utterly despondent, perhaps saying to yourself, '*Did I teach them anything*?'

As a task, marking is highly repetitive, boring and we see ourselves as creatively totally unproductive. We become tired and jaded and, perhaps, we are not marking consistently. Are we allocating marks fairly? Are we allocating marks consistently with the way other teachers are allocating marks. Do our students get the rewards they deserve?

Here are some simple ways to help us:

- Recognise that completely consistent marking is a practical impossibility. Our goal is to minimise the inaccuracies and inconsistencies.
- Always develop a marking brief *before* you start. Indeed, it is often better to develop the marking brief when you are setting the paper.
- A good marking brief will show how each mark is to be awarded in the entire examination paper.
- Always discuss your proposed marking brief with colleagues and re-adjust it in the light of their comments.
- If several different teachers are to mark scripts in the same examination, make sure there is a '*markers meeting*' before the marking starts. At this meeting, all the markers will need to ***agree*** on the marking brief so that everyone awards marks in the same way.
- After the papers are marked, a sample (maybe 10%) of papers should be '*double-marked*' so that the marking standards are checked.
- Wherever possible, use the examinations as a mechanism for teaching and learning. Give back the papers to your students and go through the entire paper with them, showing how you marked it. This is a wonderful learning opportunity. Allow students to approach you afterwards if they think you have not awarded the marks as you explained or if there is a totalling error. Do not be upset (and do not put the students down!) - we all make mistakes in marking and there is no loss of face in admitting this.

Are our 'standards' similar to national norms?

This is much more difficult and the only way forward is by means of experience. We see how our students perform in national examinations and we start to relate this to how well they performed in examinations we set. However, it is not even that easy.

The national examinations may be marked according to criteria that are different from those you use in your schools and it may be impossible to explore this. In other words, the national examination markers may be giving credit for things that you do not think are so important, and vice versa.

The only solution lies in experience. We look at national examination papers, we talk to our students after they have sat papers, we talk to other teachers and, best of all, we talk to those who are examiners and markers at national level.

In some countries, they have a brilliant policy that all national examination papers are marked by practising teachers. In the best system, every marker MUST attend a one-day markers meeting before they start marking. Here, the paper setter presents a proposed marking brief. This is discussed and amended by the team of markers and the final amended marking brief is used rigorously by all the markers. Typically, a sample of about 20% of the papers are re-marked to ensure marker consistency.

Such a system works very well and has the added advantage that many school teachers gain experience in how the national system of examinations works and the basis by which papers are set and marked. This means that it is much more likely that such teachers will set and mark their own in-school tests and examinations to a similar standard to that used nationally.

Finally, it is worth considering how our examinations relate to international patterns. It must be stressed that using international comparisons is meaningless. Such comparisons tell us very little for the examinations set by the various organisations who do this cannot control for very important variables. Different countries have different curriculum emphases, different patterns of testing, different styles of test questions. The subtleties of language add a further layer of complexity. Overall, each country has developed its own education culture and assessment culture.

This puts some countries at a great disadvantage in such international examinations and some countries at a great advantage. The setters of such international comparison examinations do study the different curricula for different countries but all that does is take content into account. It cannot allow for education culture and assessment culture.

Are the data obtained being interpreted and use validly?

We have set the examinations, the students have sat them, we have marked them. The tragedy is that the marks or grades we award can then misused by others. Of course, we are all well aware that the marks in, say, an examination in mathematics cannot be used to indicate how good a student is at, say, English. However, many people use marks obtained for one purpose to indicate success or failure in another.

It is an interesting study to look at how well school national examination marks relate to later success at university. From several studies in developed countries, it has been found that we cannot predict university success very well at all from school grades[64]. Indeed, the relationship is so poor that it raises the question on why we use school examinations to determine university entrance. Part of this reflects the very different way universities teach, compared to schools.

Part of this also reflects that school students have gone through adolescence. They need to develop their own good ways of working and, often, this goes on well into university life. Thus, they may change quite markedly in the way they study when comparing school to university. The important point is that any school examination is one very imperfect measure of student performance at one particular time under one particular set of circumstances. Thus, the predictive power of the outcomes is not very great.

Any school examination is one very imperfect measure of student performance at one particular time under one particular set of circumstances

Politicians and educational managers often use the marks obtained in one school to indicate how well the school and the teachers are teaching. In reality, the national examination marks were designed to indicate how well the school students performed. They say ***nothing*** about the quality of the school and its teachers. Doing this is a simple abuse of data and may lead to ***very misleading*** outcomes. Indeed, there is more or less no academic evidence for the value of school leagues tables, no evidence that genuine quality is enhanced by them, but much evidence of unintended consequences which have proved disastrous [65].

It is a bit like comparing two cars only by looking at their petrol consumption. One car travels 11 kilometres for each litre of petrol used while another travels 14 kilometres for each litre of petrol. It would be utterly naive to suggest that the second car was better than the first. Such a comparison ignores the fact that the cars do not cost the same, the former is designed for longer distances while the latter is designed as an urban run-about. Quite apart from that, the first car might be coloured grey while the second might be yellow, giving the latter a much higher visibility and, hence, safety rating. Of course, there are not so many who like yellow cars! We could go further. The first car might have an engine which lasted much longer but required more expensive maintenance.

Impossible Comparison

The two cars were designed for very different purposes

[64] Johnstone, A.H. (1972) *An evaluation of new Scottish chemistry syllabuses*, PhD Thesis, University of Glasgow, Glasgow.

[65] Robinson, K. (2011) *Out of Our Minds: Learning to be Creative*, Capstone Publishing Ltd, Chichester. (He makes some very interesting observations of the effects in parts of the US).

Muir, R. (2013) *The impact of league table reform on vocational education in schools,* IPPR, London.

Croxford, L. (1999) *League Tables - who needs them?* Centre for Educational Sociology, Edinburgh.

ASTI Briefing (undated) *League Tables*, Association of Secondary Teachers Ireland, Dublin. [http://www.asti.ie/about-asti/policy/league-tables/]

HeraldScotland (Dec, 2011) *School league tables show extent of competition, not quality of education,* HeraldScotland, Glasgow

For a very brief summary of key landmarks and comments, *Are school league tables good or bad?* UKEssays.com [http://www.ukessays.com/essays/education/are-school-league-tables-good-or-bad-education-essay.php

For a rigorous review, see: Meyer, H. D., & Benavot, A. (2013). *PISA, Power and Policy: the emergence of global educational governance,* Oxford, Symposium Books.

The key point is that we cannot compare cars only on the basis of petrol consumption. This tells us something about the engine size, the power-to-weight ratio but it does not really tell us much about engine efficiency. It certainly says nothing about the overall quality of cars. Nor will such a comparison be helpful to anyone wanting to buy a car.

In the same way, we CANNOT use national examination marks to compare schools. They tell us something about student performance in a particular year in a particular set of examinations. Education is very much more than examinations and the quality of the students in a school cannot be judged solely on the basis of the examinations.

Cost - Very large
Effort - Considerable
Risk - Far too large

While student attainment cannot be judged simply by looking at examination grades, such grades say absolutely nothing about the teachers or the quality of the school. Different schools have different catchments, different resources, different community needs, different local priorities, along with vast differences in educational ethos, all for perfectly acceptable reasons. Making comparisons simply on the basis of examination outcomes is likely to lead to meaningless conclusions being drawn and the good work of many schools being denigrated or even destroyed. Just to add another layer of confusion, results from one year to another in any school can varying enormously, simply on the basis of random fluctuations in the quality of the student intakes.

Probability of drawing correct conclusions

Almost zero

Bringing it Together

There are seven simple principles which will help us to develop quality assessment:

(1) Make sure we are testing what we think we are testing.

(2) Make sure that the way assessment is used allows the students to perform to their best.

(3) Make sure that we test what is really important and of benefit for the students.

(4) Do not let assessment procedures take up too much time or have impossible resource demands.

(5) Take careful steps to ensure that marking is both fair and consistent for ALL students.

(6) Gain as much experience as we can in the way national assessment works.

(7) Do not allow examination data to be abused by educational managers or politicians or the media.

Assessment Moving *into* a New Future

The Brain and Assessment

Aims

This chapter summarises the research which shows how information is handled in the brain and how the brain works in controlling thinking, understanding, problem solving and assessment success

Introduction

The study of human learning falls within the field of cognitive psychology and it is a research area in which there have been significant developments in recent years. Cognitive psychology and neuroscience have changed the way in which we consider people's knowledge, memory and learning processes.

Education has always been awash with new ideas about learning and teaching while teachers are regularly bombarded with suggestions for reform. Teachers are asked to use new curricula, new teaching strategies, and new assessments. Despite calls to base the new ideas on the clear evidence of research, many of the new ideas are simply expressions of opinion, often leaving the teachers to attempt what is impossible.

It is important to lay out some clear evidence which is soundly based on research. We need some clear pointers. Fortunately, there is much excellent research which can give this to us.

What is learning?

Long ago, Gagné (1964) stated that: '*Learning is a change in human disposition or capability, which can be retained, and which is not simply ascribable to the process of growth*'[66]. Lahey (2004) noted that, '...*learning can be defined as any relatively permanent change in behavior, knowledge, and thinking skills, which comes about through experiences*'[67] . Reid (2008)[68] referred to learning as a process that leads to any change in behaviour not explainable simply by development. We tend to limit this more to the cognitive but it could be also the psychomotor, attitudinal, or the development of skills.

---

[66] Gagné, R. M. (1964) *Problem solving*, in A.W. Melton (Ed.), Categories of Human Learning. New York: Academic Press.

[67] Lahey, B. B. (2004) *Psychology: an introductio*n, 8th ed. Boston: McGraw Hill.

[68] Reid, N. (2008) A scientific approach to the teaching of chemistry, The Royal Society of Chemistry Nyholm Lecture, 2006–2007, *Chemistry Education Research and Practice*, 9(1), 51–19

Johnstone (1997)[69] brings much of this together:

> *'Learning is the reconstruction of material, provided by the teacher, in the mind of the learner. It is an idiosyncratic reconstruction of what the learner understands or thinks (s)he understands of the new material provided, tempered by the existing knowledge, beliefs, biases, and misunderstandings in the mind of the learner.'*
>
> Johnstone (1997)

The important phrase is '*idiosyncratic reconstruction*'. Learning is ***not*** memorising. Learning involves the learner seeking to make sense of new ideas and information. Learning has to give new understandings that are meaningful for the learner. Thus, assessment must measure the extent of meaningful learning.

What is Meaningful Learning?

The key researcher here is David Ausubel (1918-2008). He showed that meaningful learning involved the student relating new ideas and information to what the learner already held in the brain. Ausubel suggested the following three conditions for meaningful learning to occur (Ausubel and Robinson, 1969[70]):

In simple terms:

- ☑ What is learned must be linked on to what the person already knows and understands
- ☑ The learner must possess ways of thought to enable this linking to occur
- ☑ The learner must have a desire to make sense of what is being presented

We all hold many ideas and understandings in our brains.

> '*Meaningful learning occurs when the learner's appropriate existing knowledge interacts with new learning. Rote learning of the new knowledge occurs when no such interaction place*'
>
> West and Fensham, 1974[71]

Ausubel *et al.* (1978)[72] have offered clear insights into the meaning of meaningful learning. In such learning, ideas are linked together in a complex matrix in the brain so that the learner can make sense of how it all fits and can then apply the knowledge in new situations in life. Thus, Johnstone (1997) described meaningful learning as, '*Good well-integrated, branched, retrievable and usable learning.'*[73]

Again, there is a key phrase here: '*usable learning*'. Meaningful learning can be employed by the learner. It can be applied, often in a novel situation, with success. This gives us another key for assessment. Assessing meaningful learning means that we have to assess how well our students can *apply* their understandings in situations which are new to them. This is **TOTALLY DIFFERENT** from the old tradition of testing recalled information.

[69] Johnstone, A.H., (1997) Chemistry Teaching-Science or Alchemy?, *Journal of Chemical Education,* 74(3), 262-268.

[70] Ausubel, D.P. and Robinson, F. G. (1969). *School Learning: an Introduction to Educational Psychology*. Holt, Rinehart and Winston, New York.

[71] West, l.H.T. & Fensham, P.J. (1974) Prior Knowledge and the learning of science: a review of Ausubel's theory of this process, *Studies in Science Education*, 1, 61-81.

[72] Ausubel, D.P., Novak, J. D. and Hanesian, H. (1978) *Educational psychology: a cognitive view*. 2nd Ed. Rinehart and Winston, New York.

[73] Johnstone, A.H. (1997), ibid.

Visions of Learning

Studies of the ways teachers see learning and assessment give rise to three broad ways which exist today. None of them is inherently superior. However, we need to look at them and see where we might fit most comfortably.

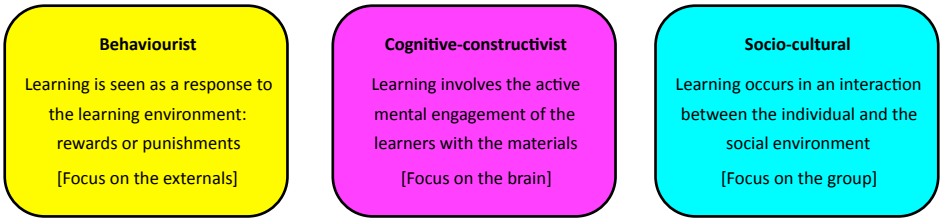


Figure 9.1 Three Perspectives on Learning and Assessment

Each of the above perspectives can be used in specific situations: for example, the first when routine skills have to be mastered, the second when we are dealing with conceptual areas, the third with areas involving community ideas, values and judgements. Indeed, all are true in all areas to some extent and can be encompassed within an information processing model of learning (see later).

Visions of Assessment

Now let us see how the three ways influence the way we develop and use assessment. Research has found the following kind of picture:

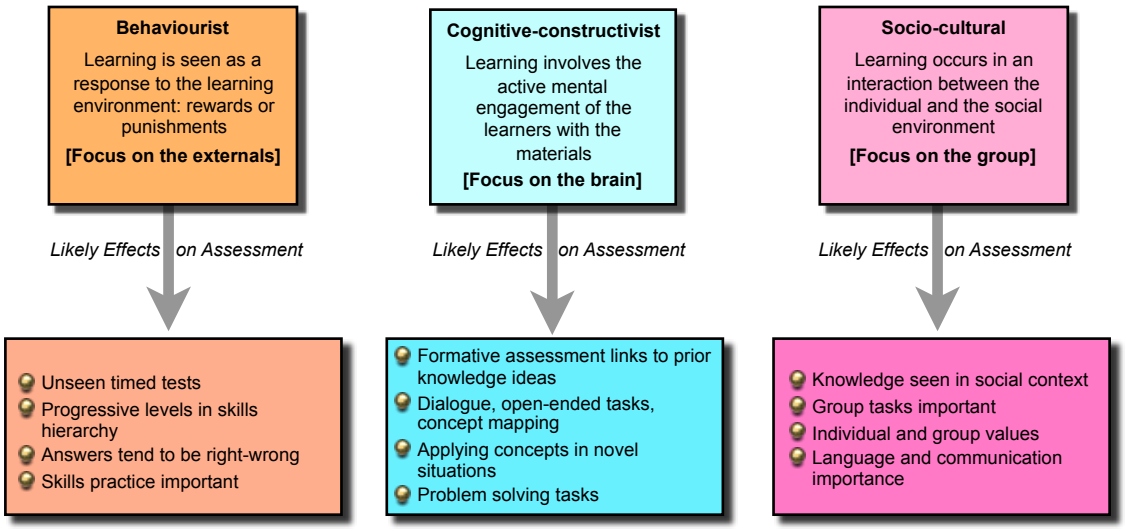
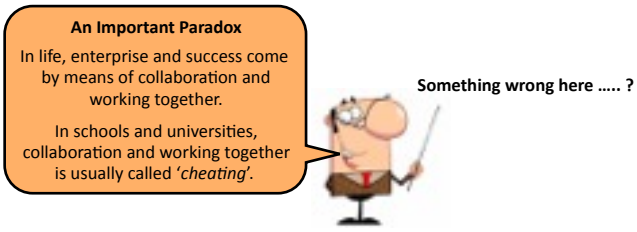


Figure 9.2 Description of learning approaches

Now think of the way you tend to assess your learners. Your discipline background and your own personal perspectives will strongly influence the way you see learning and this will, in turn, strongly influence the way you tend to look at assessment. However, there is probably a place for all three paradigms in all subject areas although the relative emphasis may vary considerably.

Behaviourism focusses on the externals. Thus, teachers teach specific skills and assessment measures these skills. There is little attempt to understand how learning actually occurs. The third perspective (socio-cultural) focusses on the group. This recognises that learning is fundamentally a social phenomenon. People organise their learning around the social communities to which they belong. Indeed, life involves the interactions between people and assessment under this perspective emphasises the social nature of learning and achieving.



It is the middle perspective that emphasises what goes on in the brain. It is the paradigm in the middle (which we have labelled, '*cognitive-constructivist*') which causes some confusion. This confusion tends to arise from the idea of constructivism.

Understanding is Constructed

Research has shown consistently that learners construct their own understandings of the world around. These understandings may be consistent or inconsistent with the generally accepted understandings.

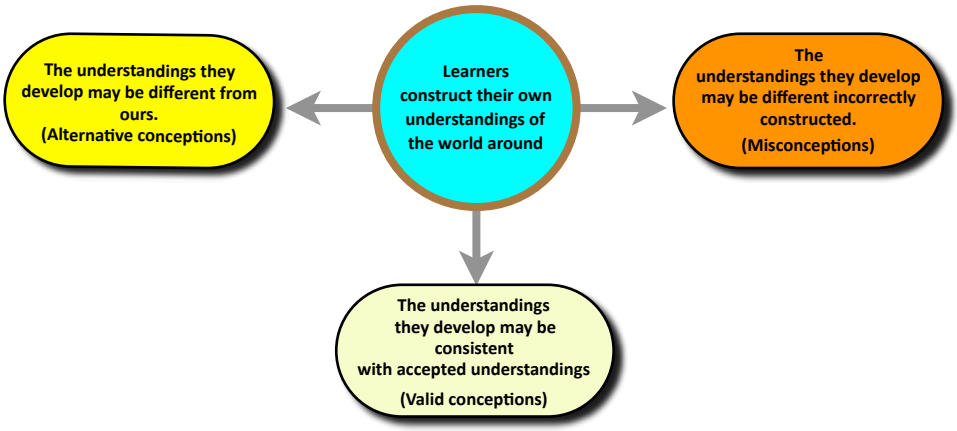


Figure 9.3 Cognitive - constructivist

Constructivism sees learning as a process of *personal* construction of understanding and the making of meaning. Research supports this unequivocally. Of course, others may well be involved but the meaning is essential in the head of the individual. Given that, how do we assess?

There is a naive understanding of what is meant by constructivism. At times, it is almost as if it is implied that ***all*** meanings which are developed by learners are ***equally*** valid. This cannot be so. Within society, there are ***accepted*** understandings. Of course, it is important to note that interpretations can vary legitimately. However, for example, the meaning of differentiation is well established, the rules for computer programming are agreed, there are accepted rules for grammar and syntax, and the understandings of the physical and biological world are agreed in considerable measure by the communities involved and the methods of handling historical evidence are widely agreed. It cannot be an assessment free-for-all.

There is considerable scope for valid interpretation and coming to variable reflections of significance. Nonetheless, the approaches for developing legitimate conclusions are largely well established and the evidence on which such understandings are to be based can be established with a fair degree of certainty.

The other weakness in most constructivist literature is the confusion between learning and teaching. ***All learners, naturally, seek to construct understandings***. This is nothing to do with teaching and the idea of *teaching constructivistically* is often discussed without any recognition of this. The phrase has little meaning. Constructivism is to do with *learning as understanding*, not primarily with teaching. Thus, no

matter what we do, the learners will construct their own idiosyncratic understandings. Certain approaches to teaching and learning may help in allowing the development of better or more complete understandings.

Some Key Principles

- Much literature on constructivism is misleading and unhelpful
- All learners, naturally, seek to construct understandings
- Constructivism is to do with learning as understanding, not primarily with teaching
- There is no such thing as contructivistic teaching
- The key to success involves working within the limits of working memory capacity

Indeed, constructivism has emerged as a good way to ***describe*** what was going on. However, as Kirschner *et al.* (2006)[74] have demonstrated, constructivism, along with numerous other strategies, cannot offer a clear way forward *unless* it takes into account the *key effect of limited working memory capacity*. The focus of their work has been on the parallel concept of cognitive load.

Constructivism is an understanding of learning founded on the premise that, by reflecting on our experiences, we construct our own understanding of the world we live in. Each of us generates our own '*rules*' and '*mental models*' which we use to make sense of our experiences.

The Research Evidence

A much more useful way to look at learning is to consider how the human brain processes knowledge. In this, we use the word '*knowledge*' in its widest sense to encompass information, ideas, understandings, skills and insights. Here are some key features from research about models of learning as they affect assessment:

(1) Learning and knowing are more about *how* knowledge is organised and processed rather than *what* is stored in long-term memory. This implies that assessment must focus on the more complex areas of knowing rather than on discrete bits of knowledge.

(2) We need to assess not only *what* is stored in long-term memory but *how* it is stored and being *used*.

(3) Knowledge in all its richness is stored as schemata which tend to be domain-based and there is little evidence of transfer of learning or generic skills. Knowledge develops in highly contextualised and inflexible ways. This is inevitable: it reflects the way information has to be processed.

(4) There needs to be an emphasis on what features make a person an '*expert*' as compared to a '*novice*'.

(5) Assessment should also focus on the strategies which students use, accepting that these may vary widely.

(6) Much understanding is embedded in a social context and assessment must take this into account. Indeed, assessment needs to explore how well students engage in this aspect of learning.

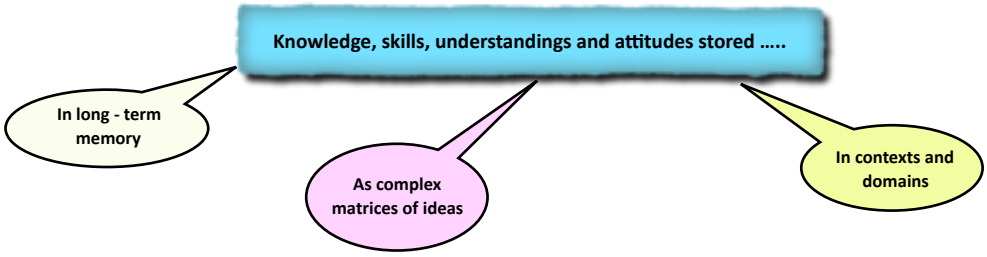


Figure 9.4 The Human Memory

Now let us look a little more at how the brain handles information.

[74] Kirschner, P.A., Sweller, J. & Clark, R.E. (2006) Why Minimal Guidance during Instruction Does not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-based, Experiential, and Inquiry-Based Learning, *Educational Psychologist*, 41(2), 75-86.

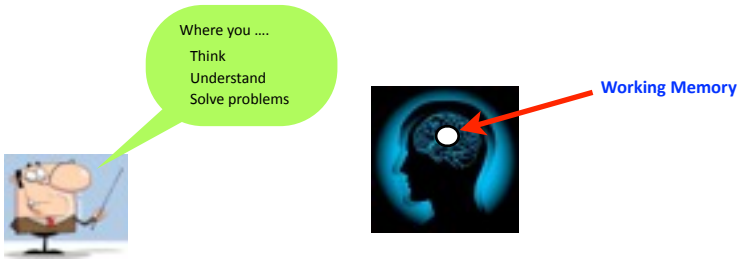
The Working Memory

In the mid 20th century, medical doctors observed that a blow to the head of a person often caused them to forget the immediate past (last 10 minutes or so) but did not affect memory of past events. This led to the clear idea that human memory had more than one component. One part had been damaged temporarily by an accident where there was a blow to the head, but the other part was unaffected. The two parts were labelled as '*short-term memory*' and '*long-term memory*'.

In 1956, Miller[75] developed a simple way to measure the capacity of what he called '*short-term memory*', his paper being one of the most quoted papers of all time. He obtained consistent results showing that the average number of pieces of information an adult could hold *at the same time* in the '*short-term memory*' was 7. His findings are highly reproducible. It was also found that the capacity of the '*short-term memory*' grew with age to about age 16 and that its capacity was fixed genetically. In other words, a person had to operate with the '*short-term memory*' they inherited from their parents. In passing, the capacity of '*short-term memory*' is nothing to do with what people have called '*intelligence*' but it may affect the success of students in the kinds of tests and examinations they face, including so-called '*intelligence tests*'.

In his brilliant work in the 1970s and 1980s, Johnstone[76] started to appreciate that all the areas in the sciences where students reported they had difficulties were those where the number of ideas that the student had to hold *at the same time* was high, if understanding was to be achieved. He then realised that this was controlled by the '*short-term memory*'. At one stroke, he had *explained why* certain topics and themes are difficult for learners. Again, his findings have been confirmed many times and his work probably represents the most important finding in all education for decades[77].

Very soon, it began to be understood that the '*short-term memory*' was not only a place where information was held temporarily (before possible storage in long term memory) but was also the location of the brain where we all think, understand and solve problems. It was renamed as the '*working memory*'. In the context of assessment, the working memory is where we work out answers to questions and solve problems. Its capacity is, therefore, critical in designing all assessment tasks.



[75] Miller, G. (1956) The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information, *The Psychological Review*, 63(2), 81-97.

[76] Johnstone, A.H. and Kellett, N.C. (1980) Learning Difficulties in School Science - Towards a Working Hypothesis, *European Journal of Science Education*, 2(2), 175-81.
Johnstone, A.H. (1981) Chemical Education, Facts, Findings and Consequences, (The Nyholm Lecture), *Chemistry in Britain,* 17(3), 130-135.
Johnstone, A.H. (1984) New Stars for the Teacher to Steer by?, *Journal of Chemical Education,* 61(10), 847-9.
Johnstone, A.H. and El-Banna, H. (1986) Capacities, Demands and Processes: a Predictive Model for Science Education, *Education in Chemistry*, 23(3), 80-84.
Johnstone, A.H. and El-Banna, H. (1989) Understanding Learning Difficulties - A Predictive Research Model, *Studies in Higher Education*, 14(2), 59-68.
Johnstone, A.H. (1991) Why is Science Difficult to Learn? Things are Seldom What They Seem, *Journal of Computer Assisted Learning*, 7, 75-83.
Johnstone, A.H. (1997) Chemistry Teaching, Science or Alchemy?, *Journal of Chemical Education,* 74(3) 262-268

[77] See Reid, N. (2009) Working Memory and Science Education, *Research in Science and Technological Education,* 27(2), 245-250.

The sustained work of Baddeley[78], working from a clinical psychological perspective, has offered the best insights into how the working memory actually operates. Thus, medicine, psychology and education[79] came together to understand that:

- ☑ Working memory is a specific part of the memory system of the brain
- ☑ Its capacity grows with age to about age 16 and remains constant throughout life
- ☑ Its capacity is fixed and cannot be altered in any way
- ☑ It is where we all think, understand, and solve problems
- ☑ Its capacity controls success in understanding, especially conceptual understanding
- ☑ The long-term memory is simply a store, of more or less infinite capacity

Now we turn to look at the implications for learning and assessment.

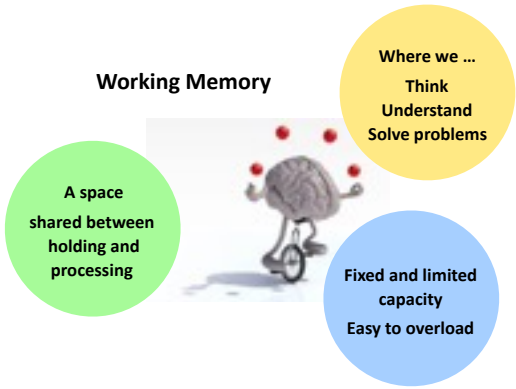
Importance of Memory

In the educational setting, students receive most of their information through sight and sound. Slavin (2000)[80] points out that most incoming information is discarded immediately and the person suddenly forgets the information. That which is not discarded is either held temporarily in the working memory and then it *may* be stored in the long-term memory.

The term '*memory*' concerns the ability to understand information, to retain it, to recall it when required and to recognise it when it is encountered again[81]. Learning relies on memory for permanence. Thus,

> '*...without the capacity to remember and learn, it is difficult to remember and learn; it is difficult to imagine what life would look like, whether it could be called living at all. Without memory, we would be servants of the moment, with nothing but our innate reflexes to help us deal with the world. There could be no language, no art, no science, no culture. Civilisation itself is a distillation of human memory...*'.[82]

Baddeley (1986)[83] noted that the working memory space is a place where we consciously think and perceive. It has a limited capacity and its function is to hold and process (there is also a time factor in processing) the information. These two functions have trade-offs: if there is too much information to hold, then less processing can take place and, if too much processing is involved, then it cannot hold much. He found that the working memory has various structures which seem to be part of it. He called these loops: the phonological loop and the visuo-spatial sketchpad. This work is important in understanding learning but we shall not pursue it further here.



---

[78] Baddeley, A.D. (1986) A Working Memory, London, Oxford University Press.
Baddeley, A.D. (1994) The Magical Number Seven: Still Magic after all these Years? *Psychological Bulletin*, 101(2), 353-356.
Baddeley, A.D. (2002) Is Working Memory Still Working, *European Psychologist,* 7(2), 85-97.
Baddeley, A.D., & Logie, R. (1999) Working memory: The multiple component model, in Miyake, A., & Shah, P. (Eds.), *Models of Working Memory* (pp. 28- 61) (New York: Cambridge University Press)

[79] St Clair-Thompson and Botten, C, (2009) Working memory and science education: Exploring the compatibility of theoretical approaches, *Research in Science and Technological Education*, 27(2), 139-150.

[80] Slavin, R.E. (2000) E*ducational Psychology* (6th Edn), Boston, USA, Allyn & Bacon.

[81] Onwumere, O. (2009). *Difficulties in Understanding Mathematics: An Approach Related to Working Memory and Field Dependency.* PhD thesis submitted to University of Glasgow, UK.

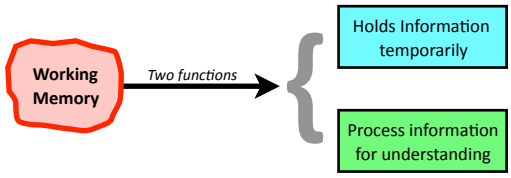
[82] Blakemore, C. (1988) *The Mind Machine*, London: BBC Publication, Page 46.

[83] Baddeley, A.D. (1986) A Working Memory, London, Oxford University Press.

Working Memory Functions

Our understandings of working memory has arisen from very large amounts of research. The main characteristics of working memory space described by Johnstone (1997) are:

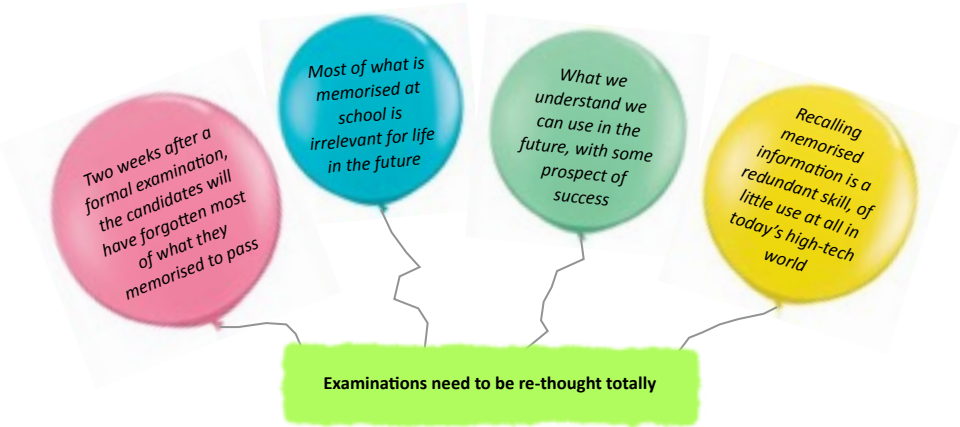
> *'The working memory space has two main functions. It is the conscious part of the mind that is holding ideas and facts while it thinks about them. It is shared holding and thinking space where new information coming through the filter consciously interacts with itself and with information drawn from long-term memory store in order to make sense. However, there is a drawback. The working space is of limited capacity… It is a limited shared space in which there is a trade-off between what has to be held in conscious memory and the processing activities required to handle it, transform it, and get it ready for storage in long-term memory store. If there is too much to hold, there is not enough space for processing; if a lot processing is required, we cannot store too much'.*[84]



The memory that each of us possesses is critical in enabling us to be the human beings we are. However, the working memory (a tiny part of our total memory) is absolutely critical for understanding. If too much information is need to be held in the working memory *at the same time* in order to gain understanding, then ***there can be no understanding***. Thus, it is imperative that every aspect of teaching takes the limited capacity of working memory into account.

If understanding proves impossible, then the learners usually resort to memorisation in order to pass the examinations. The attitudinal consequences of this are serious as Jung and Reid (2009) have demonstrated and this explains why so many choose not to study the sciences in many countries[85]. Understanding has proved so difficult that they then give up and opt out.

There is a sad comment here. It is possible to pass examinations, and often pass well, simply on the basis of the recall or memorised information. If understanding is a key goal of education (not memorisation), then it is critical that examinations are developed where it is ***impossible*** to pass simply on the basis of accurate recall. This was a goal in the Scottish Examination system of the 1960s and 70s in the sciences[86]. Although more submerged, it still continues on in some measure today.



---

[84] Johnstone, A.H., (1997) 'Chemistry Teaching-Science or Alchemy?'. *Journal of Chemical Education,* **74**(3), 262-268.

[85] Jung, E-S., and Reid, N. (2009) Working Memory and Attitudes, *Research in Science and Technological Education,* 27(2), 205-224.

[86] Scottish Examination Board (1962a) *Curriculum Papers 490 ñ Alternative Physics for Ordinary and Higher Grade*. Scottish Examination Board: Dalkeith, Edinburgh.
Scottish Examination Board (1962b) *Curriculum Papers 512 ñ Alternative Chemistry for Ordinary and Higher Grade. Scottish Examination Board: Dalkeith, Edinburgh.*

Memory Models

Many studies have explored the way the human memory works and, by the 1960s, clearer pictures were starting to become apparent. The work of Atkinson and Shiffrin (1971) on information processing is considered to offer a key foundation in this work[87]. Numerous others followed (Johnstone and Kellett, 1980; Johnstone, 1984; Ashcraft, 1994; Brunning *et al.*, 1995; Sweller et al., 1998)[88]. It has to be stressed that all these models of memory were developed based on carefully gathered empirical evidence. They are all very similar but this section seeks to find out some of the key developments.

The overall picture is shown in figure 9.5.

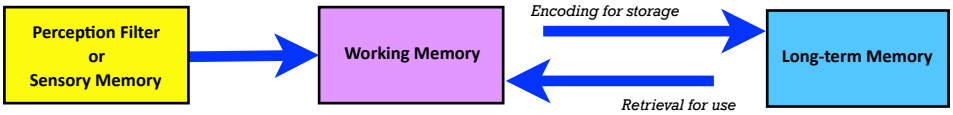


Figure 9.5 The Overall Pattern

It is interesting that the models all tended to use the language of the world of computing. Partly, this was because the new age of the computer was developing rapidly at that time. However, it is an interesting thought that humans invented computers and, perhaps, the computers are modelled on the ways by which the human mind actually works.

Figure 9.6 shows more detail of what is known:

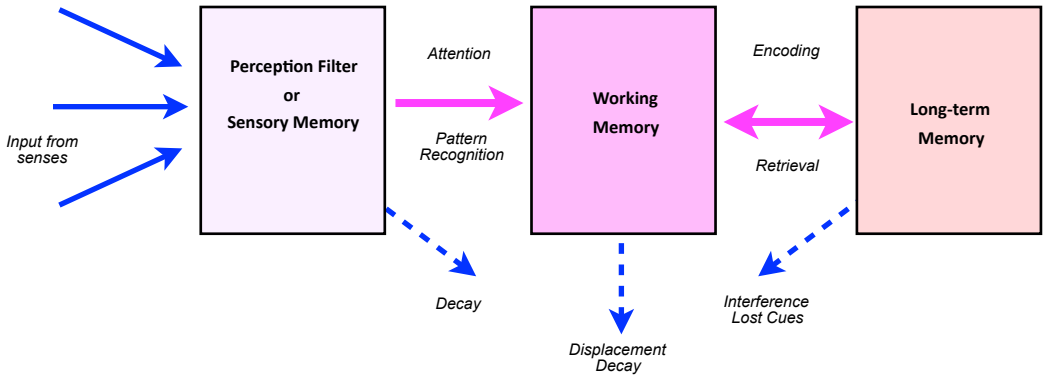


Figure 9.6 The Way the Brain Processes Information

The key feature is that the research shows the way the memory works in terms of the flow of information. We receive information through our senses - mainly eyes and ears in education. We select what we decide what is relevant or important (consciously or subconsciously) and this moves into the working memory. Here, it is processed. This may involve attempts to make sense of what has come in or to look into the long-term memory for past information that helps the sense-making process. Understandings or information itself can then be moved to the long-term memory for storage.

The model of information processing as a way to conceptualise all learning is based on very extensive research evidence gathered over many decades and is extremely well supported by the evidence. As a model, it has very considerable predictive power.

---

[87] Atkinson, R. and Shiffrin, R. (1971) The Control of Short-Term Memory. *Scientific American*, 225, 82-90.

[88] Johnstone, A.H., Kellett, N.C. (1980) Learning Difficulties in School Science - Towards a Working Hypothesis, *European Journal of Science Education*, 2(2), 175-81.
Johnstone, A.H. (1984) New Stars for the Teacher to Steer by?, *Journal of Chemical Education,* 61(10), 847-9.
Ashcraft, M. H. (1994) *Human memory and cognition*. New York, Harper Collins College.
Bruning, R.H., Schraw, G.J. & Ronning, R.R. (1995) *Cognitive Psychology and Instruction*, Englewood, USA, Prentice-Hall Inc.
Sweller, J., van Merrienboer, J., & Paas, F. G. (1998) Cognitive architecture and instructional design: Educational Psychology Architecture and instructional design. *Educational Psychology Review* , 10, 251-296.

In 1993, Johnstone[89] published his empirically-based model and this model absorbs the major components of previously research models and the key ideas of other research including that of Ashcraft (1994), Piaget (1963), Ausubel (1978), Gagne (1964), Pascual-Leone (1970), and Baddeley (1986)[90]. The model seeks to offer a description of what is happening when learning is taking place and to account for the limitations of learning. The model shows why learning is sometimes difficult or even impossible. Many of these insights have been tested and, in every case, the model has been found to predict successfully (Reid, 2008)[91].

There are three major components of this model:

(1) Perception Filter

(2) Working Memory Space

(3) Long-Term Memory

As we teach, learners are selecting from what they see, hear and do. Thus, they may come away with a very different picture of what they have learned when compared to what we intended. We often discover that the selection has been inadequate when we come to mark their test papers!

The working memory is a space of *very limited capacity* where we do all our thinking, understanding and problem solving. It is easy to measure the capacity of the working memory of any person using simple and reliable tests. For adults (age 16 and over), the average capacity is the ability to hold or handle 7 things *at the same time*. Some adults can hold more, some less. Almost everyone has a capacity lying between 5 and 9 although most can hold 6, 7 or 8.

The working memory can be '*emptied*' so that it can hold fresh information but its capacity cannot be extended. The capacity of working memory is genetically determined and is not related neatly to abilities. Younger learners have a lower average and this explains why certain skills and concepts cannot be introduced too early: there is insufficient capacity to handle them.

The long-term memory is simply a store, of more or less infinite capacity. Research shows how information can move around the three parts of the brain.

The complete model is shown:

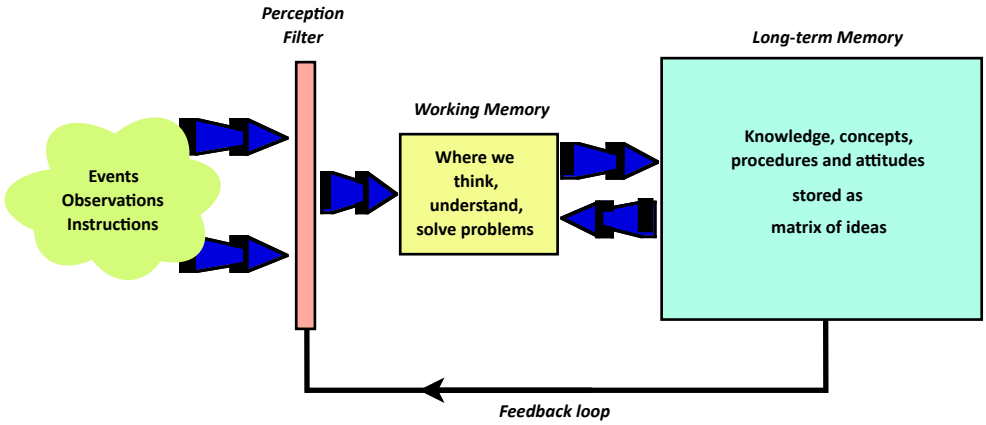


Figure 9.7 Information Processing in the brain

---

[89] Johnstone, A.H. (1993) The Development of Chemistry Teaching - A Changing Response to Changing Demand, *Journal of Chemical Education*, Vol. 70, No. 9, pp. 701-705.

[90] Ashcraft, M. H. (1994). Human memory and cognition. New York, Harper Collins College.
Ausubel, D.P., Novak, J. D. and Hanesian, H. (1978) *Educational psychology: a cognitive view*. 2nd Ed. Rinehart and Winston, New York.
Gagné, R. M. (1964) *Problem solving*. In A.W. Melton (Ed.), Categories of Human Learning. New York: Academic Press.
Baddeley, A.D. (1986) *A Working Memory*, London, Oxford University Press.

[91] Reid, N. (2008) A scientific approach to the teaching of chemistry (The Royal Society of Chemistry Nyholm Lecture, 2006–2007), *Chemistry Education Research and Practice*, 9(1), 51–19

All learners select from all that is coming at them and everything goes into the working memory. After processing, information and understandings may be stored in long-term memory for later recall. If new information is linked extensively to previous information, then this can be seen as understanding. If it is stored unlinked, it is rote memorisation. If it is linked wrongly, then we have a misconception. Evidence suggests the model covers all learning, formal and informal.

Information Processing and Assessment

The model can now be expanded to include what happens in assessment.

In terms of assessment, we are looking at what is stored in long-term memory: what the student knows, what the student understands, and what the student can do. However, the working memory is the key in that this is where the knowledge, understandings and skills are handled before being demonstrated.

The capacity of working memory is very limited and the assessment task set may demand more space than is available.

It has been shown again and again that the capacity of working memory is a critical factor in determining success in assessment tasks

However, the extent of the factor is determined by the nature of the tasks set. The factor is often very critical in highly conceptual areas of the curriculum (especially the mathematics, chemistry and physics areas) but limited working memory capacity as a factor applies to all subjects at all stages, even in tests of pure recall.

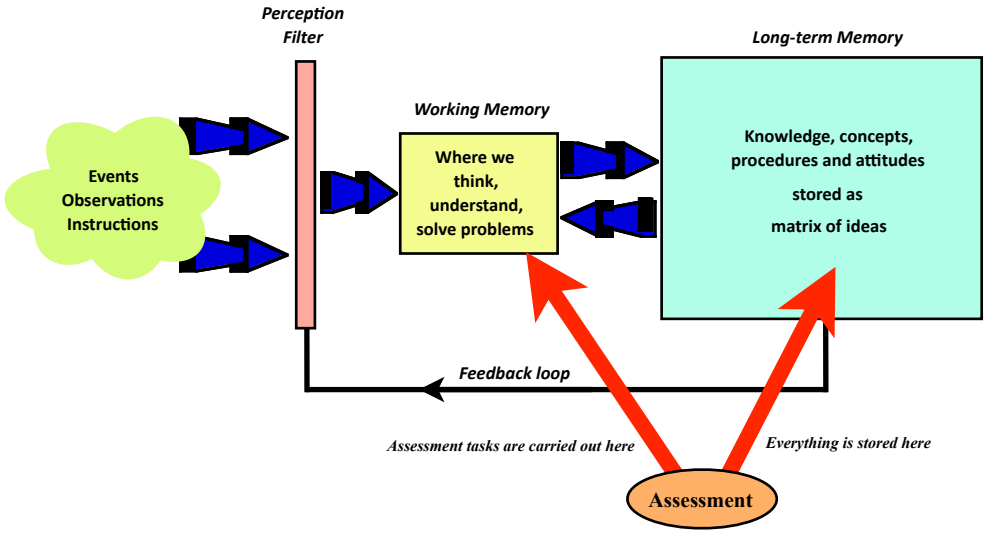


Figure 9.8 Assessment and the Brain

Thus, all assessment must be carried out within the limitations of working memory in mind. We can only get at what is in long-term memory by '*bringing it out*' through working memory: a key principle.

Assessment must work within the limitations of working memory capacity

A Summary

It must be stressed that the information model for all learning was developed on the basis of a vast accumulation of evidence derived from some very carefully conducted research. It is not simply the suggestions of others. It is the best understanding and interpretation of research evidence. Indeed, the model has been used many times to predict how learning can be improved. In each case, the prediction has been supported by subsequent research studies.

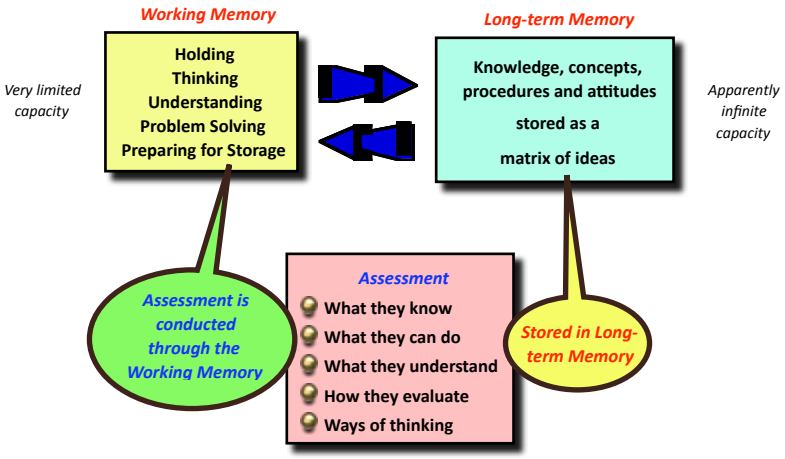
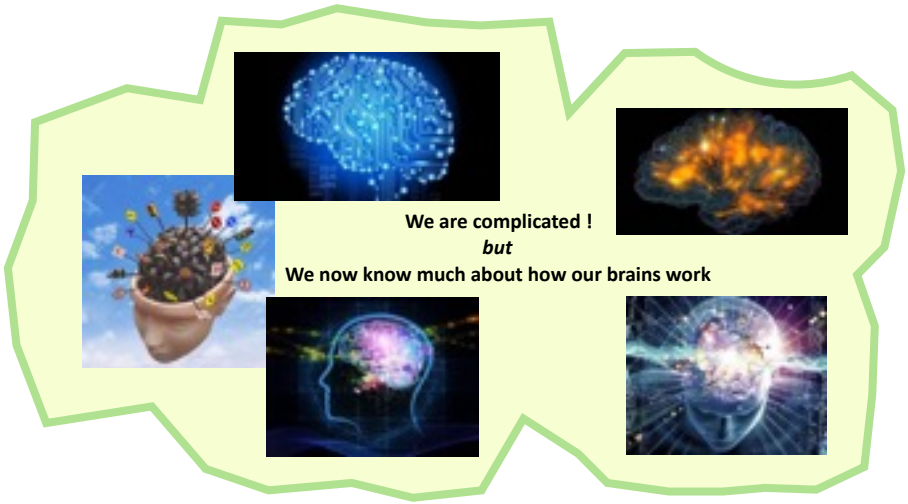


Figure 9.9 The Working Memory and Assessment

Some of the findings are now summarised, with their implications for all assessment:

Key Principles About Learning

- We all learn in essentially the same way;
- Variations are *within* the basic common processes;
- Working memory has a *very limited* capacity;
- Working memory grows with age but is fixed genetically by about age 16;
- Working memory cannot be expanded but it can be used more efficiently;
- Working memory capacity is an aspect of the way the brain works; it is not neatly related to ability;
- We can easily end up measuring the capacity of working memory and not the understanding of what has been learned;
- Assessment questions and tasks MUST be designed to take account of the limitations of working memory capacity.



Assessment Moving *into* a New Future

Skills Assessment

Aims

This chapter considers valuable skills that can be developed through education and how these skills can be assessed so that there are rewards for the students in these areas

Introduction

Assessment is a process of obtaining information about learning and achievement. In this chapter, we want to move on and show the wide range of skills that we need to consider and how we can attempt to assess them. Skill assessment is important where education or training involves performance. Some abilities require mastery, for example, conducting a chemical analysis, creating a sculpture, using a word processor, assembling a computer, preparing a meal in a hospitality course, completing a wood-work project, giving a hypodermic injection, driving a car, operating a forklift or piloting a plane. in such cases, where people must reach a given level of performance, it is imperative that it is evaluated by practical tasks.

Skill assessments are designed to measure a student's competence on some phase or operation. There are two general domains of skills:

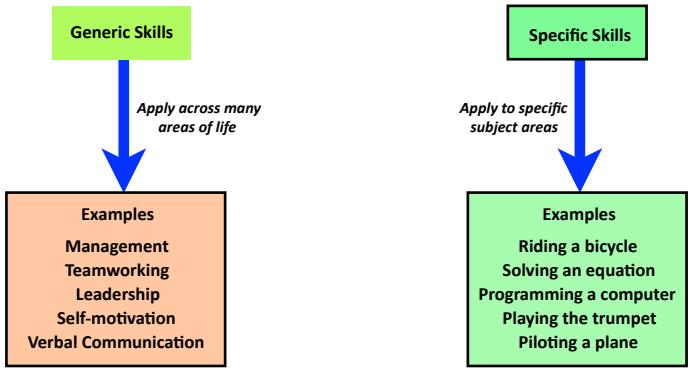


Figure 10.1 Domains of Skills

There is even the skill in being able to handle examinations questions! Long ago, Gagné (1974) said,

> *'An individual interacts with his environment by using symbols. He uses language to deal with his environment symbolically. As the learning of school subjects continues, symbols are used in more complex way: distinguishing, combining, tabulating, classifying, analysing, and quantifying objects, events, and even other symbols. This kind of learned capability is given the name 'Intellectual Skills'.*[92]

[92] Gagné, R. M. (1974) *Principles of Instructional Design*, Holt, Rinehart & Winston, Inc. USA.

Indeed, there may be many students who are extremely capable but have never mastered the art of being able to show their talents in formal written examinations and this reveals the need to re-think the whole area of assessment.

Robert Gagné (1977)[93], who studied learning as an instructor, saw learning as a process of change in human capability. Working with adults, he saw this process of learning as not dependent on the process of growth and this change in human capability persisting over a time. He says that learning is an unseen internal process but what is going on inside the learner can be inferred from the outcomes of the learning.
Gagné stated five categories of learning outcomes as shown:

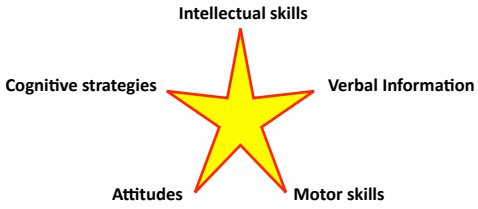


Figure 10.2 Categories of learning outcomes

In 1990, George Miller[94] proposed a classification for medical education involving four skill levels: '*Knows*' (knowledge) is at the lowest level of the pyramid followed by '*knows how*' (competence), '*shows how*' (performance), and '*does*' (action).

Sadly, in school and university education, the literature is full of suggested lists of highly desirable skills. There are very few research studies which discuss ways by which the development of such skills can be achieved and almost none that offer evidence on its assessment or achievement.

The sad picture of what is supposed to be educational literature is illustrated in figure 10.

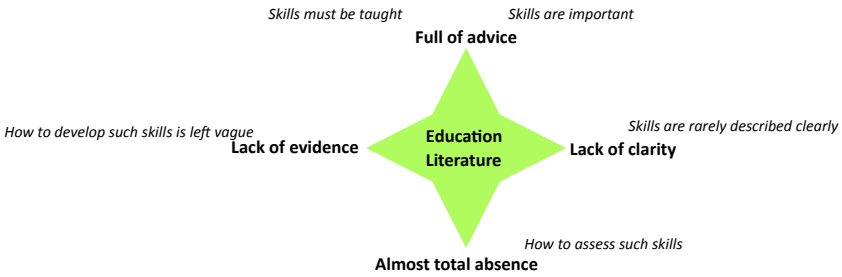


Figure 10.3 A Sad Picture

One of the most important areas relates to thinking skills and, here, there are some recent studies which can offer us much help although most of the education literature is simply speculation.

[93] Gagné, R. M. (1977) *The Conditions of Learning*, 3rd ed. New York: Holt, Rinehart and Winston, 1977.
[94] Miller, G. (1990) The assessment of clinical skills/competence/performance, *Academic Medicine,* 65(9), 63-67.

Thinking Skills

Sadly, much of the literature is simply speculation, with different authors coming up with different (sometimes interesting, sometimes very vague) ideas[95]. Thus, Griffin speaks of, '*everybody talking about thinking skills is not talking about the same thing*' while Smith (2002) speaks of '*a capacity, usually acquired through training and experience, to do something well, to perform competently certain tasks*'. Some have developed all kinds of lists of mental activities (like observation, categorisation, investigation, comparison, and interpretation) and see thinking as built of combinations of these in an attempt to reach a solution to or decision about the problem in hand (Jerwan, 2002; Zaitoon, 2003; Saadah, 2003[96]). Some view thinking skills as the outcome of a number of essential skills such as recall, deduction, analysis[97] while others have invented the idea of compound thinking skills[98].

Humans think. We all know this! Everyone thinks. You are thinking at this moment as you read these words. You might be thinking that this topic is boring or that it is not being introduced very clearly. You might be thinking that you might learn something new about thinking. You might be thinking that you already know all there is to know about thinking, and you might be wondering why you are reading this in the first place. The point is that you are thinking right now. You think because you are human, and humans think.

Let us stop for the moment and think further! We want to be able to *assess* thinking. That is a very different issue. It means that we have describe thinking in ways precise enough so that we can devise questions and tasks that might assess the presence of the skill. This is known as developing an '*operational description*'. Such a description is precise enough so that measurement can be attempted. We shall come back in a moment to see how measurement can be approached.

However, we need to remember that good thinking is hard work. Learning to think clearly and carefully takes training, patience, and practice. Thinking carefully with clarity, depth, precision, accuracy, and logic is thinking critically. Critical thinking is the process of thinking in a certain way. Creative thinking means thinking in a different way. Scientific thinking or systems thinking are other specialist ways of thinking - and they all take mental effort

---

[95] Griffin, A. H. (2001) Thinking in education yesterday, today and tomorrow. *Education,* Vol. 106, No. 3, pp. 268-280.
Smith, G.F. (2002) Thinking skills: the question of generality. *Journal of Curriculum Studies,* Vol. 24, No. 6, pp. 659-678.

[96] Jerwan, F. (2002) *Teaching Thinking: Conception and Application.* Amman, Jordan: Dar Alfiker.
Zaitoon, H. H. (2003) *Teaching thinking: Vision applied to the development of the thinking mind*. Amman, Jordan: Dar Al Shorok
Saadah, G. (2003) *Teaching Thinking Skills*. Amman, Jordan: Dar Al Shorok.

[97] Habib, M. A. (1996) *Thinking: Theoretical Foundation & Strategy.* Cairo: Arabic Renaissance Library.
Habib, M. A. (1997) Self-Reflection and Innovative Features Associated with Multi-dimensional Thinking for Undergraduate Students, *Journal of Psychology*, Nos. 40-41.

[98] Albaz, K. S. (2001) Effectiveness of the Use of Marzano's Model for Learning Dimensions in Teaching Chemistry on Achievement and Compound Thinking and Aptitude towards the Subject for the 1st Secondary Class in Bahrain, *Procedure of the 5th Scientific Conference*, Vol. 2, Egyptian Education Association, Cairo: Ain Shams University, pp. 413-448

Thinking Skills

Many curricula make reference to '*thinking skills*'. However, there is a desperate need to ***describe*** what we mean by such skills and to outline how such skills might be assessed.

First of all, we offer a simple description of three skills which are a part of thinking skills:

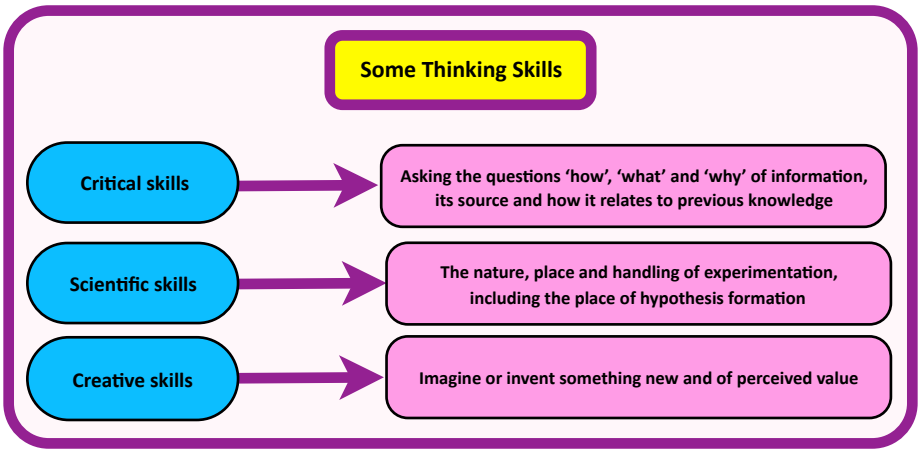
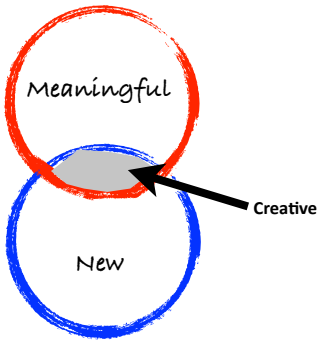


Figure 10.4 Description of Thinking Skills

These three descriptions are given in various places in the literature and methods to assess these three skills are described. It is important to recognise that creative skills have been assessed successfully for a very long time. Thus, teachers of subjects like music, art, drama, and creative writing have found good ways to assess creative skills and these ways depend on the professional judgments of subject experts, using clear criteria.



Scientific thinking skills have been analysed in some detail by Al-Ahmadi[99] and the findings published[100]. Critical thinking has been the subject of an outstanding PhD thesis[101]. Some of the key findings have been published recently[102]. The important things about all these studies is that they offer clear operational descriptions of what we mean by specific ways of thinking and they have given tested and validated examples of assessment methods. The same approach was adopted in looking at systems thinking, a set of skills of considerable importance in biology, medicine, economics and engineering.[103]

---

[99] Al-Ahmadi, F.M.A.A. (2008) *The Development of Scientific Thinking with Senior School Physics Students,* PhD Thesis, University of Glasgow, Glasgow. [http://theses.gla.ac.uk/241/]

[100] Reid, N and Serumola, L. (2006) Scientific Enquiry: The Nature and Place of Experimentation: a review, *Journal of Science Education*, 7(1), 1-15.
Reid, N and Serumola, L. (2007) Scientific Enquiry: The Nature and Place of Experimentation: some recent evidence, *Journal of Science Education*, 7(2), 88-94.
Al-Ahmadi, F. and Reid, N (2011) Scientific Thinking - What is it and can it be measured? *Journal of Science Education*, 11(2), 53-59.
Al-Ahmadi, F. and Reid, N (2012) Scientific Thinking - *Can it be Taught? Journal of Science Education*, 13(1), 18-23.

[101] Al-Osaimi, K.H. (2013) *The development of critical thinking skills in the sciences*, PhD Thesis, University of Dundee, Dundee. [http://discovery.dundee.ac.uk/portal/en/theses/the-development-of-critical-thinking-skillsin-the-sciences%28f8eb450c-0914-4e1b-aa1c-65dd9d02cfba%29.html]

[102] Al-Osaimi, K.H., Reid, N., Rodrigues, S. (2014) Critical Thinking - Can it be Measured? *Journal of Science Education*, 15(1), 30-36.
Al-Osaimi, K.H., Reid, N., Rodrigues, S. (2014) Critical Thinking - Can it be Developed? *Journal of Science Education*, 15(2), 57-75.

[103] Chandi, S.S. (2008) *Using Systems-Thinking as a Teaching and Learning Tool in Biology Education* (PhD thesis), University of Strathclyde and University of Glasgow.
Chandi, S.S., Reid, N., McWilliam. A., and Gray, D. (2009) Exploring the usefulness of a systems-based model from the perspective of biology students, *Scottish Educational Studies*, 41(1), 68-85.

Now we have some idea what we are considering in these three thinking skills areas. How about approaching assessment? We have to acknowledge that, in the present state of knowledge, assessment will not be easy. However, here is the approach adopted in many of the research studies quoted:

Look for tasks where the presence of these thinking abilities is likely to lead to greater success

In other words, when trying to assess the skills of critical thinking, scientific thinking and creative thinking, we need to develop assessment tasks where the student who possesses the skills will perform better. This is very different form rewarding the students with the best memory.

Now, let us examine the three types of think a bit further and see how assessment can be approached.

Assessing Thinking

This is the way that many teachers have considered creativity for years. They have set tasks and challenges where the presence of creative thinking skills gives the learner an advantage. Some '*product*' (intellectual, practical, or combined) emerges where there is evidence of something '*new*' and something of perceived value. This has applied in the expressive arts, physical activities as well as areas of the humanities. It could apply anywhere.

Of course, assessment involves considerable value-judgement. However, different examiners can show considerable agreement over this, suggesting that creative thinking *can* be measured by looking at what the learner has created. This can even happen in written form. The great novelists in the world are capable of highly creative thinking. We can ask our students to write an essay on a theme which can show their abilities and skills to write creatively.

Books have been written on this subject and it is generally agree that this is a desirable skill. However, rarely have they developed any operational descriptions that might guide us in assessment. Very recently, two papers have offered a way forward by showing that critical thinking can be measured and critical thinking skills can be developed[99]. The analysis is built around the simple model:

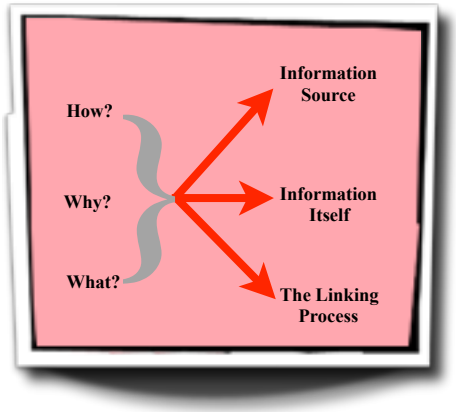


Figure 10.5 Critical Thinking : An analysis

The authors have developed a short test where being able to answer the questions successfully gives evidence of their critical thinking skills, using this model. In some ways, describing critical thinking is best seen by looking at the opposite. Figure 10.9 shows this.

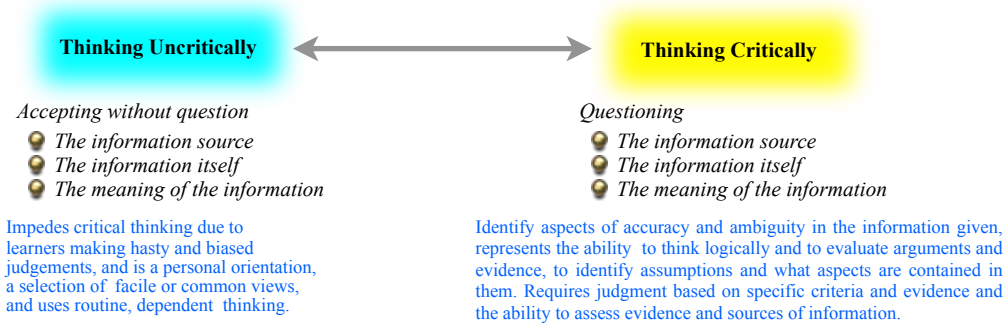


Figure 10.6 Thinking Critically and Uncritically

The key to thinking critically is being unafraid to ask questions and keep on asking questions, never accepting an answer unless it is supported by the evidence.

In many ways, scientific thinking has generated even more confusion. The literature is full of books and articles where the authors (always scientists) have attempted to describe scientific thinking using long lists of skills. In the end, their lists seem to encompass ever conceivable form of thinking and the reader is often left with the conclusion that the only people who ever think are scientists: an insult to the majority of the world's population!

However, almost never have they developed any operational descriptions that might guide us in assessment. Recently, four papers have offered a way forward by showing that scientific thinking can be measured and scientific thinking skills can be developed[104].

The one feature of all the sciences is that they gain their insights and understandings by means of experiments (that includes computer modelling). That is what makes scientific thinking unique and that led to the clear description:

The nature, place and handling of experimentation, including the place of hypothesis formation

Given that operational description, the authors were able to develop a test of scientific thinking.

With both critical thinking and scientific thinking, there is still a problem. The tests described in the literature were designed for research purposes where several hundred school students were involved. The test certainly worked well but there is no certainty that these tests are yet ready to measure the skills at an individual level although that may not prove a major problem. However, there is a greater problem: the tests were designed to fit specific age ranges and specific subject areas. They would not necessarily adapt for other contexts.

---

[104] Reid, N and Serumola, L. (2006) Scientific Enquiry: The Nature and Place of Experimentation: a review, *Journal of Science Education*, 7(1), 1-15.
Reid, N and Serumola, L. (2007) Scientific Enquiry: The Nature and Place of Experimentation: some recent evidence, *Journal of Science Education*, 7(2), 88-94.
Al-Ahmadi, F. and Reid, N (2011) Scientific Thinking - What is it and can it be measured? *Journal of Science Education*, 11(2), 53-59.
Al-Ahmadi, F. and Reid, N (2012) Scientific Thinking - *Can it be Taught? Journal of Science Education*, 13(1), 18-23.

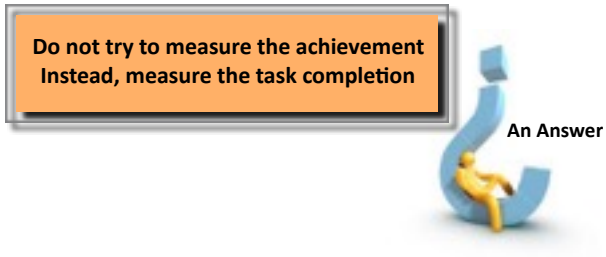
Therefore, in the immediate future, it is highly unlikely that there will be any reliable and fair tests available to test any of these skills. However, by thinking though the test tasks we set carefully, we may steadily begin to see ways by which we can obtain evidence. Much will come through observation of the processes of learning and the various activities which learners undertake in our own classrooms and workrooms.

Giving a learner a grade for, say critical thinking may be utterly unrealistic. However, we may be able to say, *with confidence* that we have seen evidence of critical thinking developing in learners as they progress through our courses. Indeed, research with teachers in the creative arts shows considerable resistance in having to give a mark or a grade for creativity. The need to give a mark or grade almost contradicts the very nature of creative thinking.

Please do not give me a grade

Creative thinking can permeate the curriculum in all kinds of ways but, perhaps, the way forward is for those subjects where it is of special importance to consider awarding a '*duly performed*'. In this, on the basis of considerable evidence gathered over a course, you would indicate that the student had made ***satisfactory progress*** and that the student had reached a ***satisfactory skill level***. The same sensible principles can be applied to critical thinking and scientific thinking.

However, there is another way forward that shows great promise:



This leads to the concept of *'duly performed*' - a powerful way to look at assessment.

The Idea of *'Duly Performed*'

Here is a very simple way to give credit for skills where we do not as yet have good ways to provide a measurement in any formal assessment. It is not new but has been used successfully for at least 50 years!!

Suppose you have an eight week long laboratory course in chemistry as part of the university chemistry course. How do we give credit for this course? The first thing to note is that there is no simple, economic way to assess the achievement of the desired outcomes of this course. In the list of desired outcomes, the course aims to:

(a) Make the chemistry covered in the lecture course real and meaningful
(b) Show how experimental work can provide the data which leads to theories and models
(c) Develop skills of planning and team-working
(d) Allow students to develop an awareness of safety considerations
(e) Develop confidence in handling key pieces of equipment

The only way to assess such skills might be to watch every student all the time during the eight-week course - a complete impossibility and totally inhumane. It is important to note that handing in laboratory reports after each experiment gives almost no evidence of the achievement of *these* goals.

Do not try to measure the achievement
Instead, measure the task completion

However, research has shown that completion of this course does enhance the desired skills. Therefore, we measure the completion of the course. If the student attends each time and completes the experimental work and hands in a brief report showing what they did, then they gain the maximum credit. If a student misses some laboratory sessions and fails to complete all the reports, then they gain less credit. If a student fails to attend for the majority of the course or fails to complete a minimum number of reports, then they are denied the opportunity to sit the final examination in the lecture course. This approach has been used successfully from the 1950s in university laboratory courses.

One of the important skills in a physics course at school level was development of the problem solving skills. The problem with problem solving skills is that such skills are highly context dependent. Therefore, we must look at problems in the context of physics. There is simply no way to measure the development of such skills with any accuracy or reliability. However, '*duly performed*' offers an easy way.

Research has shown that the completion of four group problem solving tasks develops the skills sufficiently for your purposes[105]. Each of the four group problem solving tasks is based on a real-life problem which the group has to solve and each task takes about 1.5 hours. If a student takes part fully in a group to solve four such problems, then they gain the full credit. If they complete 3 out of the 4, then they get ¾ of the credit, and so on. The very act of completion of four tasks is sufficient to develop the desired skills adequately although it is impossible to measure success by any kind of formal test at an individual level.

The key thing about research is that it shows what can be achieved with a large sample of students, on average. We use that evidence to give credit to an individual level, even although we cannot measure accurately at that level. The students feel good. They know what is required of them and, if they are committed enough to complete all the tasks, they gain the credit. Thus, problem-solving skills in some subject discipline have been rewarded.

Now let us go back to look at problem solving skills a little further.

Problem Solving

This has become a fashionable skill list in the aims of many courses today. Again, the problem is how to define a problem, in a way that allows assessment to be possible. Again, the literature is not helpful for it often fails to give a clear picture that can take us forward in our thinking.

A description given by Hayes (1981)[106] is now widely used:

' … whenever there is a gap between where you are now and where you want to be, and you don't know how to find a way to cross that gap.'

This description is excellent. They key thing is that a problem means uncertainty - we do not know how to cross the gap. That is very different from the usual kinds of problems we see in mathematics, physics or chemistry. These usually involve the learner applying some simple rule or algorithm and then processing the numbers correctly. There is no uncertainty here.

[105] See, for example, Reid, N. and Yang, M-J. (2002) Open-ended Problem Solving in School Chemistry: A Preliminary Investigation, *International Journal Science Education*, 24(12), 1313-1332.

[106] Hayes, J. R. (1981). *The complete problem solver*, Philadelphia: The Franklin Institute Press.

In simple terms, the way we have conceptualised problem-solving in the past has been far too limited and the questions which supposedly test the skills almost certainly do not do so. The problem is a lack of clear definition and that is where the work of Hayes is important. In fact, it is even more important when we realise that, throughout life, we shall all face situations, events and decisions where we simply do not know what to do: these are real problems of real life. Education needs to prepare the learners for life.

Many years ago, a simple, but clear-sighted, analysis by Johnstone (1993)[107] noted that there are three elements in *every* problem: *things we know, a goal to be reached, and a method or strategy to get there*. Each of these can either by fully known or not. This gives eight types of problems. His model is now widely used.

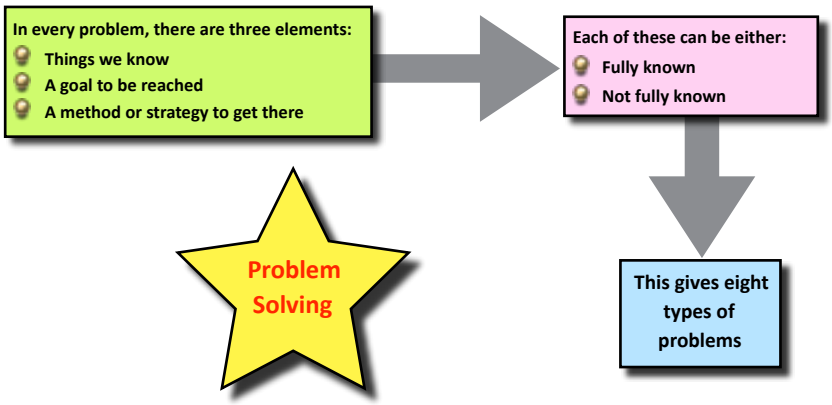


Figure 10.10 Three Dimensions of Problems

At last, we have a clear description of what problem solving means. The beauty of this analysis is that it works in every area of life, in every subject taught at school or university, in every context or circumstance. The analysis is also so straightforward.

Johnstone presented his analysis as a table:

Exercises or algorithmic tasks

Type	Data	Methods	Goals
1	Given	Familiar	Given
2	Given	Unfamiliar	Given
3	Incomplete	Familiar	Given
4	Incomplete	Unfamiliar	Given
5	Given	Familiar	Open
6	Given	Unfamiliar	Open
7	Incomplete	Familiar	Open
8	Incomplete	Unfamiliar	Open

More open-ended (or real) problems

Table 10.1 Eight Problem Types (after Johnstone, 1993)[108]

It is interesting that, in his original publication of this table, Johnstone never suggested that there was any hierarchy in difficulty. In other words, there are eight different *types* of problems but none is intrinsically more difficult. Later research showed that this was correct[109].

[107] Johnstone, A.H. (1993) In, *Creative Problems Solving in Chemistry,* Royal Society of Chemistry, London.
[108] Wood C.A. (2006) The development of Creative problem solving in chemistry, *Chemistry Education Research and Practice*, 7(2), 96-113.
[109] Yang, M-J. (2000) Problem solving in chemistry at secondary school, PhD Thesis, University of Glasgow, Glasgow. [http://theses.gla.ac.uk/2161/]

The power of the analysis means that we can place every assessment question or task into one of the eight categories. If your subject uses problem solving as one its learning outcomes, have a look at examples of assessment questions you use and see if you can place them in one of the eight categories.

Research has shown the near impossibility of developing problems for school level under types 7 and 8. Research has also shown that almost all assessment problems (even at university levels) rarely move beyond type 1 and these are scarcely problems in the sense that Hayes uses.

There are two other uncomfortable findings from research[110]:

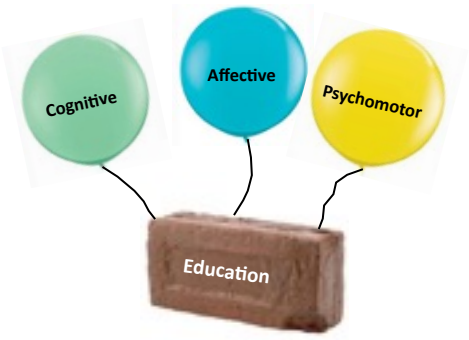
- While it is possible to teach the skills for solving algorithmic problems (exercises, type 1), it has proved to be ineffective to try to teach any strategies for solving the more open-ended problems (types 2 to 8). Their very open-endedness means that taught strategies do not really work and, indeed, strategies can actually be unhelpful.
- There is no evidence that there is such a thing as generic problem-solving ability and there is strong evidence to suggest that problem-solving is a *context dependent* skill. Thus, allowing the learners to solve problems in your subject is *no guarantee* that they will become better problem solver in any other area.

Problem Solving

- It is easy to test skills at closed problem solving
  - Set a problem question and see how can solve it
- It is very difficult to test all other kinds in an examination
  - How can we make it fair ?
- Research shows that the following factors are critical for success:
  - Working memory capacity
  - Holding many ideas in long term memory well linked together
  - Numerous cognitive characteristics like visual-spatial ability, extent of divergency
  - Confidence
- Teaching strategies for solving problems is very limited and only really applies to closed problems (are these problems?).

An Extra Thought

It is often stated that education involves three dimensions:



The three areas or dimensions are often described as:

Cognitive:	Knowledge and understanding
Pscyhomotor:	Practical skills
Affective:	Attitudes

[110] Yang, M-J. (2000) Problem solving in chemistry at secondary school, PhD Thesis, University of Glasgow, Glasgow. [http://theses.gla.ac.uk/2161/]
Reid, N. and Yang, M-J. (2002) The Solving of Problems in Chemistry: The More Open-ended Problems, *Research in Science and Technological Education*, 20(1), 83-98.
Reid, N. and Yang, M-J. (2002) Open-ended Problem Solving in School Chemistry: A Preliminary Investigation, *International Journal Science Education*, 24(12), 1313-1332.

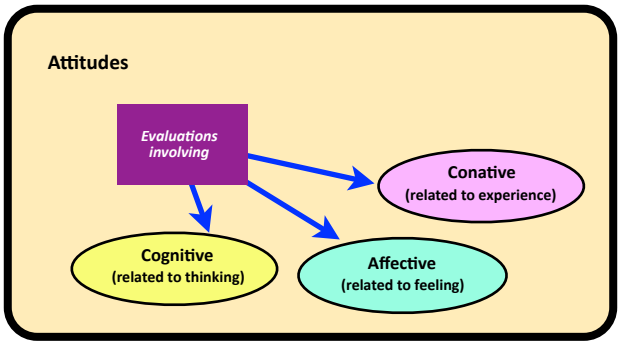
The aim of those who have emphasised this kind of approach is understandable. They want to move learning away from education's emphasis on '*head knowledge*' - filling the heads of learners with facts and information. However, the analysis above is not quite right.

The first problem relates to the word *'affective*'. This word relates to that which is involves feelings or emotions. The '*affective*' is NOT the same as '*attitudes*'. The whole field of attitude research is quite enormous and the classic review of the entire area is that of Eagly and Chaiken[111]. There have been, literally, thousands of studies relating to attitudes, their nature, purpose, and measurement. The work has developed enormously and there are now very clear pictures:

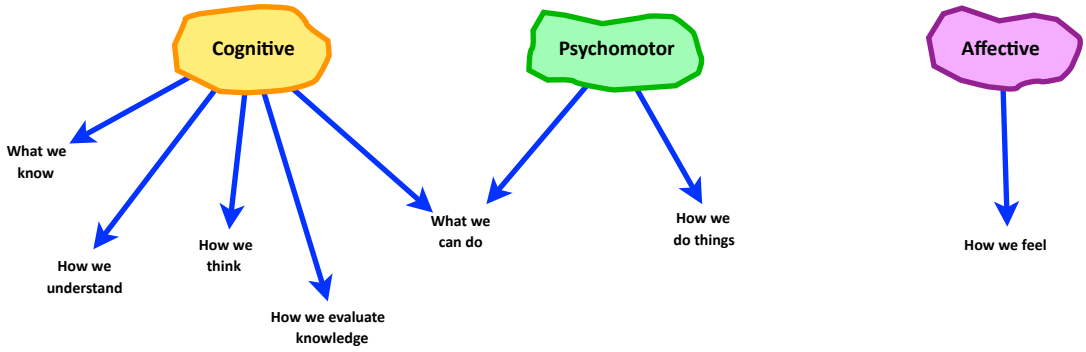
The agreed finding is an attitude can be described like this:

> *'Attitude is a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor'.*
> (Eagly and Chaiken, 1993, page 7)

The key word here is '*evaluation*'. In developing an attitude towards someone or something, we are evaluating in some way. This evaluation may be based on knowledge (cognitive), feeling (affective) or experience (known as conative) or any combination of the three in any proportions. In other words, attitudes are simply NOT the same as the '*affective*' although they may involve the '*affective*'. You can find a very short and simple summary in the section on attitudes in this monograph[112].



While it is true that education involves the total person and this must include feelings and behaviour just as much as head knowledge, the analysis commonly used is still too simplistic and is really misleading. Here is a better model for you to consider:



[111] Chaiken, S. H., & Eagly, A. (1993). The Psychology of Attitudes, Harcourt Brace College Publishers, New York.

[112] Reid, N. (2003) *Getting Started in Pedagogical Research in Higher Education*, LTSN Physical Science, Higher Education Academy, Hull, ISBN 1-903815-07-X, [online: http://www-new2.heacademy.ac.uk/assets/ps/documents/practice_guides/practice_guides/ps0076_getting_sarted_in_pedagogic_research_in_the_physical_sciences_aug_2004.pdf]

Implications for assessment

If education is to value this wide range of skills and outcomes, then it is important assessment gives credit for them. However, in many areas, there are, as yet, no reliable assessment methods which can give fair measurements.

There are good ways to assess what we know, understand and skills in evaluating knowledge. There are developing ways to look at thinking and there are some good ways to consider the things we can do. However, it is more or less impossible to measure emotions or attitudes or, indeed, some aspects of thinking and problem solving in a reliable way. The '*duly-performed*' approach may help in some areas but, while attitudes can be measured very large samples, the accuracy of individual measurement is still unacceptable.

to explore
the potential of
'duly performed' assessment approaches
in
national assessment and certification

RSVP

Assessment Moving *into* a New Future

Features of Quality

Aims

This chapter considers the basic principles that can help us to develop and apply assessments which are of quality in measuring student learning against the goals of the curriculum

Assessment Quality

We all want quality education. We all want quality assessment. The difficulty is in describing accurately what we mean by quality. The first danger is in assuming that quality assessment data can give a reliable indication of quality of education. If we use the data from formal assessment to indicate the quality of education, the quality of schools or the quality of teaching, then we distort all that goes on in schools. This is simply because formal assessment can only measure a very narrow range of skills and, by focussing on these, the entire education process will neglect other, perhaps more important, skills that cannot be measured using formal examinations.

This chapter will focus on seeking to develop quality in our assessments. Our assessment will have a large impact on the futures of our students. It is, therefore, essential that we develop the best assessment of the highest quality possible.

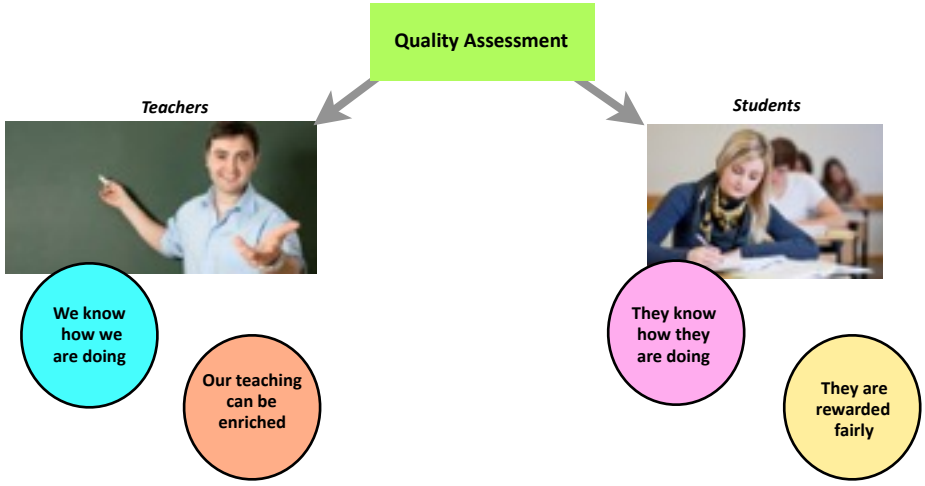


Figure 11.1 Why important assessment in quality key features

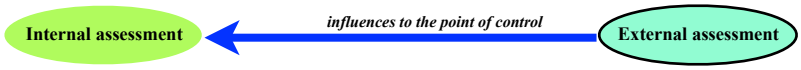
Two Kinds of Assessment

There are two broad kinds of assessment which we shall meet:

- '***Internal***' includes using regular grading, record keeping and reporting to parents and to the students themselves; at secondary level, informing decisions about courses to follow where there are options within the school.
- '***External***' includes meeting the requirements of statutory national assessment and certification by examination bodies or for vocational qualifications, selection for further or higher education.

This can be seen:

The problem is that national examinations strongly influence how we assess our students in the school for our task is to prepare our students for these national examinations.



Key Features of Quality

In this, we need to start by looking at what we want to test and ensuring that this is an appropriate set of goals. We then select the appropriate method to measure against these goals so that we have evidence of student competency. We want our students to perform at their best to show to us that they have made significant achievements. A good way forward is:

- Teachers clearly identify the key desired characteristics and learning outcomes
- Teachers inform their students of what is expected of them
- Teachers develop assessment tasks to match the desired characteristics and learning outcomes

There are four aspects we need to consider in developing quality assessment:

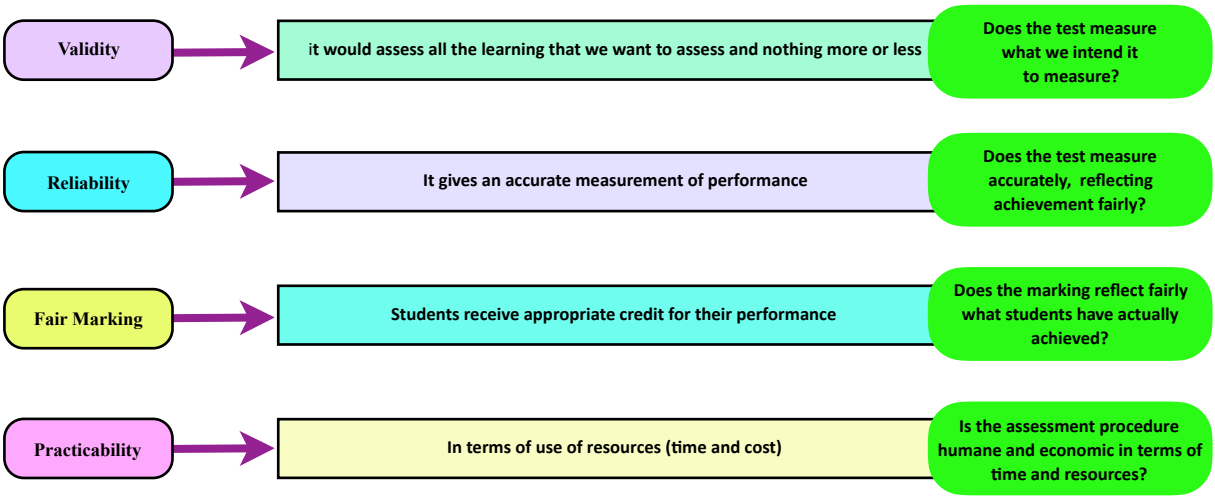


Figure 11.2 Factors of quality assessment

The ideal of satisfying all these requirements is, unfortunately, very difficult, for they are often in conflict. For instance, reliability and validity interact: the greater the demand for '*accuracy*' (high reliability), the greater the tendency for the assessment to be reduced to aspects of performance that can be unambiguously judged as correct or incorrect, which restricts the range of skills to be measured and hence the validity of what is assessed. Costs cannot be ignored: marking every assessment twice would improve accuracy but is out of the question on a large scale. We now look briefly at what each of these requirements means in preparation for considering how they can be achieved.

Assessment Confidence

Overall, the key thing: can we have confidence that the assessment reflects the achievements of the students? Assuming that a test does, in fact, test what it is intended to test, then confidence in its results relate to:

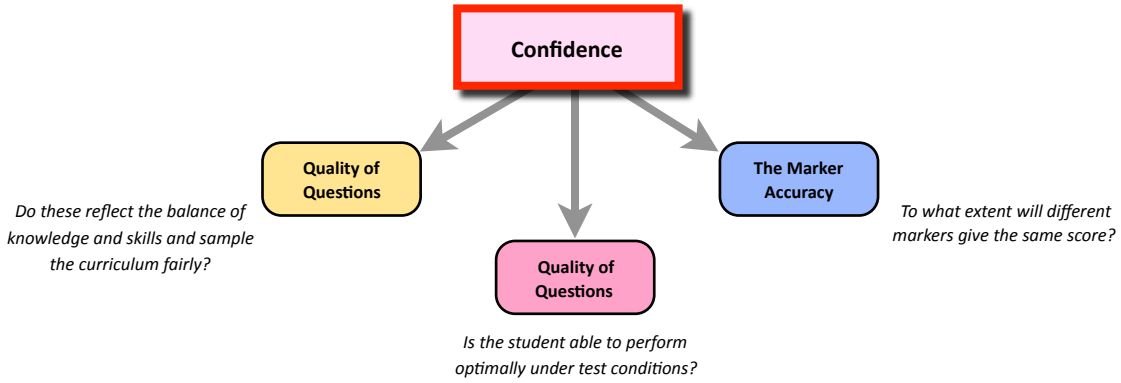


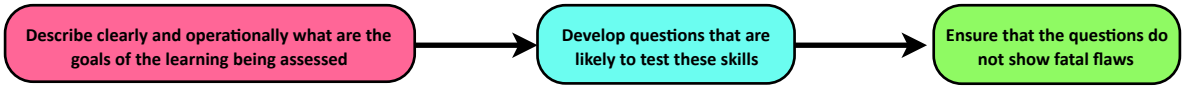
Figure 11.4 Assessment Confidence

Numerous studies have shown that different tests, or different markers, can make an enormous difference to grades, even with carefully designed tests and well briefed markers[113]. This suggest multiple measures are important, with multiple markers.

Practicalities

Any assessment is only as good as the test questions asked. Test questions cannot be good unless they are testing what we want to test in a reliable way.

Here are the key stages to explore this:



Think of an example. Imagine that your students have just completed a section of work on the topic of redox chemistry. Your aims might include the following:

The students should:

- Understand the nature of oxidation and reduction in terms of electron flow (*conceptual understanding*);
- Be able to balance ion-electron half-equations correctly (*an intellectual-practical skill*);
- Be able to combine ion-electron half-equations correctly (*a skill of application*).

[113] Spencer. E., (1981) Inter-Marker Unreliability in SCE O Grade English Composition. Is Improvement Possible? *Scottish Educational Review*, 13(1), 41-55

In order to assess *conceptual understanding*, a structural communication grid question[114] might be the best format. Multiple choice questions will be useless while short answer questions might simply test recall. However, the only way to test if they can *balance* and *combine* ion-electron half-equations correctly is to ask them to demonstrate these skills by giving them two or three such equations and asking them to show the skills. The first question might be very easy, the next a little more difficult and the next one quite difficult, allowing differences in ability to show.

Note that none of the three aims reflect recall skills although they have to remember the skills of *balancing* and *combining* ion-electron half-equations. The key is being able to *carry out* these skills correctly.

Quality Questions

Here is a simple way to look at questions to see if they are of good quality. It can be carried out on your own but it is much better if two or three of you look at the questions together and discuss how you see the questions. Imagine you have set a school examination paper in the topic of simple direct current circuits. As before, you need to start out by listing what are the aims of this part of the course. In other words, what do you want you students to be able to do at the end of this piece of work?

Now, work with two or three of your colleagues. Together look at each question in your draft test paper and ask three questions of each question:

What does this question test?	Knowledge, understanding, skills, thinking or evaluating ?
Is it worth testing?	From the perspective of the learner ?
Are there any obvious flaws in the question?	Information overload, implicit answers, ambiguity.

We often set questions which aim to test a particular outcome but, in fact, evidence shows again and again that the learners get answers by routes we did not expect. The most common flaw in questions is that they simply end up in testing recall or recognition when we think they are testing something else.

The next step is even harder to face. For the moment, forget your own subject discipline and your own commitment to that discipline. Think of the learner. For them, is the question asking something worth testing? Has it value or relevance for them? Does the evidence from this question take the learner forward in their development?

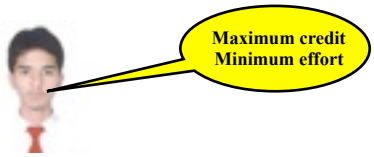
Finally, if the question passes the first two stages, look afresh at the question. It is amazing how many flaws exist in test questions. If the question is overloaded with information, we may be simply testing the capacity of the learner's working memory. There may be '*clues*' in the wording of the question which point to the obvious answers. There may be ambiguities so that the more able learner sees something deeper in the question and misses the '*right*' answer.

Good questions (or tasks) are difficult to set and the problem is greater in some subjects than others. It is useful to take the questions you have set and ask others to apply the three questions to what you have done. It is relatively easy to set questions which simply test recall and recognition. We all quickly discover how demanding it is to set good questions which test other skills.

[114] Examples of this approach will be shown later (chapter 12)

There are several steps we can take to enhance the questions we set so that they measure what we intend them to measure, they measure something useful for the perspective of the learner and they avoid most of the major flaws. The first step is to be honest with ourselves in appreciating that, as soon as we move beyond testing recall, ***good questions are difficult to develop***. This is no weakness on our part; it is simple realism!

Secondly, we need to keep the learners always at the focus of all our thinking. We need to ask how they will see the question, from the perspective of their age, experience and what they have been taught. Most learners at school stages want to gain the *maximum* credit with the *minimum* effort and they will find all kinds of devious, but usually legitimate, ways to achieve this. It is worth the effort to listen to what the learners say, perhaps after an assessment.



Thirdly, we need to work with colleagues in an affirmative way. Experienced colleagues in the same subject discipline can offer perceptive insights while colleagues in other disciplines can often see things which escape our notice altogether.

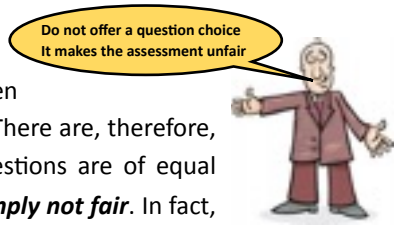
The quality is only as good as the validity of the assessment tasks set:

- **Does it test what we think it is testing?**
- **Can pupils get right answers by other routes than those we intend?**

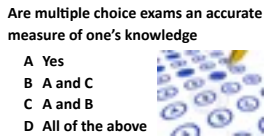
We now need to turn to our learners and how they respond as they complete the assessment. Has the test or examination (in whatever form it takes) given them an opportunity to show their genuine knowledge, understandings, skills, thinking or evaluation.

Of course, learners can cheat but this is not what we are thinking of here mainly. Let us look at some practical issues:

- If there is a choice of questions in a test paper, then all students are *not* sitting the *same* assessment. For example, if the examination paper requires the learners to complete five question out of a set of eight, then there are 56 ways by which the students can select their five questions. There are, therefore, not all sitting the same examination. Unless we know that all the questions are of equal difficulty and are testing things of equal value, then the assessment is ***simply not fair***. In fact, we cannot know either of these with certainty although the first can be handled easily using simple statistics.



- The test format we use may be inappropriate. For example, multiple choice is known to be very unreliable as a test method. In some research, it was found that many right answers were chosen for the *wrong* reasons. In another study, it was found that simply changing the order of the distractors (set of possible answers) changed the level of performance quite markedly[115].



- A test may be too difficult or too easy, giving some learners a disadvantage. Able learners may be '*insulted*' by a trivial paper and may not show their true abilities while different learners tend to panic to different extents in a too demanding paper.

[115] Friel, S. and Johnstone, A.H. (1978a) A Review of the Theory of Objective Testing, *School Science Review*, 59, 733-8.
Friel, S. and Johnstone, A.H. (1978b) Scoring Systems which allow for Partial Knowledge, *Journal of Chemical Education*, 55, 717-9.
Friel, S. and Johnstone, A.H. (1979a) Seconds Thoughts in Multiple Choice Tests in Science, *Journal of Chemical Education*, 56, 326.
Friel, S. and Johnstone, A.H. (1979b) Does the Position of the Answer in a Multiple Choice Test Matter?, *Education in Chemistry*, 16(6), 175.

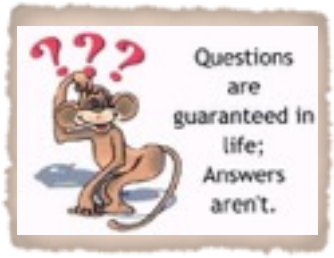
- The working memory demand of questions may favour some students. The research evidence is very clear here. In *almost all* assessments, there are questions which require the candidates to hold or process too many things *at the same time*. Those with above average working memories have a considerable advantage and this is nothing to do with their abilities[116]. It is simply a function of the way the brain is '*wired up*'.

My brain is about to explode

- The style of assessment may favour students with certain learner characteristics. There are numerous aspects to this. One example considers visual-spatial abilities. It has been found that possessing enhanced visual-spatial abilities always seems to favour better test performance[117].

Use many different styles of questions

- There is, of course, the '*good day, bad day*' aspect. Some learners perform better in the morning, some have '*off*' days, performance on specific day may be influenced by events outside the classroom.



In some subjects, it is relatively easy to mark fairly consistently so that any differences between markers are very small. In other subjects, this is difficult to achieve and differences between markers can be very large - enough to change a good pass into a '*fail*' or a borderline pass into an 'A' pass.

The problem lies in the great difficulty to specify what are the criteria for success unambiguously so that all markers can award credit in a similar fashion.

Practical Things you can do

Here are some practical ways forward:

- ***Set up a marking brief, outlining as precisely as possible how marks are to be allocated***.
  - This should be written, discussed, agreed and followed;
  - In some subjects, it is possible to allocate every mark;
  - In other subjects, it is extremely difficult:
  - Where it is difficult, try to *describe* as clearly as possible what you understand by, say, '*five marks worth*', '*three marks worth*' and '*one mark worth*' (assuming credit is out of 5).

[116] Reid, N. (2009) Working Memory and Science Education, *Research in Science and Technological Education,* 27(2), 245-250.
[117] Hindal, H., Reid, N. and Whitehead, R. (2013) A Fresh look at High Ability, *International Journal of Instruction*, 6(1), 59-76.

Enhancing reliability

- There is slippage with time: we get tired and the marking of the last paper is inconsistent with the first. To check for this, re-mark the first marked paper again at the end;
- Individualised interpretation of criteria: we need to follow *agreed* criteria. This is not easy, especially in subjects like English. We need to surrender our legitimate personal value judgements for agreed value judgements - not easy.
- Marker bias - we know the learners too well! Be aware of it and aim for objectivity. It is too easy to upgrade marks subconsciously when the paper is beautifully presented and we know the learner's commitment.

Practical Approaches

- Set a paper with clear marking brief in mind;
- Check paper and marking brief with colleagues;
- Ensure that learners know test aims and format;
- Ensure that test conditions are acceptable;
- Ensure that all markers use the marking brief.

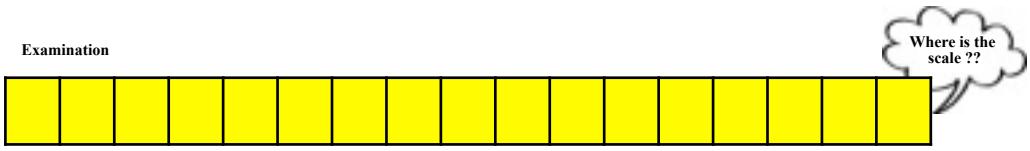
Standards

Almost every day, there are references in the media to standards in school education. Indeed, we probably all agree that we wish to develop the highest standards possible for our learners. One major problem is how do we define '*standards*'?

The danger is that we measure what we are capable of measuring and then we define our '*standards*' by that measure. Thus, it is too tempting to define standards by examination grades. This is how politicians and the media see it. Sadly, it is often how educational managers and school inspectors see it as well.

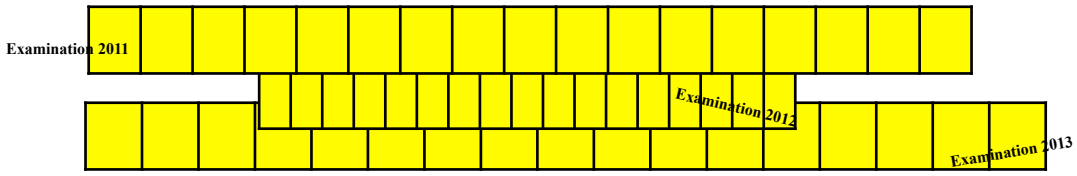
The problem is that examinations, no matter how well set and marked, are very imperfect measuring instruments. It is like trying to measure your height with a measuring tape on which no number scale has been marked. If we have a group of people, then we can place them in rough order by their height. That is all.

Think of examinations like a measuring tape with no scale on it …..



Examinations are like that. They are quite good (but by no means perfect) in placing learners in some kind of order of merit. However, where to place the '*pass mark*' or the grade boundaries is a matter of human judgement.

Therefore, statements in the media which suggest that standards have improved or otherwise on the basis of grades in some national examination are ***simply nonsense***. The examination papers are not identical from year to year. We now have more than one measuring tape and none of them has a number scale marked on it!



At the national level, it is a reasonable assumption to think that the ability of the candidates in any examination will be roughly the same from year to year. It is that which is used to decide pass marks and grade boundaries. That is all we have and, in subjects with high numbers (> 1000), it is a very sound basis for taking decisions.

At the school level, the numbers are too low to expect the general abilities from one year to the next to be similar. Indeed, it only takes a fairly superficial statistical analysis to show that the performance in an average size school at national examinations can vary by as much as ± 30% over a ten year period. This is simply caused by the natural variation in year groups and is nothing to do with the '*standards'* in the school. Therefore, any attempt to compare overall performance in a school from year to year (and then blame headteachers and staff !) is simply nonsense. It is quite unethical and totally unfair.

There is another problem in the media about examination standards. Media commentators base their knowledge of education almost entirely on their memories of their own school days. Education at all levels has changed out of recognition over the years. It is probably true to say that a person has almost certainly '*lost touch*' after no more five years of loss of direct regular contact with a school. And direct contact means working there!

The only certainty about school education is uncertainty: change is happening so fast that an outsider has no chance of keeping up to date.

Examinations place candidates in a rough order of merit
Examinations are not exact measuring instruments

A Different View of Standards

There is another aspect of standards which is closer to us. It is very difficult, perhaps almost impossible, to specify in unambiguous language the standards we wish from our learners. Even when we think we have it, a colleague may interpret what we write in a slightly different way. This means that criteria laid down in curriculum specifications are incredibly difficult to interpret unambiguously. It is also very difficult to interpret the criteria into test questions where we can be sure that then question tests the desired outcome and that it tests it at the standard we wish.

There are two ways we resolve this problem. The first is to look at previous examination papers or exemplar examination questions. The second is to rely on our experience and the experience of our colleagues.

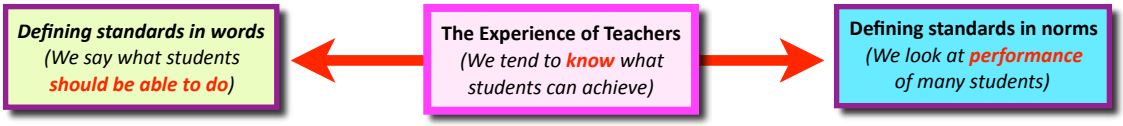
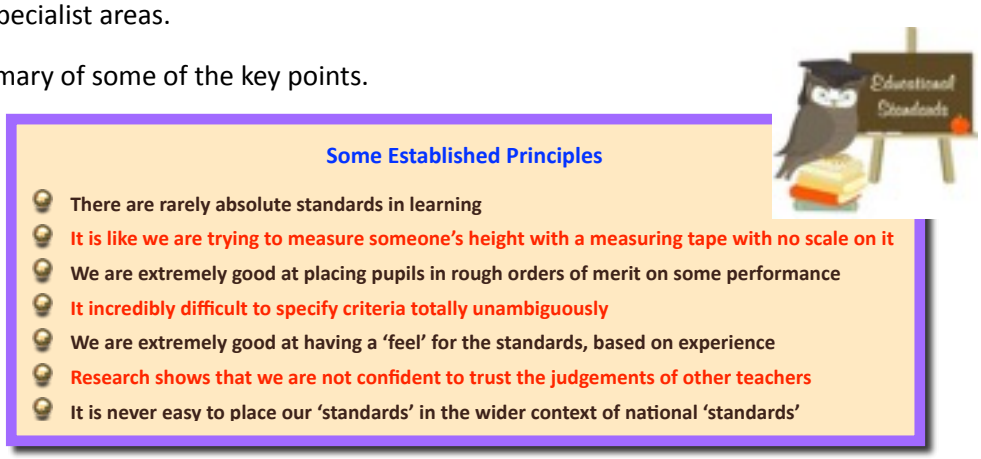


Figure 11.5 Defining the Standards

From experience, we tend to know what to expect, we tend to know what is the likely performance of the more able and what is the critical performance that just gives the '*pass*'. However, that is a value judgement and not exact. Nonetheless, a group of experienced teachers can usually agree on the '*standards*' required in their own specialist areas.

Here is a summary of some of the key points.



In assessment, the following features are important:

Figure 11.6 Features to Achieve Quality Assessment

We start by stating exactly what we want to assess and the standards we expect from our students. We then start to create assessment questions and tasks, continually checking if these are likely to measure against the objectives we have set for the assessment. It is important to check that the questions we set are realistic and meaningful for the students. It is also vital that we check that the questions do not give some students unfair advantages. Thus, we need to be sure that we are not overloading the working memories of students who happen to have slightly smaller working memory capacities.

Setting our Own Assessments

In some parts of the world, *all* assessment at *all* levels in secondary schools is carried out internally within the school. They have developed a system which has been found to assess the learners performance in a robust and authentic way. Universities are found to be able to trust the assessments. Indeed, wider society trusts the assessments. Let us take this step by step and see how it can be done. Firstly, what does research show?

(1) The best way to describe and define the criteria for assessments is for teachers to work together to draw up these criteria, with some support from outside.

(2) Teacher assessments are remarkably reliable much of the time while the greatest reliability is obtained by combining teacher assessments with external assessments.

(3) Teachers working affirmatively together cannot only define criteria but they can also agree on assessment tasks which are valid and reliable.

(4) National examinations are not as reliable as most assume: they are one picture of performance at one moment of time with one sample of questions.

(5) The predictive value of national assessments in term of future performance (eg university degree class) is fairly poor.

(6) In many subject areas (interestingly, the sciences, technologies, engineering and mathematical subjects), Universities are less interested in content covered than in ability to work independently and show key thinking skills and positive attitudes.

Assessment Moving *into* a New Future

Methods of Assessment

Aims

This chapter looks to focus on numerous ways to assess, giving practical examples and also ways of marking so that we reward skills beyond recall

Introduction

The focus of this chapter is on the practical ways we can assess our students in the classroom and in the tests and examinations we develop in our schools and universities. We need to recognise the enormous power of our assessments:

> *'Assessment acts as a mechanism to control students that is far more pervasive and insidious than most staff would be prepared to acknowledge'.*
>
> *Danili and Reid, 2005*[118]

Of course, there are numerous ways by which we can assess the progress of our students. In simple terms,

Interviews	Written Examinations	Observations	Task Completion
Our students tell us what they know	Our students write down what they know	Our students show us what they know	Our student undertake tasks of known value
SPEAK	**WRITE**	**DO**	**UNDERTAKE**
We talk to our students	We read what they write	We watch what they do	We check that the tasks are completed
We LISTEN	**We READ**	**We WATCH**	**We CHECK**
From all this, we deduce what they know, understand, and can do as well as how they think and evaluate			

Many people think that assessment is limited to the paper and pencil tests given in class to see if students have learned the content.

- Assessment is far more than pencil and paper tests:

  We can observe our students in the classroom; we can talk to our students before and during the completion of a task.

- Assessment is far more than measuring what they can recall:

  We need to emphasis their understanding and what they do successfully. We need to consider their skills in thinking and their ability and willingness to evaluate knowledge.

[118] Danili E. and Reid N., (2005) Assessment formats: do they make a difference? *Chemistry Education Research and Practice*, 6, 204-212.

> ***Assessment performances are day-to-day activities that can also be authentic and engaging demonstrations of students' abilities to grapple with the central challenges of a discipline in real life contexts.***
>
> Kulieke et al., (1990)[122]

Classroom Assessment

Here is an important principle:

> ***'Classroom assessment provides valuable information that allows teachers to adapt instructional procedures to the learning needs of their students.'***
>
> (Koyalik, 2002 as cited in Eggen and Kauchak, 2004)[120]

All assessment, including classroom assessment involves a process:

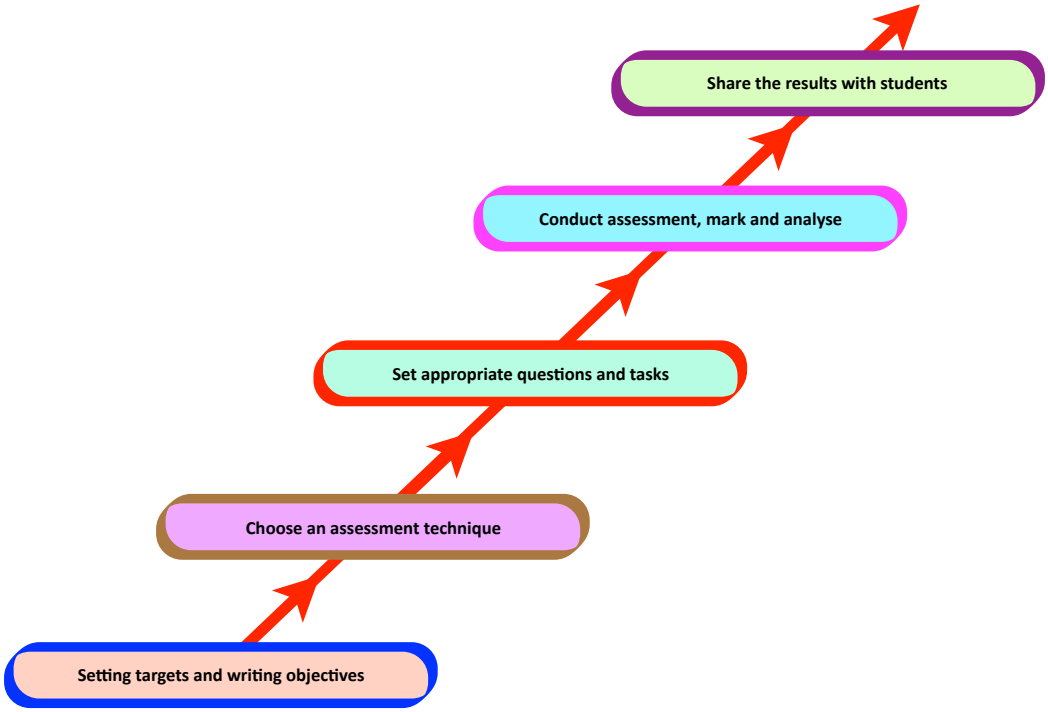


Figure 12.1 Processes in Assessment

This can be summarised:

- ☑ We need to decide want we want to measure
- ☑ We need to decide how we are going to measure it
- ☑ We need devise appropriate ways to measure it
- ☑ We measure and we mark and analyse
- ☑ We tell our students how they got on

Methods and tools may include oral questioning, quizzes, tests, examinations, assignments, projects, essays, observations of performance, and any other product or sample of behaviour that might provide information about how well a student or group of students is doing (Earl and Cousins, 1996)[121].

---

[119] Kulieke, M., Bakker, J., Collins, C., Fennimore, T., Fine, C., Herman, J., Jones, B.F., Raack, L., & Tinzmann, M.B. (1990). *Why should assessment be based on a vision of learning?* [online document] NCREL, Oak Brook: IL. Available online: http://www.ncrel.org/sdrs/areas/rpl_esys/assess.htm

[120] Eggen, P. and Kauchack, D. 2004. *Educational psychology: windows on classrooms*. 6th ed. New Jersey: Pearson.

[121] Earl, L., and Cousins, J. B. (1996). *Classroom Assessment*. Canada: Ontario Public School Teachers' Federation.

Assessment and Testing

We tend to think of assessment as a formal test or examination where students write down what they can recall about some topic or theme. It is critical to remember that the formal written test or examination is **ONLY ONE** kind of assessment. Dietel *et al*., (1991)[122] define assessment as '*any method used to better understand the current knowledge that a student possesses*'. The important word here is '*any*'. However their description is still inadequate. Here is a better way:

Assessment

Any acceptable technique that gives us information about how well our students know, understand and demonstrate their skills, as well as showing how well they are thinking and evaluating.

Assessment affects students. Assessment affects us, as teachers. The way it affects students and teachers is described as feedback. Look at figure 12.2 below. Start in the middle, with the green box entitled '*teaching and learning*'. This leads to assessment. However the assessment feeds back to the students and to ourselves:

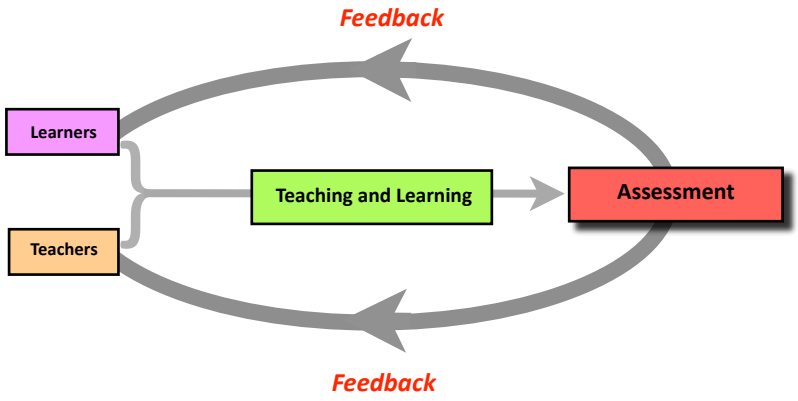


Figure 12.2 Assessment Feedback

Now let us look at the possible effects of that feedback (figure 12.3).

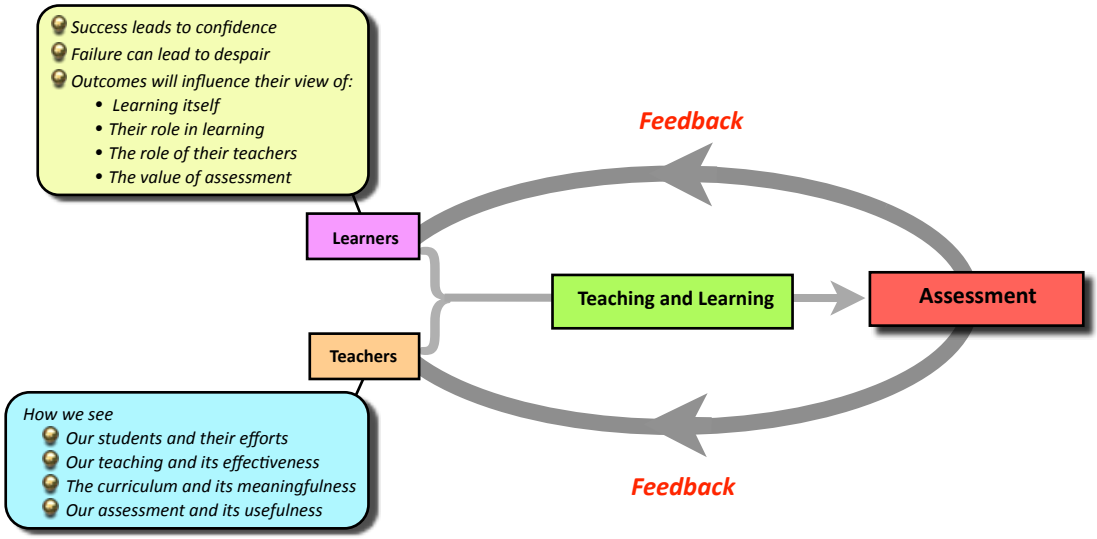


Figure 12.3 Assessment Feedback in Detail

[122] Dietel, R. J., Herman, J. L., & Knuth, R. A. (1991). *What does research say about assessment?* NCREL, Oak Brook. Available online: http://www.ncrel.org/sdrs/areas/stw_esys/4assess.htm

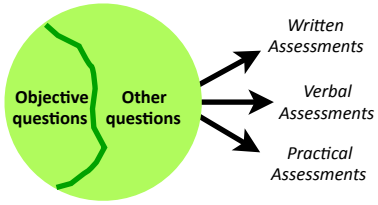
Widening our Horizons

We can obtain assessment evidence from numerous sources:

- Course-embedded assessment (uses exams, class activities, and assignments)
- Portfolio analysis (collection of student work)
- Performance-based measures (activities such as writing an essay, making a presentation, completing a problem -solving exercise, giving a performance, and simulations)
- Surveys, interviews, and focus groups of students, alumni, and employers

Until now there is no one perfect or ideal method of assessment. Each method has its advantages and disadvantages. However, the suitable methods of assessment to be used depend mainly on the outcomes required to be assessed and, therefore, teachers have to think deeply before choosing any assessment method.

We shall now look at examples of the various methods open to us. Let us think of questions set in written test papers. There are two broad categories:



It has to be stressed that all questions are set by us. The only thing that is objective about objective questions is that they can be marked by a machine. However, the machine still has to be '*told*' what are the '*right*' answers. There is ***nothing special about objective questions***.

The table below looks at some of the more common formats of questions, with some comments on their potential. Under '*objective formats*', partial knowledge multiple choice and structural communication girds have very considerable potential in some areas of the curriculum and examples are offered later.

Objective Formats		Other Formats	
Example	**Comment**	**Example**	**Comment**
Multiple Choice	They tend only to indicate recall-recognition while reliability is highly suspect	***Short Answer***	Can indicate what they know, can do or, to a limited extent, what they understand.
Partial knowledge multiple choice	More reliable, can be used to give some indication of understanding	***Extended answer***	Versatile, can be used to explore understanding, skills, thinking and evaluation.
Structural Communication Grids	Powerfully diagnostic, especially good at conceptual areas; cannot be used for everything	***Practical Tests***	Can assess cognitive skills (as in mathematics procedures) as well as practical skills.
There are several other objective formats but these have limited application in courses at school and further education stages. There is ongoing work in exploring formats for assessment which use the power of computers. Much of this work has sadly never moved beyond multiple choice variants, with their well established limitations. However, new formats are being considered and, in due course, may have wider applications. Indeed, electronic assessment is a growing area where we can expect developments in the near future. There may come a day when assessments use mobiles, on-line tasks and iPads. Nonetheless, e-assessment faces the major problem about uncertainty of whether an assessment is completed unaided by others.		***Observation***	Very insightful but very difficult to reduce to scores.
		Calculation	Important in many subject areas to test procedural skills and, occasionally, understanding
		Essay	Can assess almost anything but the problem is how to mark fairly. The marking time demand is considerable.
		Dissertation	Offers enormous scope. Marking needs careful thought but good ways to mark have been developed.
		Project	This can reflect extended work and offers many insights. Reducing to a score needs thought.
		Verbal Presentation	Very time-consuming but gives rich insights. Marking needs careful thought but good ways to mark have been developed.
		Duly Performed	Credit is given for task completion and no traditional marking is involved. A very useful way forward to assess some intractable skills.

In making a choice of which format to use, there are two important aspects to be considered:

- Different formats have different strengths and weaknesses. It is important to select the format which will suit the task in hand;
- In selecting a format, ask yourself if the learner answers will provide you with the information you want. This needs to be related to your aims in the assessment.

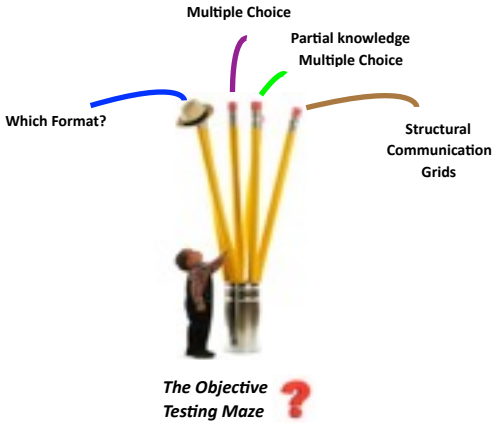
We now look at each of these assessment methods and discuss their strengths and weaknesses, giving examples of them where they are less familiar.

Objective Assessment Formats

Some argue that objective testing removes tester bias (Currie, 1986;[123] Leonard *et al.*, 1996[124]). However, this is not really true as we have to be very clear about the word *objective*. In objective testing, the curriculum goals that are intended to be tested are *subjectively* chosen (chosen by the teachers). The questions themselves have been *subjectively* written to match these goals (a professional judgment made by the teacher). The wording used is *subjective* (chosen by the teachers to test the goals), and the interpretation of the scores is *subjective* (the results are *subjectively* judged by the examiner, as well as parents and employers).

The only objective part of objective testing is the scoring because machine scoring can be used

The three most common examples are multiple choice, partial-knowledge multiple choice and structural communication girds. However, other possibilities exist: true-false questions, matching questions, fill-in questions. However, these will not be considered as they are far too limited for use in schools and universities.



[123] Currie, R. (1986) 'Objective tesing' in Lloyd - Jones, R., Bray, E., Jonson,G. and Crrie, R (eds,) *Assessmest from Principles to Action*, London, Macmillan Education Ltd., 47-54.

[124] Leonard, L., Macteer. E. and Wilson, R. (1996) *Report of objective testing study group,* Glasgow, Institute of Biomedical and Life Sciences, University of Glasgow.

Multiple Choice Questions

These are familiar to you and involve a question with four or five possible answers [known as the *key* (right answer) and *distractors* (wrong answers)] being offered, only one of which is right. Let us look at the advantages and disadvantages:

Advantages

- They allow a wide range of areas to be tested fairly quickly.
- With computers, they can be marked almost instantly.
- With computers, patterns of choices can be seen very easily.

Disadvantages

- It is very difficult and time-consuming to set good questions.
- Such questions tend to test mainly recall-recognition skills, despite many attempts to go beyond this.
- Many right answers are chosen for the wrong reasons and many wrong answers are chosen for good reasons.
- Changing the order of the distractors can alter performance quite markedly (showing lack of reliability).
- Guessing is a serious problem and using 5 distractors shows little advantage over 4.
- It is far too easy to offer more than one right answer, leaving the candidates in confusion.
- It is possible to develop skills in students so that they can do much better in such questions without any increase in subject knowledge.

Recommendation

- Reduce the use of multiple choice as this method shows serious flaws.
- Adapt multiple choice questions to partial knowledge multiple choice questions is a good way forward.

Here are two examples:

A Bad Example	***A Better Example***
If ammonium chloride crystals are dissolved in water, the pH of the solution will be:	Look at the vanadyl ion: VO_2^+ What is the oxidation state of vanadium?
A Above 7 B 7 C Below 7 D None of these	A +1 B +3 C +4 D +5

Comments

In the first question, there can be only three possibilities. Therefore, D is a meaningless choice.

In the second question, this seeks to test if the candidates really understand the meaning of the concept of oxidation state.

Option A suggests that there is confusion with charge.

Option B suggests there is confusion with valency.

Option C suggests there is confusion with bonding.

Partial Knowledge Multiple Choice Questions

Research showed many years ago (Friel and Johnstone[125]) that a major problem with multiple choice lay in the fact that they gave no indication of partial knowledge. An answer was either right or it was wrong. Numerous methods were developed to reduce this problem and one this is illustrated here. Your own multiple choice questions can easily be adapted to this format.

Suppose four answers are offered.

(B) is the correct answer.

The candidate is asked to mark with a tick (✔) the correct answer and put a cross (✘) against the ***two*** answers which are definitely wrong.

Consider three answers offered by three students:

	Student 1	*Student 2*	*Student 3*
A	✘	✘	
B	✔		✘
C	✘	✔	✘
D		✘	✔

Student 1 Has ticked the correct answer and crossed two wrong answers: full marks (perhaps 1)

Student 2 Has ticked a wrong answer but has put a cross against the other two wrong answers: partial marks (perhaps 0.5)

Student 3 Has ticked a wrong answer but put a cross against the correct answer: no marks (0)

Advantages

- They allow a wide range of areas to be tested fairly quickly.
- With computers, they can be marked almost instantly.
- With computers, patterns of choices can be seen very easily.
- They give credit for partial knowledge.

Disadvantages

- It is difficult to set good questions.
- Such questions tend to test mainly recall-recognition skills, despite many attempts to go beyond this.
- It may be possible to develop skills in students so that they can do much better in such questions without any increase in subject knowledge.

Recommendation

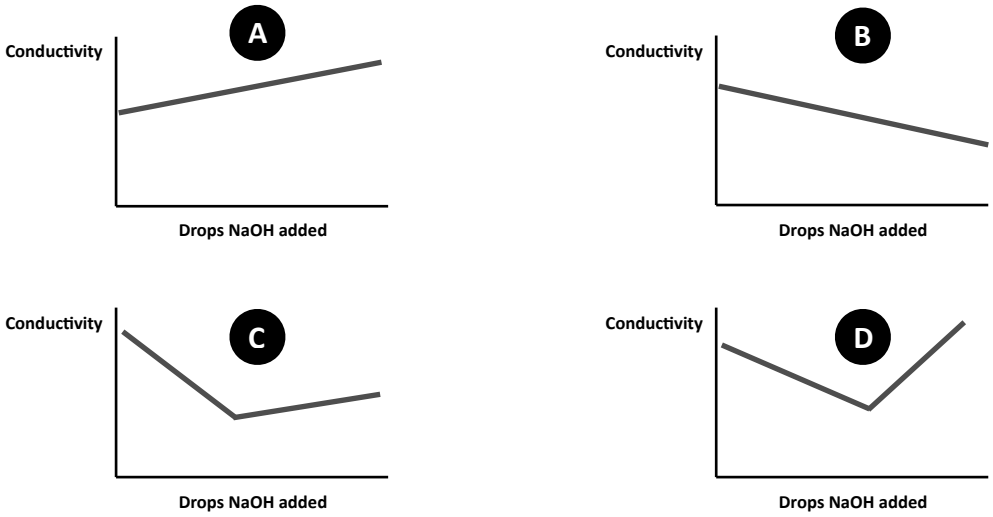
- Adapt your multiple choice questions to partial knowledge format.

[125] Friel, S. and Johnstone, A.H. (1978a) A Review of the Theory of Objective Testing, *School Science Review*, 59, 733-8.
Friel, S. and Johnstone, A.H. (1978b) Scoring Systems which allow for Partial Knowledge, *Journal of Chemical Education*, 55, 717-9.
Friel, S. and Johnstone, A.H. (1979a) Seconds Thoughts in Multiple Choice Tests in Science, *Journal of Chemical Education*, 56, 326.
Friel, S. and Johnstone, A.H. (1979b) Does the Position of the Answer in a Multiple Choice Test Matter?, *Education in Chemistry*, 16(6), 175.
Friel, S. and Johnstone, A.H. (1988) Making Test Scores Yield More Information, *Education in Chemistry,* 25 (2), 46-49.

A Good Example

Mark with a tick (✔) the correct answer and put a cross (✘) against the ***two*** answers which are definitely wrong.

If dilute sodium hydroxide is added drop by drop to a solution of very dilute sulphuric acid, which graph shows how the conductivity changes?



Comments

The conductivity is controlled by the number of free ions present in the solution and the mobility of these ions.

More ions are being added and, in option A, the candidate sees the conductivity rising, ignoring any chemical reaction.

In option B, the candidate sees the hydroxide ions linking to the hydrogen ions (as water) and the ion concentration falling, but ignores the reaction completion.

In option C, the candidate appreciates the reaction in removing water but then grasps that further addition of sodium hydroxide increases the number of ions present.

In option D, the candidate appreciates the reaction in removing water but then grasps that further addition of sodium hydroxide increases the number of ions present, but fails to see that sodium and hydroxide ions are less mobile than the hydrogen ions just removed.

The pattern of ticks and crosses reveals where the confusion lies in the candidate's mind.

Exploring genuine understanding …..

Structural Communication grids

This method was first described long ago in 1972 by Egan[126]. It was used by many and found to be a very powerful assessment technique[127]. The method is now widely used in formal examinations and in research, but the use tends to be in the science subjects[128]. However, there are good possibilities in some other areas of the curriculum.

In structural communication grid questions, the data are presented in the form of numbered boxes laid out for convenience in a grid format. The contents of the data can be numbers, words, phrases, pictures, equations, formulas, or diagrams.

1	**2**	**3**
4	**5**	**6**
7	**8**	**9**

In essence,

- You provide an array of information of any style
- Students have to make a selection to answer questions
- The information is usually set out in a grid arrangement
- Students can also be asked to arrange their selection in order

Respondents are asked in response to a question by considering the content of each box and decide which box or combination of boxes constitutes the most appropriate solution(s) to the question. In answering the questions by selecting the appropriate boxes, a respondent (Johnstone, 1988[129]):

> *'…has stamped his structure upon the random boxes of information to communicate his understanding of the material being tested: hence the name 'Structural Communication'…'*

The appropriate size of the grids can be chosen according to the age of the population using it. For age 12-15, grids with nine boxes (3 x 3) have been found to be appropriate[130] (Johnstone and Ambusaidi, 2001; 2000; Johnstone *et al*., 2000). The larger grids (3 x 4 or 4 x 4) can be used with higher level and the largest grids that was used on undergraduates contain twenty boxes (4 x 5)[131] (Bahar, 1999).

---

[126] Egan, K. (1972) Structural communication - A new contribution to pedagogy, *Programmed Learning and Educational Technology*, 1, 63-78.

[127] Johnstone, A.H. (2003) Effective Practice in Objective Assessment, available at: https://www.heacademy.ac.uk/resource/effective-practice-objective-assessment

Bahar, M. and Hansell, M. H. (2000) The relationship between some psychological factors and their effect on the performance of grid questions and word association tests *Educational Psychology*, 20(3), 349-367.

Johnstone and Ambusaidi (2001) Fixed response questions with a difference, *Chemistry Education Research and Practice*, 2(3), 313-327

[128] Chen and Whitehead (2009) Understanding physics in relation to working memory, *Research in Science and Technological Education*, 27(2), 151–160.

Hassan, A.K., Hill, R.A. and Reid, N. (2004) Ideas Underpinning Success in an Introductory Course in Organic Chemistry, *University Chemistry Education*, 8, 40-51.

Johnstone, A. H., Bahar, M. and Hansell, M. H. (2000) Structural communication grids: A valuable assessment and diagnostic tool for science teachers. *Journal of Biology Education,* 34(2), 87-89.

Bahar, M. (1999) Investigation of Biology Students' Cognitive Structure through Word Associated Tests, Mind Maps, and Structural Communication Grids, PhD Thesis, University of Glasgow.

Scottish Examinations Authority (1997): *www.sqa.org.uk/*

Johnstone, A.H. and Mughol, A.R. (1979) Testing for Understanding, *School Science Review*, 61, 174-50.

Duncan, K. D. (1974) *Analytical techniques in training design*. In Edwards, E. and Lees, F. P. (Ed) The Human Operator in Process Control, London, UK: Taylor and Francis, 283-319.

[129] Johnstone, A. H. (1988) *Methods of assessment using grids*. Lab Talk, 10, 4-6.

[130] Johnstone and Ambusaidi (2001) Fixed response questions with a difference, *Chemistry Education Research and Practice*, 2(3), 313-327.
Johnstone, A. H., Bahar, M. and Hansell, M. H. (2000) Structural communication grids: A valuable assessment and diagnostic tool for science teachers. *Journal of Biology Education,* 34(2), 87-89.

[131] Bahar, M. (1999) *Investigation of Biology Students' Cognitive Structure through Word Associated Tests, Mind Maps, and Structural Communication Grids*, PhD Thesis, University of Glasgow.

Advantages

- Guessing is virtually eliminated because the student does not know in advance how many boxes contain correct answers.
- Text, diagrams, pictures, numbers....can all be used.
- With computers, patterns of choices can be seen very easily.
- The correct responses reveal something of students' insights of conceptual understanding, area of interest, or student knowledge gaps.
- The wrong answers reveal something of student insights of misunderstandings and misconceptions.
- The approach is a powerful way to assess conceptual understanding, avoiding simply testing recall-recognition skills.
- You can ask many questions using one grid, gaining useful insights into many aspects of some concept or area of interest.
- The problems associated with working memory overload tend to be minimised.

Disadvantages

- It is difficult to set good questions but easier than multiple choice.
- Students find such questions mentally taxing and they tire easily: tests cannot contain many such questions.
- They do not work in all areas of the curriculum.

Recommendation

- Develop such questions and use them.

Scoring

Egan[132] suggested a formula:

$$\text{Score} = \frac{\text{Number of correct answers selected}}{\text{Total number of correct answers}} - \frac{\text{Number of wrong answers selected}}{\text{Total number of wrong answers}}$$

This has been used and found to work well[133]. This is the score for each individual question. The maximum score obtainable is +1 and the minimum is -1. Scaling may be necessary to ensure all students gain positive scores. It was been found that the method is highly discriminating: It separates very markedly those students who really understand the topic from those who do not.

A computer can handle all this very well and the method gives vital information of areas where there are confusions in understanding, thus helping teachers to modify teaching to address these areas of difficulty.

Because the method is unfamiliar, three examples are now shown.

[132] Egan, K. (1972) Structural communication - A new contribution to pedagogy, *Programmed Learning and Educational Technology*, 1, 63-78.

[133] Hassan, A.K., Hill, R.A. and Reid, N. (2004) Ideas Underpinning Success in an Introductory Course in Organic Chemistry, *University Chemistry Education*, 8, 40-51.

Three Structural Communication Grids

A Mass of a mole is 18g	**B** Mole of molecules of hydrogen	**C** Mass of a mole is 1g
D Mass of a mole is 17g	**E** Two hydrogen atoms linked together	**F** A mole of molecules of water
G A mole of atoms of hydrogen	**H** Two hydrogen atoms linked to one oxygen atom	**I** A mixture of hydrogen and oxygen atoms

Select ***all*** the box(es) where there are statements which are:

(a) **True** about the formula: **H_2O**? .....

(b) **True** about the formula: **H_2**? .....

A Slightly	**B** Practice	**C** Went
D Attractive	**E** Inevitable	**F** Gently
G Practise	**H** Share	**I** Set

Select ***all*** the box (es) which contain words that can function as:

(a) Adjectives .....

(b) Nouns .....

(c) Verbs .....

(d) Adverbs .....

A Sulphur	**B** Zinc	**C** Titanium
D Oxygen	**E** Magnesium	**F** Nitrogen
G Calcium	**H** Carbon	**I** Hydrogen

Select ***all*** the box (es) which contain elements which

(a) Are metals ……..

(b) Are gases at room temperature ……..

(c) Present in limestone ……..

(d) Form dioxides ……..

(e) Are present in the human body in large amounts ……..

(f) Are present in ammonium nitrate fertilisers ……..

Short and Extended Answer Questions

This is an area with which we are very familiar. It can vary from a question worth 2 marks right up to an essay worth, perhaps, 20 marks.

We have, perhaps, faced questions like:

(a) Explain what is meant by osmotic pressure and how it affects living cells (4 marks)

(b) Describe two outcomes that can be observed as evidence of glaciation (4 marks)

(c) Explain what is happening when an iron nail is dropped into a solution of copper sulphate (3 marks).

…. right up to huge (and demanding) questions like:

(d) Write an essay on some of the key voting systems used in elections in the world, showing relative strengths and weaknesses (20 marks).

(e) Give an account of Newton's First Law and show its importance in everyday life (10 marks).

The interesting issue is how to mark such questions.

One way is to give one mark for each correct statement, comment or explanation. This may work well for short questions. It is not very good for marking a long question worth 20 marks. Here is a possible way to mark a longer question.

Imagine your students have just complete a course on sulphuric acid, perhaps the single most important chemical in chemical industry. Imagine this question:

Write an essay on the industrial and social importance of sulphuric acid. (25 marks)

Here is a possible marking brief.

Outcomes	Evidence	Mark	Mark Allocation
Awareness of social importance	Has the candidate shown that they are aware of where sulphuric acid has a vital importance in life	5	5 - for clear evidence that candidate is fully aware 3 - for some evidence that candidate is fully aware 1 - for limited evidence that candidate is fully aware
Awareness of key elements of industrial production	Key raw materials, energy problems, production methods, environmental issues	5	5 - for clear evidence that candidate is fully aware 3 - for some evidence that candidate is fully aware 1 - for limited evidence that candidate is fully aware
Awareness of importance of the acid in wider industrial production	Has the candidate shown that they are aware how the acid is essential for other industrial production	5	5 - for clear evidence that candidate is fully aware 3 - for some evidence that candidate is fully aware 1 - for limited evidence that candidate is fully aware
Good start, good conclusion		5	5 - arresting start, focussed conclusion 3 - adequate start and conclusion 1 - poor start and conclusion
Sequence of ideas	Evidence of logical flow of ideas and clear thought	5	5 - excellent logical flow of ideas and clear thought 3 - adequate logical flow of ideas and clear thought 1 - poor logical flow of ideas and clear thought
	Maximum Mark =	**25**	

Note: Although the description for allocation do 1,3 or 5 marks is shown, 2 or 4 marks can also be allocated as '*in-between*' performance.

Assessing Skills of Doing

We may assess our students by asking them to demonstrate a skill, by observing them carrying out some procedure, or by asking them to give a brief presentation.

The key thing is:

(a) Define exactly what you want to measure.

(b) Find a way to make the measurement in a humane and reliable way.

(c) Develop a way to allocate marks.

Let us look at examples:

Suppose we want our students to develop the skill of being able to present their understandings and findings on a topic by giving a very short talk.

We give them 3 blank overhead projector transparencies and pens.

The talk is to last no more than 5 minutes and must be presented using the overhead projector.

Imagine that you have asked each student in your class to prepare this talk where they will choose a chemical element and present all that they have found out about this element (using the internet, books etc), with special emphasis on the importance of the element for humans today. Suppose each student in your class has been allocated a different element. If you have 30 students then, in 2½ hours, your entire class has taught each other about 30 elements!! And they will have fun as well.

Here is a possible way to score the talk.

Aspect	Mark	Criteria to be Rewarded	Marking Guideline
Good introduction	2	Mark in terms of gaining audience interest	2 - excellent; 1 good; 0 poor
Good ending	2	Mark in terms of a memorable conclusion	2 - excellent; 1 good; 0 poor
Relevant content	5	Mark in terms of the way the presenting the information to show the importance of that element for humans today	5 - coverage was balanced and relevant; 3 - coverage was adequate; 1 coverage was poor
Quality of OHP transparencies	5	Mark for clarity in communicating	5 - excellent; 3 - adequate; 1 - poor
Clarity of verbal presentation	5	Mark for lucidity, logical clarity, holding attention of audience	5 - excellent; 3 - adequate; 1 - poor
Overall effectiveness	6	Mark in terms the overall effectiveness of the presentation in meeting its goals (which may vary from element to element)	6 - excellent; 4 - good; 2 - just adequate; 0 - poor
Maximum Mark	**25**		

Dissertations and Projects

The marking of dissertations and projects can be illustrated by a method used at university level:

Essays at Masters Level

	Excellent	Acceptable	Poor	Not present
	3	2	1	0
Presentation of essay (its overall accessibility for reader)				
Quality of start and conclusion (purpose and achievement of purpose)				
Logical flow (sequencing of relevant ideas)				
Coverage (was the field covered adequately?)				
Clarity of expression (could it be read for meaning easily?)				
Literature coverage (adequate coverage and relevant use)				
Critical thought (relating to literature and ideas developed)				
Originality (evidence of fresh thinking)				

Total (out of 24):

Grade awards:

>19	A	*Excellent*
18-19	B	*Very Good*
16-17	C	*Good*
14-15	D	*Satisfactory*
12-13	E	*Weak*
<12	F	*Poor*

Dissertations at Masters Level

	Excellent	Acceptable	Poor	Not present
	3	2	1	0
Presentation of dissertation (its overall accessibility for reader)				
Quality of start and conclusion (purpose and achievement of purpose)				
Logical flow (sequencing of relevant ideas)				
Coverage (was the field covered adequately?)				
Clarity of expression (could it be read for meaning easily?)				
Literature coverage (adequate coverage and relevant use)				
Critical thought (relating to literature, ideas developed, or empirical areas)				
Originality (evidence of fresh thinking)				
Empirical relevance (does research question lead to experimental lead to conclusions?)				
Empirical validity (sampling, methodology, experimental structure)				
Data handling (is it reasonable?)				
Contribution to knowledge (does it take us forward?)				

Total (out of 36):

Grade awards:

>27	A	*Excellent*
26-27	B	*Very Good*
23-25	C	*Good*
20-22	D	*Satisfactory*
20-21	E	*Weak*
<20	F	*Poor*

It is possible to use a similar system at school level:

	Excellent	*Acceptable*	*Poor*	*Not present*
	3	2	1	0
Presentation of report (its overall accessibility for reader)				
Quality of start and conclusion (purpose and achievement of purpose)				
Logical flow (sequencing of relevant ideas)				
Coverage (was the field covered adequately?)				
Clarity of expression (could it be read for meaning easily?)				
Literature coverage (adequate coverage and relevant use of resources)				
Experimental (competent report of undertakings)				
Critical thought and originality (anything fresh or new)				
Overall impression as a piece of work				

Project: where a student undertakes a piece of work unique to them and writes up a report summarising what they have studied.

Dissertation: where a student undertakes a long period of work on their own, perhaps involving considerable reading or perhaps involving experimental work. The bring together their findings in a report.

Duly Performed

Look back to page 88 to see a description of this approach. It is incredibly valuable and effective as a way of assessing important but intractable skills. Instead of assessing the product, we assess the completion of a task or set of tasks.

Conclusions

Here is a suggested way to think about choice of test format:

Testing				
What they know	***What they understand***	***What thy can do***	***How well they think***	***How well they evaluate***
Multiple Choice Partial knowledge multiple choice Short answer	Structural communication grids Essay Projects or reports Verbal presentation	Projects Dissertations Duly performed Observation Verbal presentation	Duly performed Dissertation Essay	Essay Project Dissertation Duly performed
Remember: Never depend on one test or examination Use multiple short assessments during and at the end of a course Use many different formats in any assessment programme Always plan assessments with a clear specification of what you want to assess Avoid testing just recall - its is an unimportant skill in the internet age Marks have no absolute meaning at all				

Assessment Moving *into* a New Future

Handing Test Data

Aims

This chapter looks to show the meaning of examination marks and how marks can be handled easily to give us more useful information. This may involve spreadsheets and their use is described.

Introduction

All the examination scores data you have collected on the performance of students will have to be analysed. In this chapter we will focus on the analysis and interpretation of the data you have collected about the knowledge, skills and attitudes of your students.

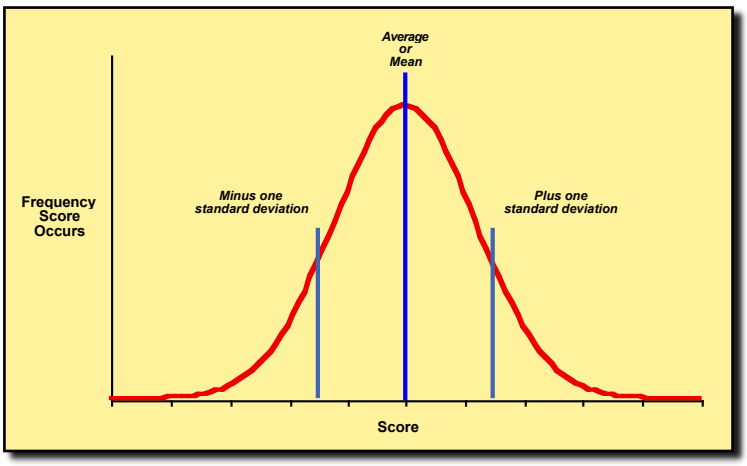
You analyse and interpret the information you have collected about your students quantitatively and qualitatively. For the quantitative analysis of data, various statistical tools are used. For example, statistics are used to show the distribution of scores on a Biology test and the average score obtained by students.

After a diet of examinations set by the Examination Board, there is a need to collect and analyse the data. Some will need to be shared with students, their parents, administrators, government authorities, potential employers and the wider community. The level of detail included in the result report will depend on the demands of the audience and whether it will be use effectively.

For some people, the word '*statistics*' is often associated with numbers while others are of the opinions that one has to be good in mathematics to understand and use statistics. Both these perceptions are not altogether accurate. Statistics is much more than numbers. Indeed, it does involve the assembling, classifying and tabulating of numbers but more important is how we analyse data for purposes of making decisions and generalisations. In fact, many of us are using statistics without being aware of it. For example, when we talk about average mathematics score, per capita income of parents in the district, the percentage of students who scored 'A', we are talking the language of statistics.

Simple Statistical Ideas

If we set an test or examination and large numbers sit it, then the marks will always tend to form a distribution as shown below. Most candidate will gain marks near the average and fewer will gain very high or very low marks.



The average mark is known as the *mean* and the extent to which the marks are spread out is given by what is called the *standard deviation*. In a subject like mathematics, marks may be well spread out while in a subject like English, the marks will probably be much more bunched near the middle. This simply reflects the nature of these two subjects and the way tests and examinations can be marked.

A mark of 50% has absolutely no meaning. In a difficult examination, it might be a very good mark while, in an easy examination, the candidate scoring 50% may well be near the bottom of the range of marks. It is an interesting thought that a 50% simply means that the candidate managed to score half the possible marks in that selection of questions on that day. Think of an air pilot or a surgeon. When they sat their examinations, is 50% enough?

When the marks for a subject are obtained, any examination board will obtain a pattern like that shown in the picture above. The real question is where we make the pass mark and where we divide the 'A's from the 'B's etc. This is a value judgement and is taken on the basis of experience and the knowledge that the ability of the population sitting the examination is likely to be very similar from year to year. Thus, if the population sitting is regarded as similar, then the same proportion will tend to obtain the various grades of passes from year to year.

If we use percentages, then a mark of 70% in English may well be worth much more than a 70% in mathematics simply because the candidate is near the top in English but may be well off the top in mathematics. Thus, a mark in the English paper is worth more than a mark in the mathematics paper.

If marks are to be added, then we must ensure that that mark on one paper is worth approximately the same as mark in the other. This is easily achieved using standardisation. In standardisation, we are making the means the same and the spread of marks the same. It does not alter the order of merit in any subject paper. Thus, a mark in one paper is worth the same as a mark in the other. Standardisation can be carried out easily using any standard spreadsheet. There is another advantage of standardisation. It means that the learners can compare their performance across various subjects more easily, giving more valid information to the candidates and to their parents.

For the mathematically inclined, the formula for standardisation can be illustrated:

Imagine that we have set an examination paper out of 100. We have a group of, say, 50 learners. If we place all the marks in a column on a spreadsheet, then the spreadsheet will calculate the mean and the standard deviation (almost all spreadsheets have this facility built in). Suppose we find that we have a mean of 53% and standard deviation of, say, 7%.

Suppose we want to make our marks have a mean (average) of, say, 60% and a standard deviation of, say, 12%. This spreads our marks in such a way that the vast majority lie between 36% and 84%. This is a reasonable spread. The formula is:

$$\left[\left(\frac{\text{Student Mark} - \text{Mean}}{\text{Standard Deviation}} \right) \text{X } 12 \right] + 60$$

Thus, for a candidate who obtained 67% originally, we take away 53, to give 14. We divide 14 by 7 to obtain 2. This is multiplied by 12 to give 24 and 60 is added on, giving a standardised score of 84%.

We could, of course, have used any mean and standard deviation we liked. Here we simply *chose* 60 and 12 as this gives a convenient mark range.

Key Issues - Human Judgement

The way the teachers grade students and their work is always a contentious topic. It is directly relevant to all instructors and teachers, especially those who are required to grade within a formal system.

Grading is the most formal part of the process of student assessment. It is associated with the judgments or decisions that you make on the basis of the assessment results:

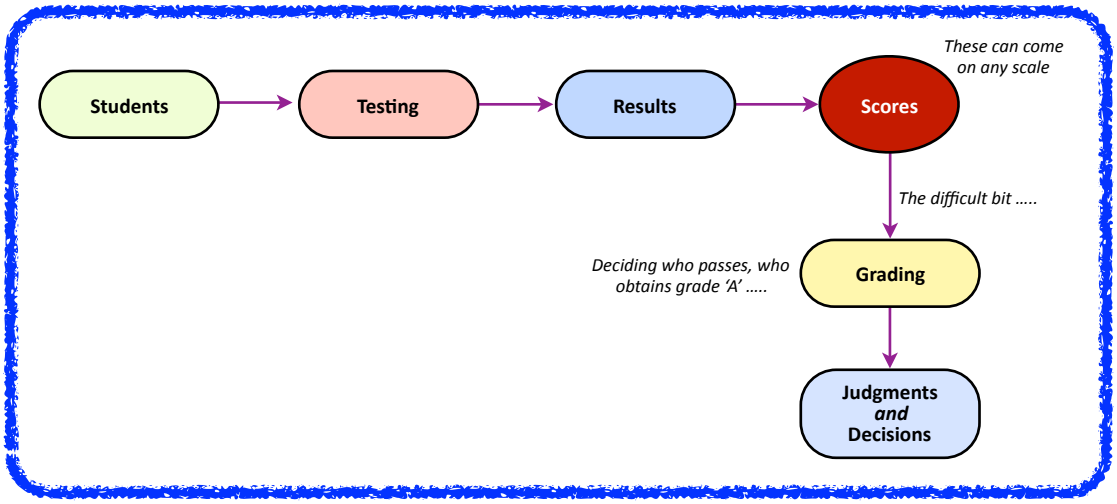


Figure 13.1 Student Assessment Process

Assessment results are the basis of grading. They come to you as marks or the original test scores (also known as raw scores).

Although scores are the marks obtained by a student in a test, they cannot be interpreted as units of knowledge. We attach marks to each question but the sum of these numbers depends entirely on the difficulty of the test questions and the ability of the students. At best, the scores give you only a general sense of the direction of a person's achievements. In addition, there are some important limitations in the traditional use of scores in testing:

Examinations scores or grades

- **Place candidates in a rough order of merit**

They do NOT tell us:

- **What the candidate can actually do**
- **The quality of their learning**
- **Anything absolute**
- **How well the candidate will perform in the future**

It is important that we recognise the weaknesses inherent in any examination and that we do not ask test scores to do too much and to do things for which they were not really designed. Sadly, in wider society, there is a naive belief in the value of all test scores, their exact validity and accuracy. There is an important task to be carried out by Examination Boards in educating the wider public, including parents, media and politicians.

Key Issues - Meaning of marks

The most important point is that marks carry no absolute meaning at all. In other words, if you gain 90% in an examination, that does not necessarily mean that you have done very well. If most of the students taking the examination are obtaining 90% or more, then you may indeed have done badly!

All that an examination does is to place the candidates in an approximate order of merit

It is important to remember that the order is only approximate and that the marks obtained are only an estimate of performance, reflecting how a particular student performed on one occasion in one examination.

Key Issues - Frequency of assessments

There are two important principles:

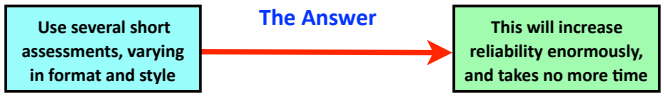
Principle 1

Imagine you are teaching a course or module. There is real danger in only have one final assessment at the end of the course or module. This will be an *unreliable* assessment of student performance. This is because no examination is free from flaws, students can have good days and bad days, and the examination paper might suit some students more than others. It is MUCH better to use several assessments: perhaps, an essay or report to be handed in during the course, a mid-course assessment, a project to be completed, a verbal presentation to be given, plus a final examination. All count towards the final overall grade. This is much more reliable.

Principle 2

You cannot have too many assessments simply because it will take you too much time and, more importantly, it will take too much time for your students.

This leads to the answer:



One *'test'* might be a short report, maybe a summary of some literature or a section of a textbook (done as '*homework*'. Two *'tests'*, at different stages of the course, might each involve ten multiple choice questions (takes 10 minutes each). One *'test'* might be an essay on a theme (maybe 30 minutes of class time). One *'test'* might involve 2 structural communication grids and a few short-answer questions (taken half way through course) - maybe lasting 45 minutes. The final *'test'* might have a mixture of formats involve short answer questions and longer questions of an essay type - this might last 90 minutes. The overall course assessment is found by combining all these outcomes. However, combining test outcomes is not quite so easy and we discuss this next.

Collation of several assessments

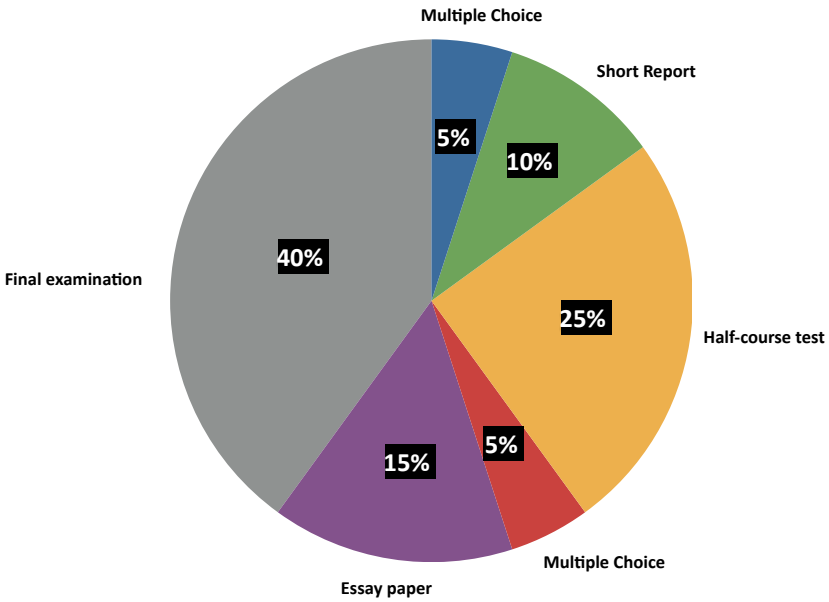
Many teachers convert scores or marks to percentages and then add the results from different tests to gain the final score. However, not all the tests might be of equal value or importance and not all the test will be of equal difficulty.

Let us look at the set of assessments we have just suggested for a course you are teaching. They are now set out in a table:

Your decision

Test	Time	Description	Total Mark	Credit
Short multiple choice test (10 items)	10 minutes	Covering a short section of work	10	5%
Short Report	Done as 'homework'	Extraction of information from internet and books, to summarise some theme	25	10%
Half-course test	45 minutes	2 structural communication grids, plus some short answer questions	40	25%
Short multiple choice test (10 items)	10 minutes	Covering a short section of work	10	5%
Essay paper	30 minutes	Two short essays on two topics	30	15%
Final examination	90 minutes	Short answer and longer answer questions	60	40%
				100%

This is offered simply as an example to show how to handle the test data.



The outcomes for each student are simply typed into a spreadsheet which then does all the calculations for us. Here is the what the top of the spreadsheet might look like:

	A	B	C	D	E	F	G	H
1	Candidate	Multiple Choice 1	Short Report	Half-way Test	Multiple Choice 2	Essay Test	Final Exam	Total
2		10	25	40	10	30	60	100
3	1	6	18	21	7	16	34	57
4	2	7	19	18	8	17	28	54
5	3	4	14	15	3	12	31	45
6	4	7	22	29	9	22	45	76
7	5	6	16	31	8	25	41	73
8	6	9	21	31	9	22	51	82
9	7	5	15	27	5	19	48	69
10	8							
11	9							

The spreadsheet needs programmed to calculate the total marks but this is easy. The formula used here is:

$$=(B3/2) + (C3/2.5) + (D3*25/40) + (E3/2) + (F3/2) + (G3*4/6)$$

This is for the first student and then it applies to all the students. If you are not sure, then ask someone who has experience with spreadsheets. The use of spreadsheet is outline later in this chapter. Once set up, you can use it year after year. All that have to do is type in the actual marks - the spreadsheet does it all for you! The next question is what you do with the total marks and this is now discussed.

Marks or Grades ?

Look at the total marks in the spreadsheet above. They range from 45% to 82%. Where do we make the passmark? After talking to colleagues, you might decide that 55% is the passmark and that above 65% gives a 'B' pass, with above 80% giving an 'A' pass. The spreadsheet can be completed now:

	A	B	C	D	E	F	G	H	I
1	Candidate	Multiple Choice 1	Short Report	Half-way Test	Multiple Choice 2	Essay Test	Final Exam	Total	Grade
2		10	25	40	10	30	60	100	
3	1	6	18	21	7	16	34	57	C
4	2	7	19	18	8	17	28	54	Fail
5	3	4	14	15	3	12	31	45	Fail
6	4	7	22	29	9	22	45	76	B
7	5	6	16	31	8	25	41	73	C
8	6	9	21	31	9	22	51	82	A
9	7	5	15	27	5	19	48	69	B
10	8								
11	9								

Many teachers indicate that they find the assigning of grades to be one of their most difficult tasks. It is a difficult task to be both an instructor as well as an assessor. It requires a combination of objectivity as well as the ability in developing trust between teacher and students. These are demanding skills to achieve but the best teachers do them well and can generate a very level of integrity as well as having a great empathy for their learners.

Predictive Value of Marks?

Society holds examination marks and grades with great respect, especially final national examinations. The evidence shows two main things very clearly:

- These marks or grades are only *estimates* of student abilities on *one* day in *one* examination.
- They are very poor predictors of long-term academic success.

Indeed, it is foolishness to think that we can summarise a whole semester's learning in a single letter of the alphabet or a single mark. Learning is much more complex than that!

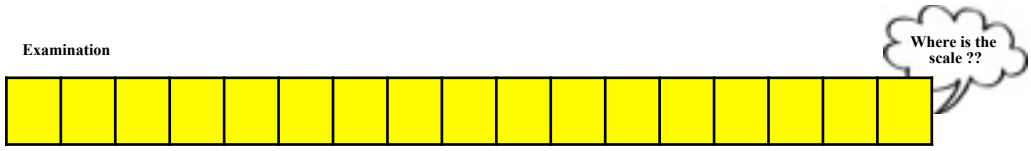
In other words, assessment is a very difficult task and, with current knowledge, we are not very good at it. Therefore, we need to treat assessment results (marks and grades) with great caution. They do not give a very accurate picture of student performance and they do not relate very well to longer terms success. In simple terms, the quality of degree from a university does not relate too exactly with the grades set student achieved at school. This is caused by two factors:

(1) Examination marks and grades are intrinsically not very accurate.

(2) Students change and develop with time and this may bring about considerable deterioration or improvement in academic performance.

Setting pass marks and grade levels

The accuracy of examinations was discussed in chapter 11 (see page 104). No matter how well set and marked, they are very imperfect measuring instruments. It is like trying to measure your height with a measuring tape on which no number scale has been marked. If we have a group of people, then we can place them in rough order by their height. That is all.



Examinations are like that. They are quite good (but by no means perfect) in placing learners in some kind of order of merit. However, where to place the '*pass mark*' or the grade boundaries is a matter of human judgement.

Therefore, statements in the media which suggest that standards have improved or otherwise on the basis of grades in some national examination are simply nonsense. The examination papers are not identical from year to year. We now have more than one measuring tape and none of them has a number scale marked on it!

Using Spreadsheets in Assessment

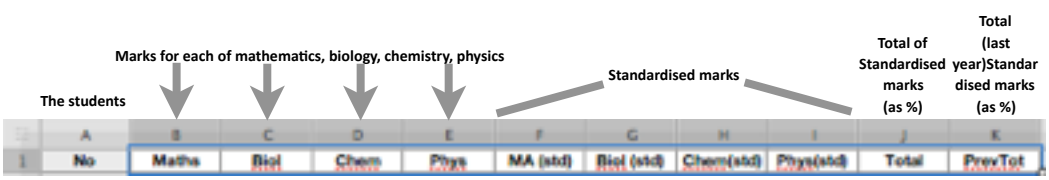
The spreadsheet can make our lives very much easier when handing assessment data. A spreadsheet is a two dimensional array of boxes. Every box is labelled by its column position (a letter) and its row position (a number). In the spreadsheet below, the box selected is box E6.

Suppose you want to analyse the examination results for several subjects. You have obtained marks (as percentages) for four subjects (Mathematics, Biology, Chemistry and Physics) for, say 50 students.

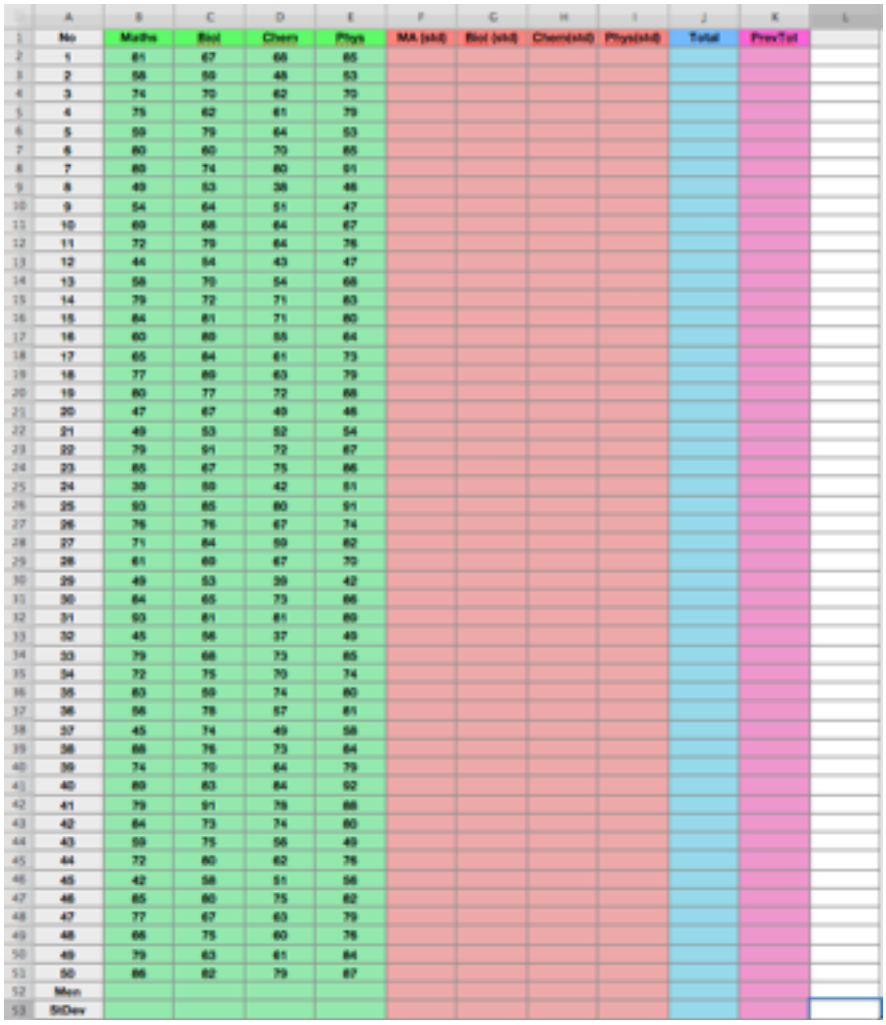
You set up spreadsheet like the one shown here. The ONLY data you enter is in columns B,C,D and E and the total mark for these students for their previous year (column K).

	A	B	C	D	E	F	G	H	I	J	K	L
1	No	Maths	Biol	Chem	Phys	MA (std)	Biol (std)	Chem(std)	Phys(std)	Total	PrevTot	
2	1											
3	2											
4	3											
5	4											
6	5											
7	6											
8	7											
9	8											
10	9											
11	10											
12	11											
13	12											
14	13											
15	14											
16	15											
17	16											
18	17											
19	18											
20	19											
21	20											
22	21											
23	22											
24	23											
25	24											
26	25											
27	26											
28	27											
29	28											
30	29											
31	30											
32	31											
33	32											
34	33											
35	34											
36	35											
37	36											
38	37											
39	38											
40	39											
41	40											
42	41											
43	42											
44	43											
45	44											
46	45											
47	46											
48	47											
49	48											
50	49											
51	50											
52	Men											
53	StDev											

The meaning of all the columns is shown below:



Now let us imagine you have entered the marks. The spreadsheet now looks like this, with colour added to make it easier to follow down the columns.



	A	B	C	D	E	F	G	H	I	J	K	L
1	No	Maths	Biol	Chem	Phys	MA (std)	Biol (std)	Chem(std)	Phys(std)	Total	PrevTot	
2	1	81	67	68	85							
3	2	58	59	48	53							
4	3	74	70	62	70							
5	4	75	62	61	79							
6	5	59	79	64	53							
7	6	80	60	70	85							
8	7	89	74	80	91							
9	8	49	53	38	46							
10	9	54	64	51	47							
11	10	69	68	64	67							
12	11	72	79	64	76							
13	12	44	54	43	47							
14	13	58	70	54	68							
15	14	79	72	71	83							
16	15	84	81	71	80							
17	16	60	69	55	64							
18	17	65	84	61	73							
19	18	77	80	63	79							
20	19	80	77	72	88							
21	20	47	67	49	46							
22	21	49	53	52	54							
23	22	79	91	72	87							
24	23	85	67	75	86							
25	24	39	59	42	51							
26	25	93	85	80	91							
27	26	76	76	67	74							
28	27	71	84	59	82							
29	28	61	69	67	70							
30	29	49	53	39	42							
31	30	84	65	73	86							
32	31	93	81	81	89							
33	32	45	56	37	49							
34	33	79	68	73	85							
35	34	72	75	70	74							
36	35	83	59	74	80							
37	36	56	78	57	61							
38	37	45	74	49	58							
39	38	88	76	73	84							
40	39	74	70	64	79							
41	40	89	83	84	92							
42	41	79	91	78	88							
43	42	84	73	74	80							
44	43	59	75	56	49							
45	44	72	80	62	78							
46	45	42	58	51	56							
47	46	85	80	75	82							
48	47	77	67	63	79							
49	48	66	75	60	76							
50	49	79	63	61	84							
51	50	86	82	79	87							
52	Mean											
53	StDev											

The spreadsheet will carry out all the calculations for you, provided you tell it what you want it to do.

Look at the foot of the spreadsheet, boxes A52 and A53.

These are labelled: Mean; StDev. The mean is simply the average of all the marks in each column while the standard deviation (StDev) tells us how well the marks have been spread out. To calculate the mean of column B, type the following in box B52:

= AVERAGE(B2:B51)

The '*equals sign*' tells the spreadsheet you want it to do a calculation. The word 'AVERAGE' tells the spreadsheet to find the average. In the brackets is B2:B51. You are instructing the spreadsheet to find the average over that range of boxes.

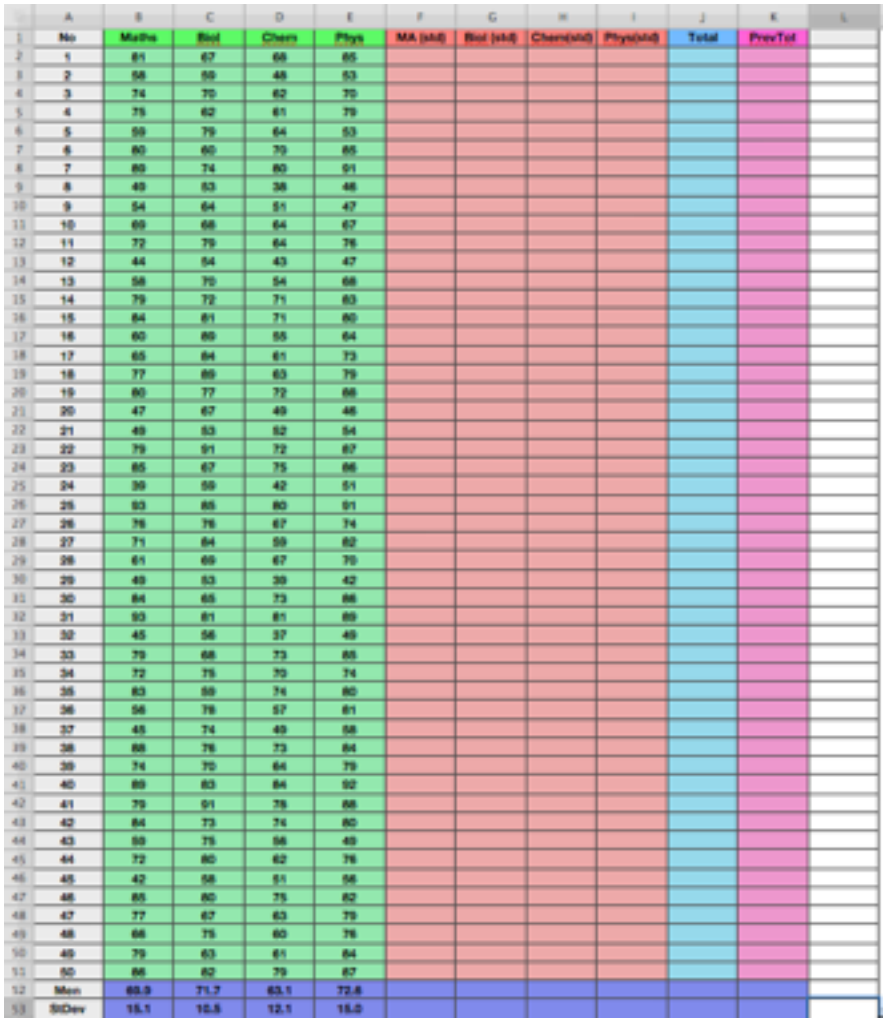
For box B53, the formula is:

= STDEV (B2:B51)

This calculates the standard deviation of the marks in boxes B2 to B51.

The clever things is that we do not need to re-type it into boxes C52 and C53 and all the boxes in line 52 and 53.

We select boxes B52 and B53. You will notice a small dot at the lower right hand corner. Place the cursor over this and pull gently to the right to cover boxes C52, C53 across to E52 and E53. The calculations are done automatically for you. This is what it looks like.



	A	B	C	D	E	F	G	H	I	J	K	L
1	No	Maths	Biol	Chem	Phys	MA (std)	Biol (std)	Chem(std)	Phys(std)	Total	PrevTot	
2	1	81	67	68	65							
3	2	58	59	48	53							
4	3	74	70	62	70							
5	4	75	62	61	79							
6	5	59	79	64	53							
7	6	80	60	70	85							
8	7	89	74	80	91							
9	8	49	53	38	48							
10	9	54	64	51	47							
11	10	69	68	64	67							
12	11	72	79	64	76							
13	12	44	54	43	47							
14	13	58	70	54	68							
15	14	79	72	71	83							
16	15	84	81	71	80							
17	16	60	69	55	64							
18	17	65	84	61	73							
19	18	77	89	63	79							
20	19	80	77	72	88							
21	20	47	67	49	46							
22	21	49	53	52	54							
23	22	79	91	72	87							
24	23	65	67	75	86							
25	24	39	59	42	51							
26	25	93	85	80	91							
27	26	76	76	67	74							
28	27	71	84	59	82							
29	28	61	69	67	70							
30	29	49	53	39	42							
31	30	84	65	73	86							
32	31	93	81	81	89							
33	32	45	56	37	49							
34	33	79	68	73	85							
35	34	72	75	70	74							
36	35	83	89	74	80							
37	36	56	78	57	81							
38	37	45	74	49	58							
39	38	88	76	73	84							
40	39	74	70	64	79							
41	40	89	83	84	92							
42	41	79	91	75	88							
43	42	84	73	74	80							
44	43	59	75	56	49							
45	44	72	80	62	76							
46	45	42	58	51	56							
47	46	85	80	75	82							
48	47	77	67	63	79							
49	48	66	75	60	76							
50	49	79	63	61	64							
51	50	86	82	79	87							
52	Mean	68.9	71.7	63.1	72.6							
53	StDev	15.1	10.5	12.1	15.0							

Now you can see that the means (averages) are not the same for the four subjects. Clearly, the chemistry paper was more difficult than those for biology or physics while the mathematics paper was the easiest. That means that each mark in chemistry is worth more than the marks in other subjects.

Standardisation of Marks

This is a complex word that simply means that we put all the marks for the four subjects onto the same scale. This means we can add them up and compare across subjects.

We have to decide what scale we shall use. A convenient one is to have a mean (average) of 60 and a standard deviation of 12. This places the marks on a convenient scale.

Now look at box F2. To standardise the mathematics marks, we place a set of instructions in this box:

$$= \quad (((B2-B\$52)/B\$53) * 12) + 60$$

Do not worry too much about this formula. When this is typed in and we press the '*return*' button, we find the value of 69 appears in box F2.

We select box F2 and place the cursor over the dot in the lower right-hand corner. Pull it across to I2. The standard marks appear for this student in all four subject.

The clever thing now is to select boxes F2, G2, H2 and I2. Place the cursor over the lower right-hand corner and pull the boxes down to I51. The standard scores for all subjects for all students appear, just as if by magic! This is the incredible power of spreadsheets. They save hours of labour.

You can now pull across from boxes E52 and E53 to obtain the means of the standard scores. They are all showing a mean of 60 and standard deviation of 12, just as we set them. This checks we have it right.

This is what you get:

	A	B	C	D	E	F	G	H	I	J	K	L
1	No	Maths	Biol	Chem	Phys	MA (std)	Biol (std)	Chem(std)	Phys(std)	Total	PrevTot	
2	1	81	67	68	85	69	55	66	70			
3	2	58	59	48	53	51	46	45	44			
4	3	74	70	62	70	63	58	59	58			
5	4	75	62	61	79	64	49	58	65			
6	5	59	79	64	53	51	68	61	44			
7	6	80	60	70	85	68	47	67	70			
8	7	89	74	80	91	75	63	77	75			
9	8	49	53	38	46	43	39	35	39			
10	9	54	64	51	47	47	51	48	39			
11	10	69	68	64	67	59	56	61	55			
12	11	72	79	64	76	62	68	61	63			
13	12	44	54	43	47	39	40	40	39			
14	13	58	70	54	68	51	58	51	56			
15	14	79	72	71	83	67	60	68	68			
16	15	84	81	71	80	71	71	68	66			
17	16	60	89	55	64	52	80	52	53			
18	17	65	84	61	73	56	74	58	60			
19	18	77	89	63	79	66	80	60	65			
20	19	80	77	72	88	68	66	69	72			
21	20	47	67	49	46	42	55	46	39			
22	21	49	53	52	54	43	39	49	45			
23	22	79	91	72	87	67	82	69	72			
24	23	85	67	75	86	72	55	72	71			
25	24	39	59	42	51	36	46	39	43			
26	25	93	85	80	91	78	75	77	75			
27	26	76	76	67	74	65	65	64	61			
28	27	71	84	59	82	61	74	56	68			
29	28	61	69	67	70	53	57	64	58			
30	29	49	53	39	42	43	39	36	35			
31	30	84	65	73	86	71	52	70	71			
32	31	93	81	81	89	78	71	78	73			
33	32	45	56	37	49	40	42	34	41			
34	33	79	66	73	85	67	56	70	70			
35	34	72	75	70	74	62	64	67	61			
36	35	83	59	74	80	70	46	71	66			
37	36	56	78	57	81	49	67	54	67			
38	37	45	74	49	56	40	63	46	48			
39	38	66	76	73	84	74	65	70	69			
40	39	74	70	64	79	63	58	61	65			
41	40	80	83	84	92	76	73	81	76			
42	41	79	91	76	86	67	82	75	72			
43	42	84	73	74	80	71	62	71	66			
44	43	59	75	56	49	51	64	53	41			
45	44	72	80	62	76	62	69	59	63			
46	45	42	56	51	56	38	44	48	47			
47	46	85	80	75	82	72	69	72	68			
48	47	77	67	63	79	66	55	60	65			
49	48	66	75	60	76	57	64	57	63			
50	49	79	63	61	84	67	50	58	69			
51	50	86	82	79	87	73	72	76	72			
52	Mean	60.9	71.7	63.1	72.6	60.0	60.0	60.0	60.0			
53	StDev	15.1	10.5	12.1	15.0	12.0	12.0	12.0	12.0			

Now all examinations marks are on the same scale. Therefore, we can add them up.

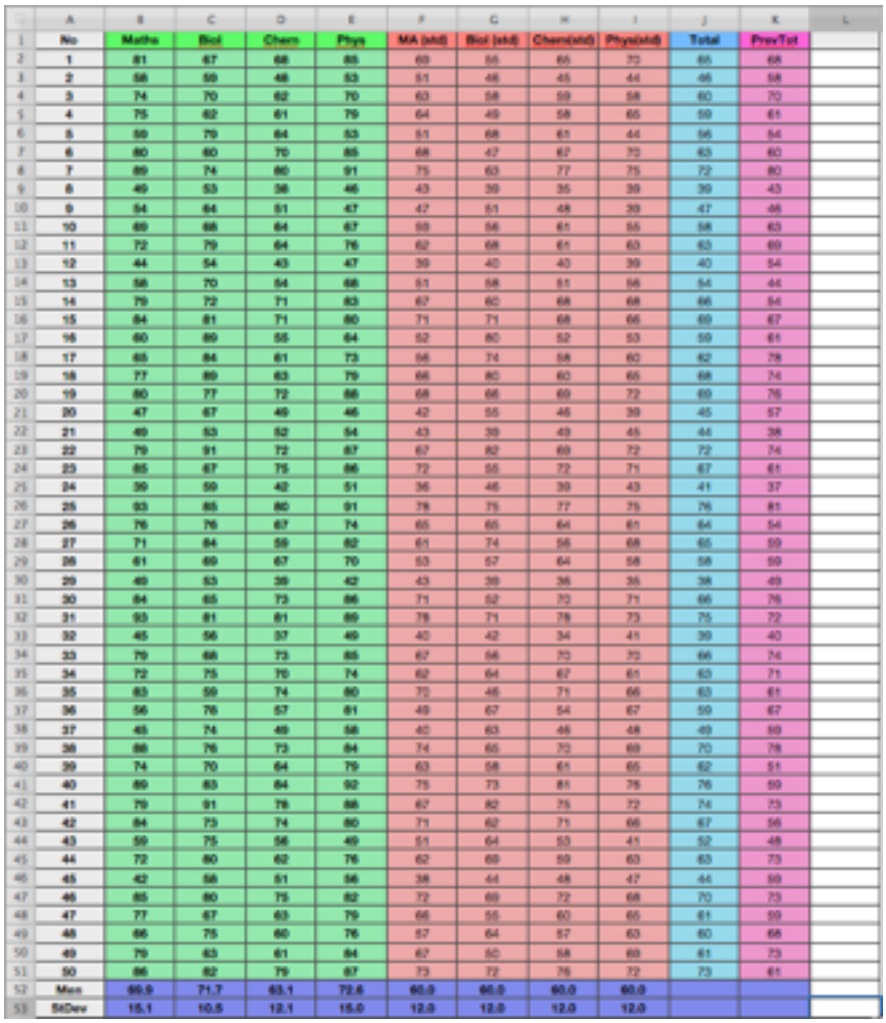
It is important to note that standardising marks does NOT change the order of marks at all. What it does is put the marks for all subjects on to the same scale (allowing for some examinations giving lower marks than others because the examination paper happen to be more difficult).

The final step is to complete column J by adding them up and then dividing by 4 (to get an average percentage).

At the same time, the average marks for the previous year are typed in to column K. We can now see which students have improved and which have been less successful. By using standard marks, we are comparing like with like and comparisons are valid.

The next page shows the final spreadsheet.

Completed Spreadsheet



	A	B	C	D	E	F	G	H	I	J	K	L
1	No	Maths	Biol	Chem	Phys	MA (std)	Biol (std)	Chem(std)	Phys(std)	Total	PrevTot	
2	1	81	67	68	85	69	55	65	72	65	68	
3	2	56	59	48	53	51	46	45	44	46	58	
4	3	74	70	62	70	63	58	59	58	60	70	
5	4	75	62	61	79	64	49	58	65	59	61	
6	5	59	79	64	53	51	68	61	44	56	54	
7	6	80	60	70	85	68	47	67	70	63	60	
8	7	89	74	80	91	75	63	77	75	72	80	
9	8	49	53	38	46	43	39	35	39	39	43	
10	9	54	64	51	47	47	51	48	39	47	46	
11	10	69	68	64	67	59	56	61	55	58	63	
12	11	72	79	64	76	62	68	61	63	63	69	
13	12	44	54	43	47	39	40	40	39	40	54	
14	13	58	70	54	68	51	58	51	56	54	44	
15	14	79	72	71	83	67	60	68	68	66	54	
16	15	84	81	71	80	71	71	68	66	69	67	
17	16	60	69	55	64	52	60	52	53	59	61	
18	17	65	84	61	73	56	74	58	60	62	78	
19	18	77	69	63	79	66	60	60	65	68	74	
20	19	80	77	72	86	68	66	69	72	69	76	
21	20	47	67	49	46	42	55	46	39	45	57	
22	21	49	53	52	54	43	39	49	45	44	38	
23	22	79	91	72	87	67	82	69	72	72	74	
24	23	85	67	75	86	72	55	72	71	67	61	
25	24	39	59	42	51	36	46	39	43	41	37	
26	25	93	85	80	91	79	75	77	75	76	81	
27	26	76	76	67	74	65	65	64	61	64	54	
28	27	71	84	59	82	61	74	56	68	65	59	
29	28	61	69	67	70	53	57	64	58	58	59	
30	29	49	53	39	42	43	39	36	35	38	49	
31	30	84	65	73	86	71	52	70	71	66	76	
32	31	93	81	81	89	78	71	78	73	75	72	
33	32	45	56	37	49	40	42	34	41	39	40	
34	33	79	68	73	85	67	56	70	70	66	74	
35	34	72	75	70	74	62	64	67	61	63	71	
36	35	83	59	74	80	70	46	71	66	63	61	
37	36	56	76	57	81	49	67	54	67	59	67	
38	37	45	74	49	58	40	63	46	48	49	59	
39	38	86	76	73	84	74	65	70	69	70	78	
40	39	74	70	64	79	63	58	61	65	62	51	
41	40	89	83	84	92	75	73	81	76	76	69	
42	41	79	91	78	88	67	82	75	72	74	73	
43	42	84	73	74	80	71	62	71	66	67	56	
44	43	59	75	56	49	51	64	53	41	52	48	
45	44	72	80	62	76	62	69	59	63	63	73	
46	45	42	58	51	56	38	44	48	47	44	59	
47	46	85	80	75	82	72	69	72	68	70	73	
48	47	77	67	63	79	66	55	60	65	61	59	
49	48	66	75	60	76	57	64	57	63	60	68	
50	49	79	63	61	84	67	50	58	69	61	73	
51	50	86	82	79	87	73	72	76	72	73	61	
52	Mean	69.9	71.7	63.1	72.6	60.0	60.0	60.0	60.0			
53	StDev	15.1	10.5	12.1	15.0	12.0	12.0	12.0	12.0			

We can now see precisely which students have improved and which have deteriorated. This enables us to offer encouragement and help as needed.

Moving into a New Future

Aims

This chapter brings together the main ideas covered in this monograph, suggests key initiatives that need to be taken and points to the kinds of research that still needs to be carried out.

Introduction

The educational system is facing substantial change in a rapidly evolving global context. This monograph draws on the research literature to identify current possible ways forward so that assessment can address immediate and future demands.

It has to be recognised that assessment more or less defines what goes on in schools and university classes. It controls what is taught and learned. It controls how material is taught and how learners seek to master it. It has to be remembered that school teachers do ***not*** control the curriculum or the resources available to them. More importantly, they do not control the assessment used at provincial and national levels. If teachers are to be released to move education forward, then the stranglehold of assessment needs to be re-thought completely. This will involve everyone involved in education at all levels working together.

This monograph has explored how educational institutions might use assessment more effectively to promote student learning. At a world level, there have been massive curriculum developments in many countries along with exciting developments in teaching and learning. However, assessment has lagged behind. It is now time for Pakistan to allow every level of assessment to be renewed and re-thought and brought up to the best standards of the world.

In this monograph, we have attempted to bring together the main findings from research on learning and assessment in order to develop helpful guidelines. This chapter seeks to summarise the overall findings and to suggest some key issues for the future. The chapters that then follow will outline the practical skills needed to implement the findings from research evidence in order to equip setters, markers and examiners to fulfil the aspirations in achieving the best possible examinations system.

The aim is assessment literacy:

Figure 14.1 Assessment literacy

An Examination Board has great responsibilities in developing assessment literacy.

- It can set exemplary high standards in the quality of questions used and the way the examinations are carried out.
- It can offer training and support for school teachers.
- It can offer guidance and insight for politicians and the media.
- It can develop communications for parents and employers, explaining in straightforward terms what quality assessment is all about.

Assessment

Of course, we need to be clear about what assessment really involves:

Process	Purpose	Direct beneficiary	Function	Focus
Assessment is the process of gathering, interpreting and using evidence to make judgements about the achievements of students in learning	Assessment should enhance learning	It should inform the learner (student) It should inform the Instructor (teacher)	It can direct support future learning. It is important for purpose of marking, grading, certification	Primarily on the individual learner but also on groups of learners

Table 14.1 Assessment Purposes

Assessment can be used in many ways and by many people.

For example,

- Assessment can be used as a vehicle for giving some feedback on teaching or learning;
- Assessment can inform us how well our student are doing;
- Assessment may be needed to award a certificate or grade on a course.
- Assessment may simply be needed by educational administration.

There are two key words that are important: quantity and quality:

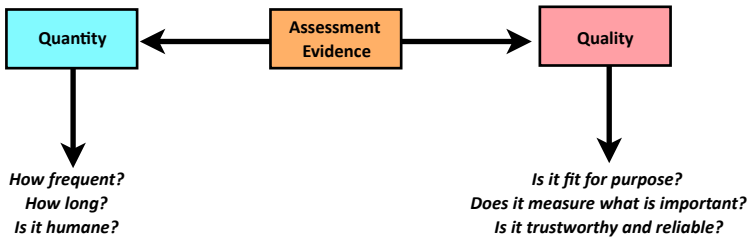


Figure 14.2 Quality and Quantity

One of the biggest problems with assessment data is that it can be used for purposes for which it was never intended. If assessment which has been designed for final grading and certification is used to evaluate the quality of teachers, schools or even countries, then the conclusions drawn will almost certainly be invalid. We simply cannot assess the quality of education by looking at examination outcomes and all international comparison are essentially invalid because there is no way to allow for the diversities in education provision and cultures in different countries.

We need to think about who will use assessment data. There are several main use groups:

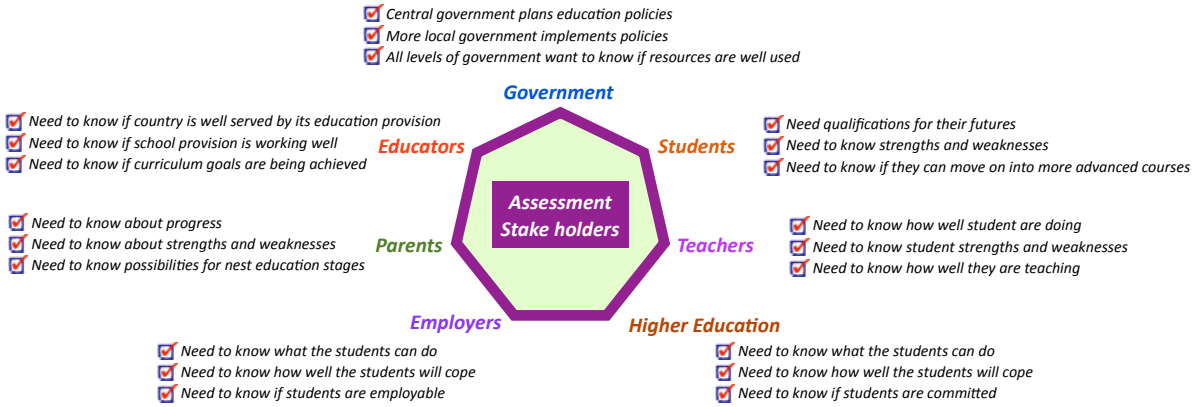


Figure 14.3 Assessment stake holders

Assessment and learning

Assessment is an integral part of all learning. *Students* need to know how well they are doing. *We* need to know out well our students doing. More importantly, *both our students and ourselves* need to know where there are strengths and weakness, so that future work can be directly effectively.

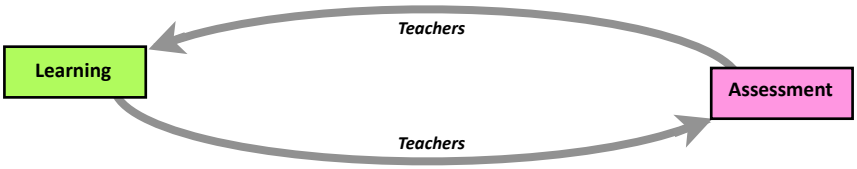


Figure 14.4 Assessment and Learning

However, assessment feeds back into every aspect of learning:

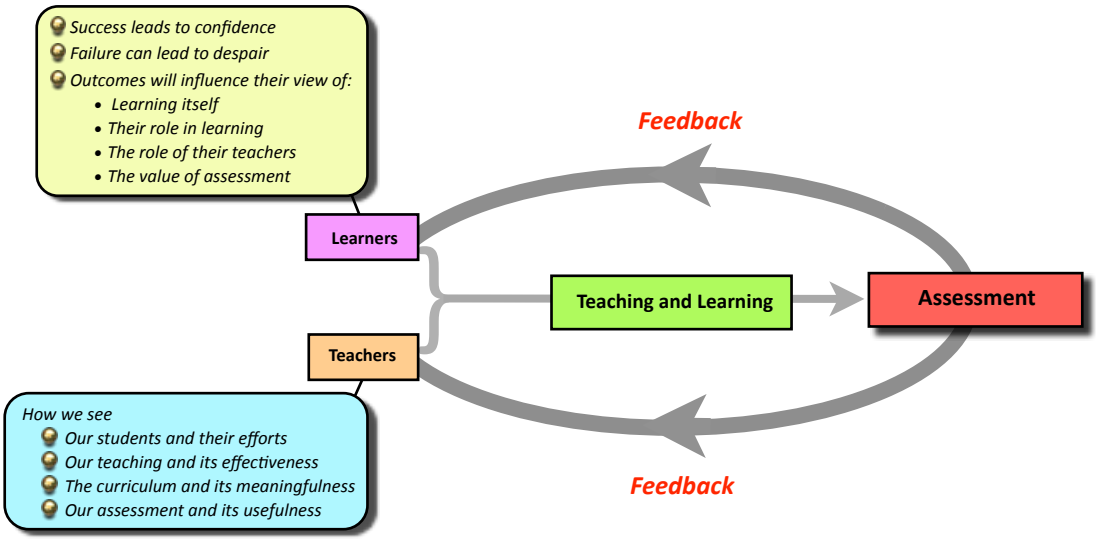


Figure 14.5 Assessment Feedback in Detail

Principles of Assessment

Here are five key principles. These follow each other in a sequence:

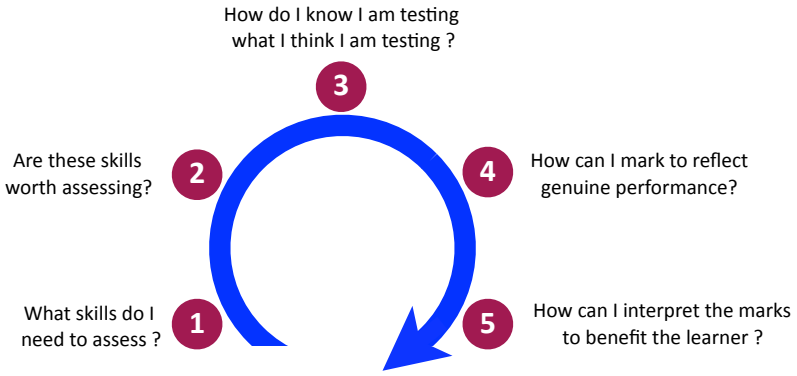


Figure 14.6 Principles of Assessment

What are we to Assess

This is the key question. So often today, we are simply testing the ability of our students to recall information or procedures. We need to widen our horizons considerably. Any the outset, based on the most recent analyses, five bread areas were suggested:

What they **know**	What the student knows (facts, concepts, skills) or can access
What they **understand**	Described in terms of the extent to which the student can apply their knowledge in novel situations with some prospect of success
What they can **do**	Skills (practical or procedural) which the student can demonstrate successfully
How they can **think**	The extent to which students can think creatively, critically or scientifically in relation to the material being studied
How well they can **evaluate**	The extent to which the student can ask the questions why? what? and how? of new information, its sources and the way it links to which is already known

This was then extended to bring in two other important skills: how well they can communicate and the extent to which they see their studies in relation to the world around. Bring this all together suggests seven main goals:

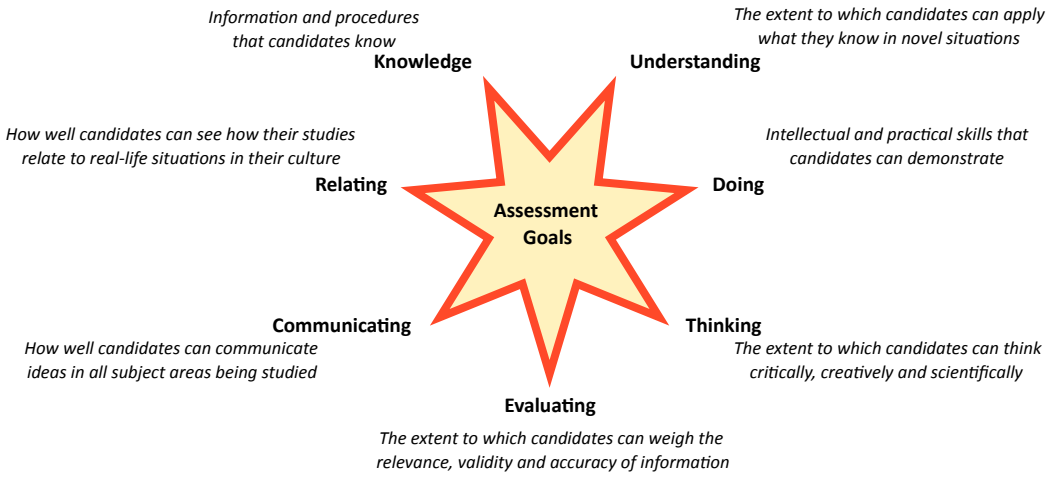


Figure 14.7 Seven Goals for Assessment

The aim of education has to include the development of all the skills and abilities of the students as well as encouraging them to learn how to apply their skills and understandings, with confidence and competence. The students need to be equipped to play a full part in their society, bringing benefit to that society because they can think critically and creatively but also have deep empathy for those around. This is a much greater picture than simply seeing education as imparting of information and examinations as a mechanism to measure how well that information has bene memorised. This can be seen in terms of the future lives of our students:

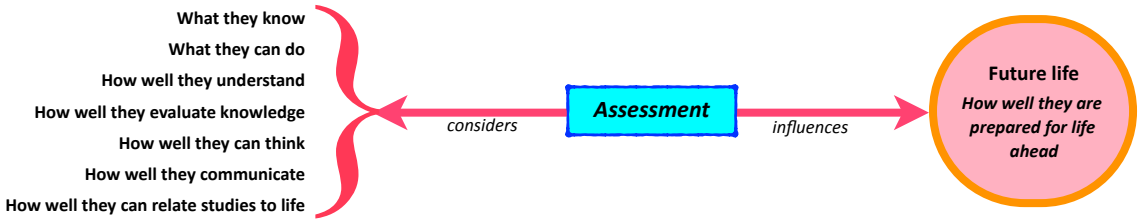


Figure 14.8 The role of assessment for life

Some Key Questions

Any assessment we use must be:

☑ Valid	Are we measuring what we think we are measuring?
☑ Reliable	Are we measuring consistently and accurately?
☑ Humane	Is the measurement procedure civilised, fair and kind?
☑ Economical	Can the measurement be carried out in reasonable time and at reasonable cost?
☑ Beneficial	Does the assessment bring benefit to learners and also to others?

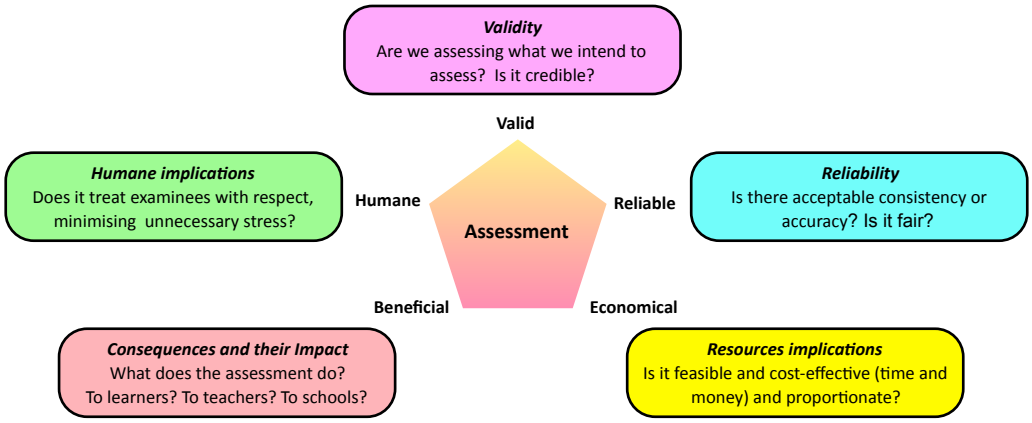


Figure 14.9 Key Features of Good Assessment

Process-Product Assessments

For too often assessment take place at the end of the course of study and reflects the '*product*', ignoring most of the developments on the educational journey.

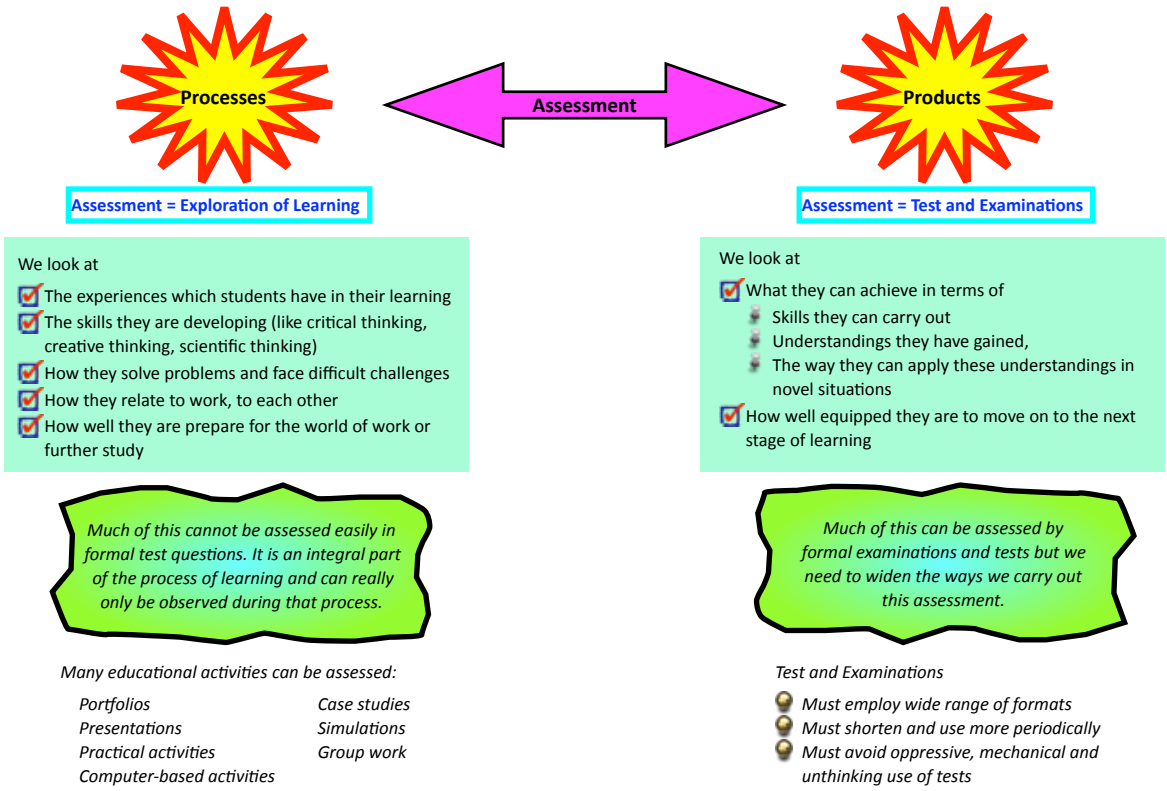


Figure 14.10 More about Process-Product Assessments

Questions of Quality

We want our assessments to be the best that they can be. Here are some questions to help:

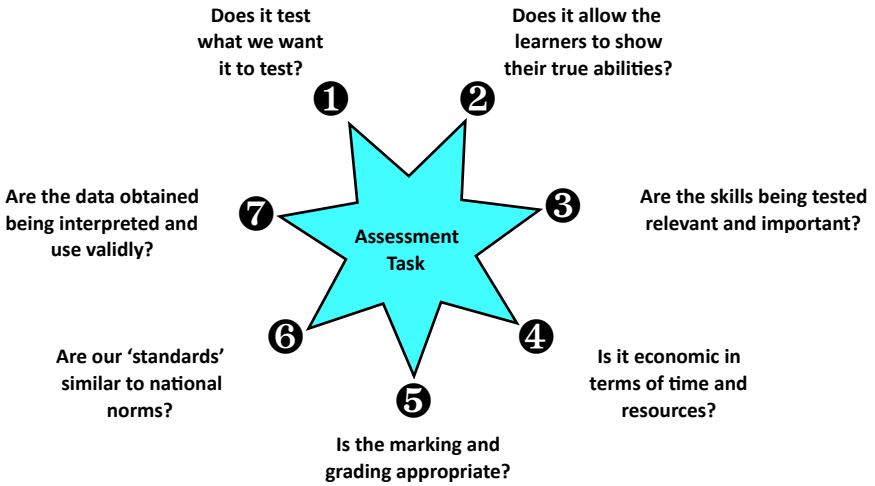


Figure 14.11 Assessment Task

One of the greatest problems we face is that assessments today tend only to reward the recall of correct information or correct proceeds. These have been memorised by the learners. Education is far, far more than memorisation.

Standards in Assessment

Another major area of difficulty lies in setting standards. We need to remember that examinations marks, by themselves, mean absolutely nothing. A easy examination gives high marks, a difficult examination gives lower marks. The problem is how to decide the pass mark and the mark required for various grades, like a 'A' pass.

The best way forward is to remember:

> *The standards of learner performance will not vary much from year to year, provided our number of student is high (best over 1000). However, with smaller numbers, learner performance may vary quite a bit from year to year. To decide the pass mark and grade marks, the best way is to discuss this with colleagues and agree on the standard required for a 'pass'. Remember, ignore the actual mark. Another approach is to define the criteria and these set the standards for pass and grades. However, setting the criteria requires experience and it is best carried out by a small group of experienced teachers discussing it together.*

Thus, we have two approaches to setting standard in our assessments:

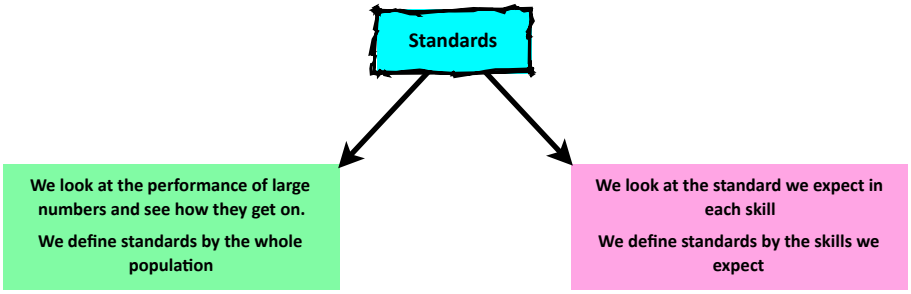


Figure 14.12 Criterion and Norm-referenced Assessment

The green box describes norm-referenced assessment while the purple box describes criterion-referenced assessment.

For an examination board, where there are large numbers of candidates, the more straightforward way forward is to look at norm-referenced assessment. The abilities of the candidates will change little from year to year. However, at some point in time, an initial decision must be taken. For example:

Imagine the following scenario:

(a) The examination results are expressed as grades (from 'A' to 'E' with 'F' representing a failure).

(b) An initial decision is taken to award a pass grade ('A to 'E') to 70% of the candidates in each subject.

(c) Within this,

Award of 'A' pass	5%
Award of 'B' pass	10%
Award of 'C'pass	15%
Award of 'D 'pass	20%
Award of 'E'pass	20%
Award of 'F' fail	30%

These *'decisions'* are offered merely to illustrate the process.

However, whatever is done, passes at 'A' and 'B' should only be offered to small groups, thus allowing the examinations to identify the most able students very clearly

(d) Another way forward involves the use of standarisation. Here the 'raw' marks in every subject examination are standardised so that the mean mark is 60% and the standard deviation is 12 (see page 120-121). This means that the pass mark (to give a something close to a 70% pass rate) is 50%. The enormous advantage of this method is that it stops the media misusing the examination board data to try to suggest that standards have risen or fallen.

(e) Every year, in every subject, marks are standardised and the pass mark is maintained at about 50%.

There are two other advantages of the kinds of systems suggested above.

- They allow universities to add the marks up (legitimately) and then decide each year the total mark required for entry to any specific course.
- If every examination board followed the same procedures, then it would be a major step forward (although not perfect) in establishing common standards, with enormous advantage for the public understanding of examination awards and fairness to candidates. However, this assumes large numbers and that each examination board has candidates from roughly similar educational backgrounds.

In many countries, there have been moves towards greater use of criterion-referenced assessment and, in some subject areas, this has clear advantages. This could be explored in the longer run.

Assessment Confidence

Overall, the key thing: can we have confidence that the assessment reflects what is actually the state of affairs in the learner's brain? Assuming that a test does, in fact, test what it is intended to test, then confidence in its results relate to:

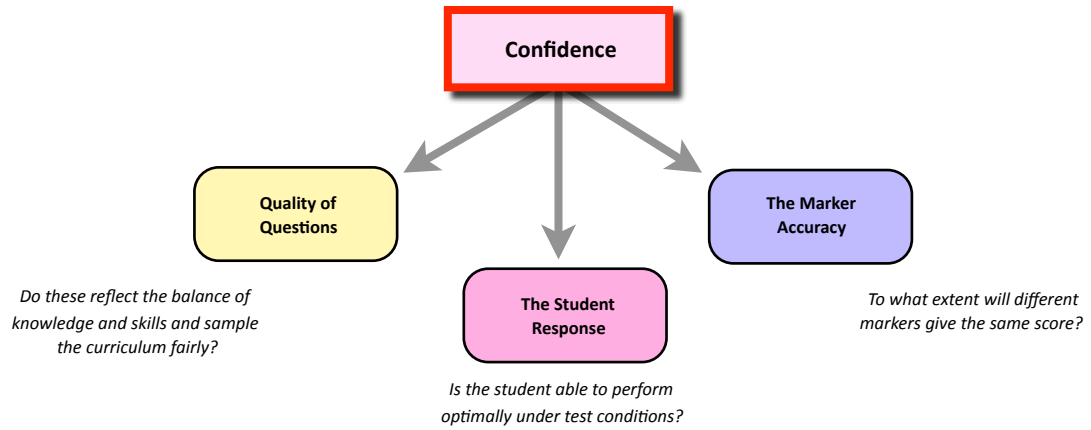
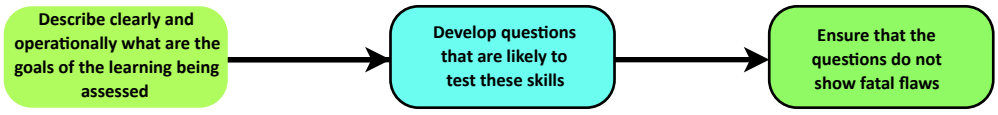


Figure 14.13 Assessment Confidence

Numerous studies have shown that different tests, or different markers, can make an enormous difference to grades, even with carefully designed tests and well briefed markers. This suggest multiple measures are important, with multiple markers.

Any assessment is only as good as the test questions asked. Test questions cannot be good unless they are testing what we want to test in a reliable way.

Here are the key stages to explore this:



One way of approach question quality is to ask three questions of every question you set. This is best done in a small group:

What does this question test?	Knowledge, understanding, skills, thinking or evaluating, communicating, relating to life ?
Is it worth testing?	From the perspective of the learner ?
Are there any obvious flaws in the question?	Information overload, implicit answers, ambiguity.

There is no way national or international examination data can be used as an accurate measure of the standards of education. However, this is how politicians and the media often see it. Sadly, it is often how educational managers and school inspectors see it as well. We need to take very opportunity to educate those who abuse examination data in this way and to show them that education is far more than examinations results and, indeed, there are many factors that will influence examination results much more than the quality of teaching. We also need to discourage our political masters from wasting precious resources in international comparison or making comparisons between schools based on examination data. The evidence shows this approach is invalid, unreliable and very destructive on quality education.

For ourselves, we need to seek, all the time, to develop quality assessment tasks and questions:

Figure 14.14 Features to Achieve Quality Assessment

The Brain and Assessment

All assessment seeks to look inside the brains of human learners and find out what they know, understand, can do, as well as how well they are thinking and evaluating. All this is stored in long-term memory but it has to be accessed through the working memory which has a very small capacity.

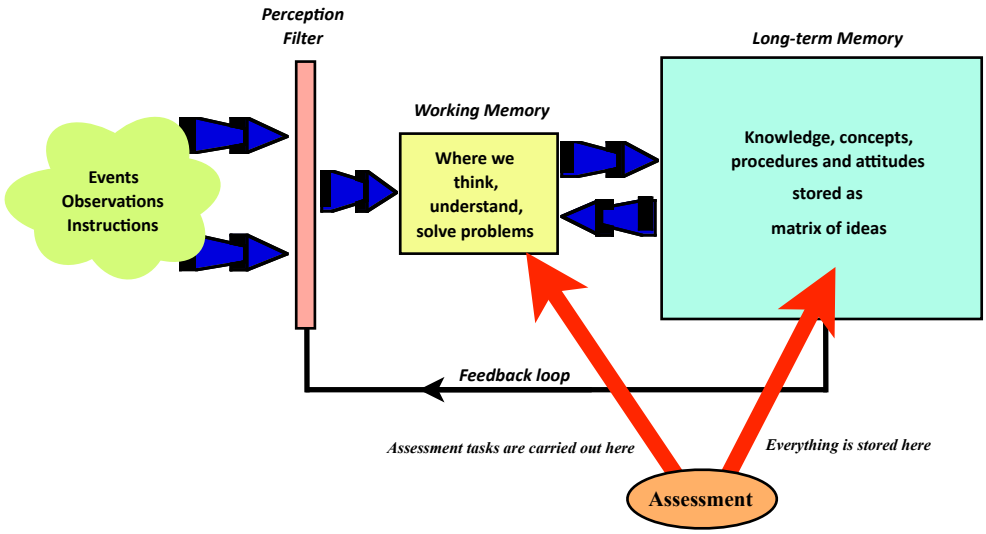


Figure 14.15 The Brain and Assessment

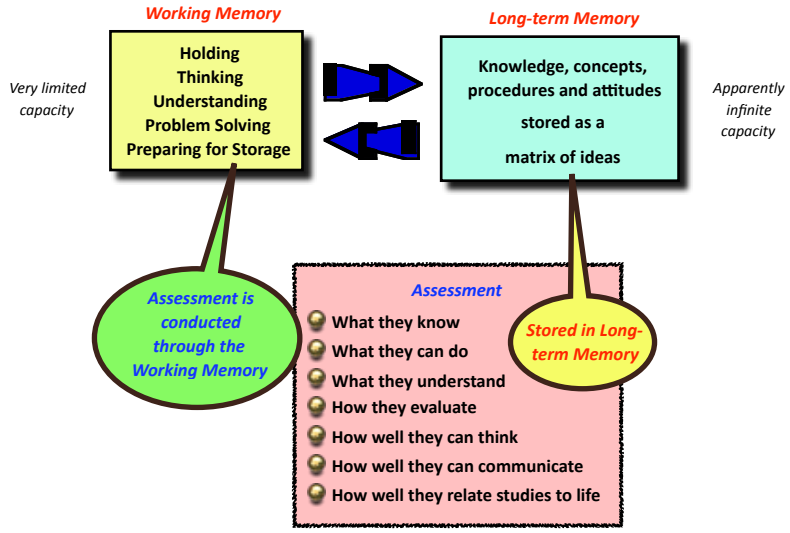


Figure 14.16 The Working Memory and Assessment

The key message is every assessment task or question we set must be capable of an answer *without* working memory overload. If we fail to do this, then we are in danger of simply measuringe the working memory capacities of our students. These capacities are genetically determined and do not relate neatly to ability.

Thinking Skills

In chapter 10, it was emphasised that there are key skills that are of increasing importance in life. A good assessment system must offer rewards for the development of these skills. At the same time, assessing these skills directly is not easy using traditional end-of-course examination papers. This is the place where '*duly-performed*' assessment holds great promise (see pages 88-89). In this way, evidence can be gained of the development of the skills of critical thinking, creative thinking, scientific thinking as well as problem-solving.

These skills can be described:

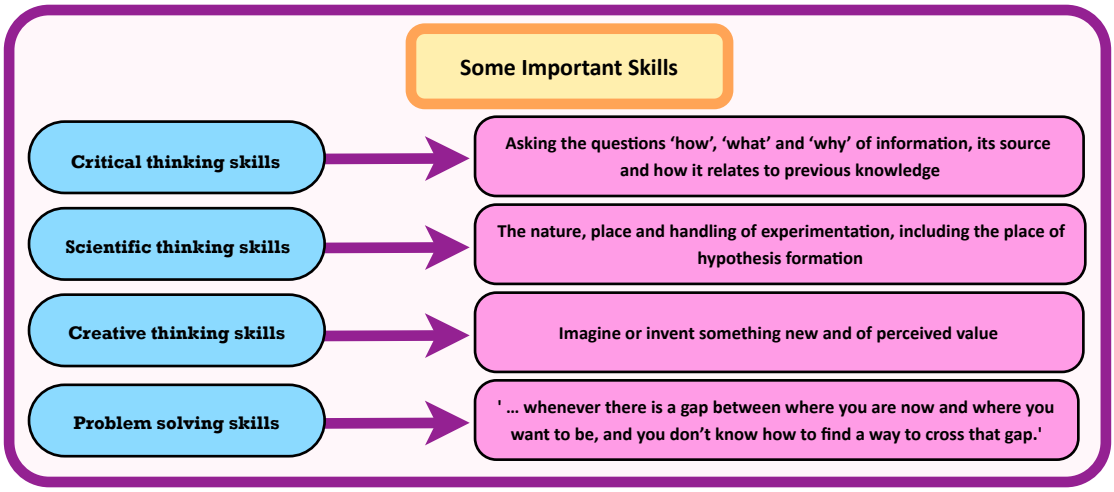


Figure 14.17 Thinking Skills

Figure 14.17 suggests '*operational descriptions*' for three major thinking skills. An operational description allows us to develop assessment tasks which will be able to measure the skill.

Brain designed for thinking

Practicalities of Assessment

There are four broad approaches:

Interviews	Written Examinations	Observations	Task Completion
Our students tell us what they know	Our students write down what they know	Our students shows us what they know	Our student undertake tasks of known value
SPEAK	**WRITE**	**DO**	**UNDERTAKE**
We talk to our students	We read what they write	We watch what they do	We check that the tasks are completed
We LISTEN	**We READ**	**We WATCH**	**We CHECK**
From all this, we deduce what they know, understand, and can do as well as how they think and evaluate			

Looking at these a little more:

Objective Formats		Other Formats	
Example	**Comment**	**Example**	**Comment**
Multiple Choice	They tend only to indicate recall-recognition while reliability is highly suspect	***Short Answer***	Can indicate what they know, can do or, to a limited extent, what they understand.
Partial knowledge multiple choice	More reliable, can be used to give some indication of understanding	***Extended answer***	Versatile, can be used to explore understanding, skills, thinking and evaluation.
Structural Communication Grids	Powerfully diagnostic, especially good at conceptual areas; cannot be used for everything	***Practical Tests***	Can assess cognitive skills (as in mathematics procedures) as well as practical skills.
There are several other objective formats but these have limited application in courses at school and further education stages. There is ongoing work in exploring formats for assessment which use the power of computers. Much of this work has sadly never moved beyond multiple choice variants, with their well established limitations. However, new formats are being considered and, in due course, may have wider applications. Indeed, electronic assessment is a growing area where we can expect developments in the near future. There may come a day when assessments use mobiles, on-line tasks and iPads. Nonetheless, e-assessment faces the major problem about uncertainty of whether an assessment is completed unaided by others.		***Observation***	Very insightful but very difficult to reduce to scores.
		Calculation	Important in many subject areas to test procedural skills and, occasionally, understanding
		Essay	Can assess almost anything but the problem is how to mark fairly. The marking time demand is considerable.
		Dissertation	Offers enormous scope. Marking needs careful thought but good ways to mark have been developed.
		Project	This can reflect extended work and offers many insights. Reducing to a score needs thought.
		Verbal Presentation	Very time-consuming but gives rich insights. Marking needs careful thought but good ways to mark have been developed.
		Duly Performed	Credit is given for task completion and no traditional marking is involved. A very useful way forward to assess some intractable skills.

Table 14.6 Different Kinds of Assessment Methods

Remember: there are powerful ways to assess which are not yet being used much in Pakistan. We strongly recommend:

(a) Structural Communication grids - for assessing understanding, especially conceptual understanding

(b) A reduction in the use of multiple choice and an increase in partial knowledge multiple choice

(c) The use of '*duly performed*' as a powerful way to assess skills that are almost impossible to measure using traditional types of examinations: very cost-effective.

(d) Marking essay, projects and dissertations using the kinds of structured rubrics suggested in chapter 12 will measure important skills with increased validity

Test Format - Some Recommendations

Here is a suggested way to think about choice of test format:

What they know	Multiple Choice Partial knowledge multiple choice Short answer
What they understand	Structural communication grids Essay Projects or reports Verbal presentation
What thy can do	Projects Dissertations Duly performed Observation Verbal presentation
How well they think	Duly performed Dissertation Essay
How well they evaluate	Essay Project Dissertation Duly performed
How well they communicate	Essay Project Dissertation Duly performed
How well they relate to life	Essay Project Dissertation Duly performed

Remember:

- Never depend on one test or examination
- Use multiple short assessments during and at the end of a course
- Use many different formats in any assessment programme
- Always plan assessments with a clear specification of what you want to assess
- Avoid testing just recall - its is an unimportant skill in the internet age
- Marks have no absolute meaning at all

Warning:

- Assessment powerfully frames how students learn and what students achieve. It is one of the most significant influences on student experiences of education and all that they gain from it.
- Assessment is the making of human judgements about how the achievements of students.
- Assessment plays a key role in the certification of students and can open or close doors for further opportunities in life.

The key findings can be summarised:

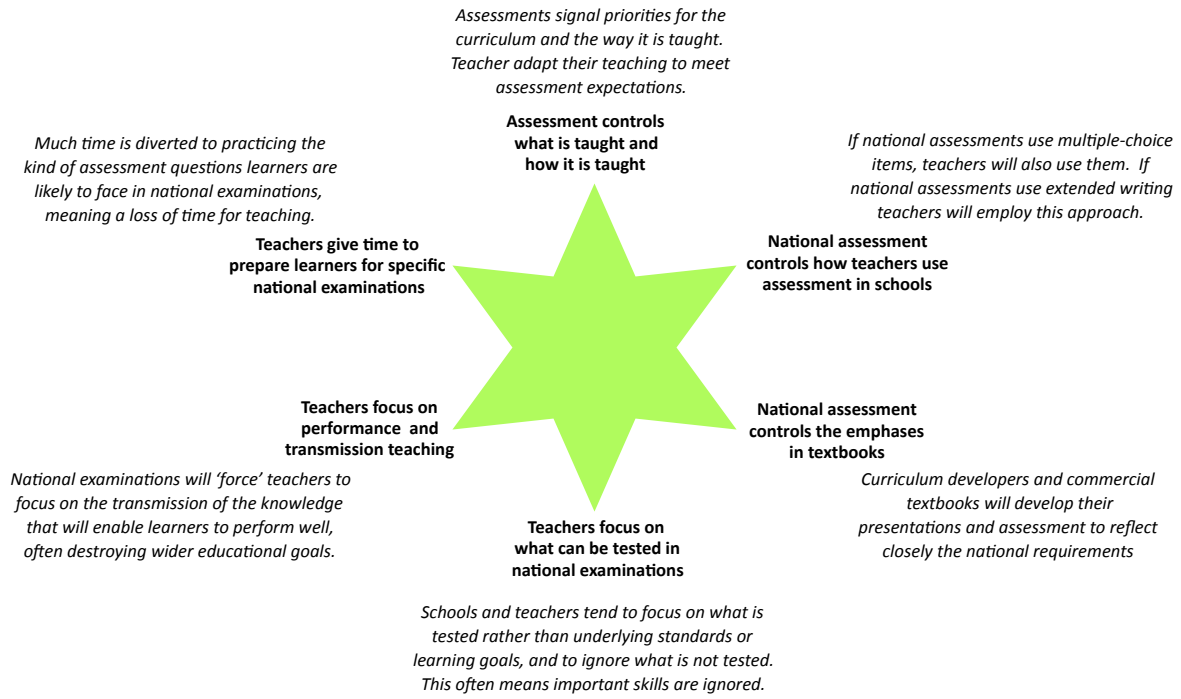


Figure 14.18 Power of National Assessment

The research clearly shows that whatever is measured matters. Educators tend to model and mimic in their curriculum and instruction the content and format of high visibility assessments and to use a significant amount of classroom time for special test preparation activities. In some countries, however, testing has become dominated by routine, and highly predictable, items which are also often short and highly structured, thus reducing the expectation that students should apply knowledge, skills and broader capabilities demanded by today's world.

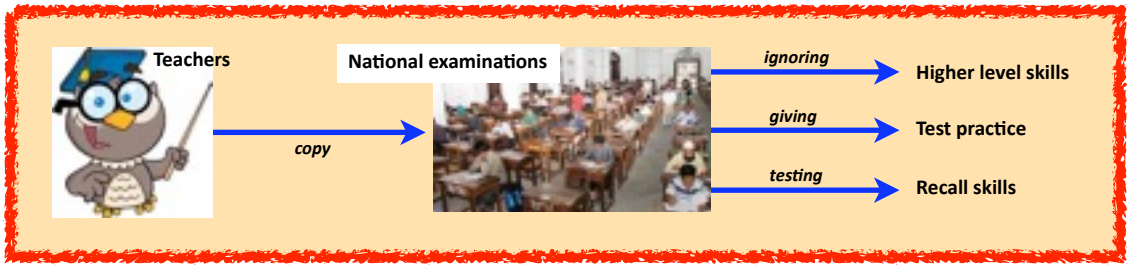


Figure 14.19 National Assessment can be anti-educational

Key Principles

Here are ten key principles:

(1) Most assessment is assessment of learning that has taken place. There is a need to develop assessment approaches that will assist future learning and this must be employed in such a way that useful feedback on learning can be offered to learners so that future learning can be enhanced.

(2) Assessment today must reflect the needs of learners as they move out into society where information is easily available but where understanding and the wise application of that knowledge is what societies need.

(3) Assessment must not over-dominate the educational experience, and it should demand excessive time or resources, including teacher resources.

(4) Assessment needs to be aligned with instruction but it must never control what is taught and how it is taught.

(5) Assessments need to be designed against curriculum goals and it needs to be checked if the evidence obtained does, in fact, reflect skills relating to these goals.

(6) Assessment needs to focus on what learners understand and can do rather than be seen as a kind of '*certificate of failure*', focussing on gaps in the understanding and experience of learners.

(7) Assessment, used to assist learning or for final certification, must NEVER be used for quality control as this will distort the assessment process.

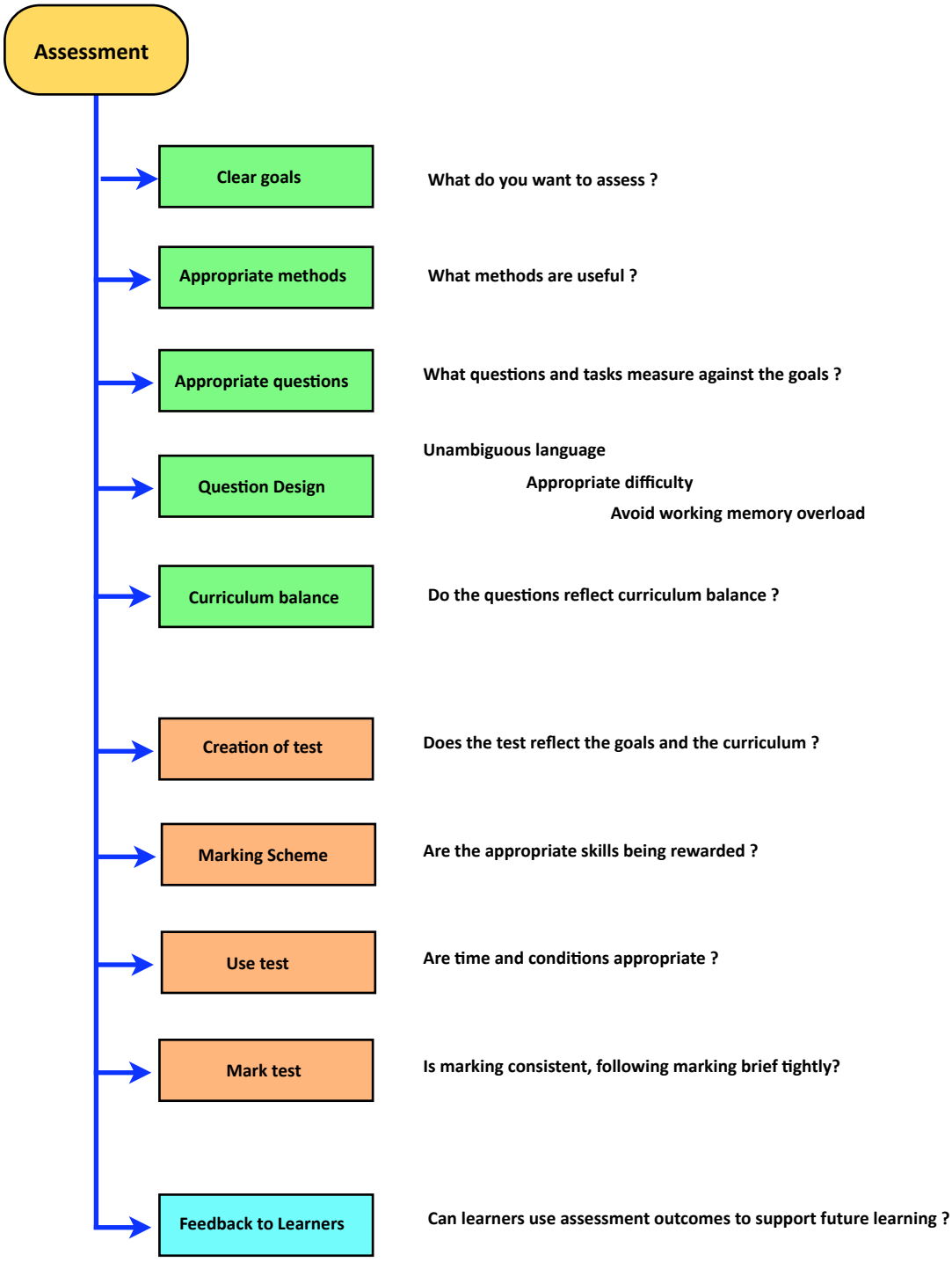
(8) No course assessment should rely on one assessment like an end-of-course examination. All courses should employ multiple assessments, preferably not too long and employing a variety of assessment approaches. The use of multiple choice needs to be reduced considerably as it is known that this approach is limited in scope and is unreliable.

(9) Assessments should be developed that employ new technologies where appropriate and where the new technologies enable assessment to be more effective or more efficient or both.

(10) It has to be remembered that all examination marks are merely estimates of performance in a specific examination paper or task at a specific time. Therefore, assessments are usually poor predictors of future success.

The Assessment Process

In this diagram, start at the top and work down…..



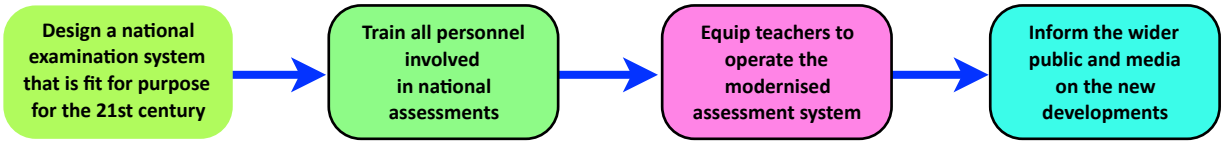
These question for each stage of the process offer guidance so that quality assessments can be developed, employed and interpreted. The impact of such assessment on the overall quality of education in Pakistan will be simply enormous, for national assessment controls most of what will happen in the schools.

Key Issues for the Future

In order to move the country forward in education, it is recognised that assessment holds the key. It is, therefore, essential that all assessment at school and university levels is informed by the widest range of research and adopts the most powerful and effective ways to assess learning. To that end, the following are recommended:

- Radical overhaul of national assessment to build these on the findings of research, employing a much wider range of assessment formats and reducing the emphasis on recall for success.
- Developing mandatory courses for those involved in managing, setting, marking and data handling with national and provincial assessments to enable new policies and procedures to be developed, all based on research evidence and high quality practices.
- Developing new courses on assessment for all initial teacher education and continuing professional development, aiming to enhance the assessment skills of teacher.
- Exploring how to develop visual and written materials on assessment with the aim of educating parents and, even more importantly, the wider media.

This is a sequence of operations:



but ....

leads to a ....

In the light of what has been drawn from the research evidence, the next sections of the monograph offer an outline of the practical ways forward, the aim being to bring the vision outlined in the first 14 chapters into being. The emphasis will be on practicalities related to implementation.

The Academic Examiner
Developing the Practical Skills

How do we do it ??

Assessment Moving *into* a New Future

The Big Decisions

Aims

This chapter considers the decisions needed to be taken at Board level, making suggestions about possible ways forward related to examination structures related to agreed objectives

We start by looking at the key issues that require decisions by an Examination Board. Specific suggestions are made as the basis for discussion and the implications are then worked though in practical terms. The key point, however, is that decisions have to be taken before we can move to the practicalities of examinations setting and the decisions suggested are offered to *illustrate* a practical way forward.

The Goal of Formal Examinations

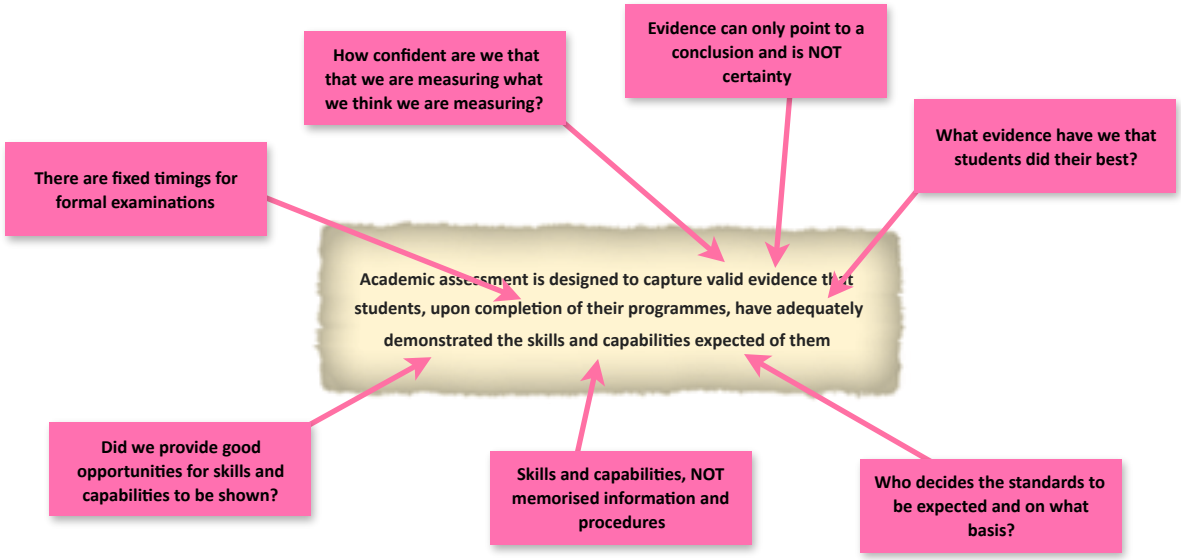
The Examiner's Goal

Most countries set formal examinations at various stages towards the end of school education, often over the ages of 16 to 18. Most countries have set up formal examination boards or authorities to conduct the entire process. Most of these boards or authorities employ skilled examiners to set the examination papers and to oversee the entire examination process. In many countries, teachers are employed as examination setters and markers and trained appropriately. In many countries, independent staff are employed to oversee the actual examination processes ins schools and to supervise and check the marking processes. However, it has to be recognised that, in some parts of the world, there is no formal examination process and the entire certification of school students rests in the hands of school teachers. The evidence shows clearly that the process works well and has generated remarkable public confidence. There is considerable evidence that shows that the *most reliable* assessment arises when formal national assessment involve in-school teacher assessments. This is what is proposed here but a very cautious introduction is suggested.

The goal for all this work can be stated:

Academic assessment is designed to capture valid evidence that students, upon completion of their programmes, have adequately demonstrated the skills and capabilities expected of them

There is more to this than might appear:



There is considerable human judgement in all this. Formal examinations and tests are not as accurate or precise as many think. We need to:

- Check that we are testing what we want to test (too often, we end up rewarding recall skills).
- Recognise that even the best examinations are only approximate measurements on a specific day.
- Know that standards of performance are decided subjectively based on experience of teaching.
- Ensure that our examinations offer opportunities for students to show their best.
- Take every step possible to develop assessment tasks that measure skills and capabilities, not recall.

The aim of this part of the monograph is to outline practical ways to design, develop and interpret formal examinations which will be employed to certificate school students.

Examination Specifications

The curricula for each subject discipline has been prefaced with sets of goals, objectives and specific objectives. These need to be *interpreted* in terms of sets of objectives against which we construct our assessments for the formal examinations. It is probably better if these sets of objectives for every subject to be examined are specified in the same broad terms. This offers consistency across the entire examinable curriculum.

One way forward is to develop these in seven general areas (see page 135 for a justification for these seven areas), giving the 7 key skills areas that students have to demonstrate in formal examinations:

Key Skill	What they students needs to be able to demonstrate
Know	Showing that they know key information and procedures
Understand	Showing understandings of key concepts, insights and procedures
Be able to do	Demonstrating relevant procedures (practical or intellectual)
Be able to think	Demonstrating specific kinds of thinking important in a specific discipline
Evaluate	Be able to assess information, conclusions and understandings critically
Communicate	Being able to communicate ideas coherently and logically
Relate to life	Being able to show how what they have studied relates to life

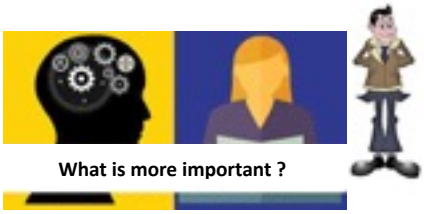
We can give examples of how such a table might work in some subject areas:

Mathematics	
Key Skill	**What they students needs to be able to demonstrate**
Know	Key procedures for specific computations listed in the curriculum
Understand	Being able apply key mathematical ideas in novel situations
Be able to do	Being able to carry out key procedures with acceptable accuracy
Be able to think	Being able to demonstrate logical thought in carrying through procedures
Evaluate	Being able to judge whether an *'answer'* is reasonable or consistent
Communicate	Being able to show that they can communicate mathematical ideas clearly
Relate to life	Being able to see the significance of specific procedures in ordinary life

Biology	
Key Skill	**What they students needs to be able to demonstrate**
Know	Key facts and procedures important in biology
Understand	Being able apply key biological ideas in novel situations
Be able to do	Being able to carry out key procedures with acceptable accuracy
Be able to think	Being able to demonstrate scientific thought in biological enquiry
Evaluate	Being able to judge whether an *'answer'* is reasonable or consistent
Communicate	Being able to show that they can communicate biological ideas clearly
Relate to life	Being able to see the significance of specific understandings in ordinary life

English	
Key Skill	**What they students needs to be able to demonstrate**
Know	Key words and language structures
Understand	Being able apply key principles in novel situations
Be able to do	Being able to read, write, listen and speak to agreed standards
Be able to think	Being able to demonstrate creative thought in communicating
Evaluate	Being able to demonstrate critical thought in evaluating communications
Communicate	Being able to show that they can communicate ideas clearly orally and in writing
Relate to life	Being able to see the significance of English globally

All of the above is a matter of examination board policy and needs to be specified, following agreement across subject areas. However, there is another aspect that needs policy agreement. Are these seven objectives of equal importance ?

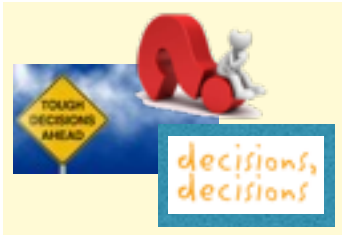


Here is a suggestion as an immediate possibility:

	Objectives		%
Evidence of student skills in relation to:			
1	Know	What they can recall or recognise	40%
2	Understand	Can they apply knowledge in novel situations?	20%
3	Do	Can they carry out key procedures correctly?	20%
4	Think	Evidence of creative, critical and scientific thinking?	5%
5	Evaluate	Can they weigh information for its accuracy, validity, usefulness?	5%
6	Communicate	Can they explain key ideas clearly?	5%
7	Relate	Can they appreciate the importance and impact in life	5%
		Total	**100%**

However, the Board can develop its policy over time and here is a suggested set of goals for, perhaps, 5 to 10 years ahead:

	Objectives		%
Evidence of student skills in relation to:			
1	Know	What they can recall or recognise	30%
2	Understand	Can they apply knowledge in novel situations?	20%
3	Do	Can they carry out key procedures correctly?	20%
4	Think	Evidence of creative, critical and scientific thinking?	10%
5	Evaluate	Can they weigh information for its accuracy, validity, usefulness?	10%
6	Communicate	Can they explain key ideas clearly?	5%
7	Relate	Can they appreciate the importance and impact in life	5%
		Total	**100%**



Setting agreed specific objectives with the percentage weightings is not easy. It needs careful consultation across all subject disciplines and there then needs to be consultation with experienced teachers. It is also something that cannot be implemented immediately in that teachers will be teaching to meet *previous* examination criteria and it takes a year or two for teachers to teach a new cohort of students using the new objectives.

Time must be allowed for adequate consultation, a general measure of broad agreement and then time to allow teachers to teach in line with the new objectives.

Here is a suggested timeline showing what needs to be done although the timings given are only indicative, However, it stresses that the process will take time. The endpoint is the time for the first examination under the revised system.

Stage	Task	Defining task	Approach	Time
1	Agreed policy at Board level	The goals for ALL examinations	This may be based on the seven goals suggested above or some adaptation of them	2 months
2	Subject Consultation	Board committees for each subject interpret the goals	Essential to check if proposed goals will work in all areas of the curriculum	2 months
3	School Consultation	All schools informed of proposals and teachers invited to submit comments by a given date	Use electric submissions and set up a central email to receive these and distribute them appropriately across Board committees	2 months
4	Goals Modification	In the light of submissions, goals are adjusted	The most likely greatest need will be clarification of wording and interpretation	1 month
5	Inform Schools	Circulate revised goals to all schools	Prepare teachers with a clear timetable for implementation	1 month
6	Exemplar examinations	Develop exemplar examinations papers using new goals, for all subjects	This will take time for questions development, question shredding, formation of final paper, all set against the specified goals	5 months
7	Implementation time	Allow adequate time before the new assessment procedures are implemented in national examinations	This allows teachers the time to adjust their teaching, their in-school assessments	12 months
				Nearly 2 years

From Objectives to Examination Formats

The next big decision is to move from lists of objectives to decisions about the way examinations are to be structured.

Several major decisions must be taken:

Key Decisions

in-School Assessments ?

It is suggested that 10% of the total credit be awarded by schools and the way this can be done with fairness is outlined later

90% of the credit comes for the end-of-course formal examinations

The proportions can be adjusted over time, the evidence showing that combining external assessment with teacher assessment gives the most valid and reliable data

Use of Objective Questions ?

It is suggested that the use of multiple choice be reduced, perhaps to nearer 10%

Modifying multiple to partial knowledge multiple choice may be a better way forward

The use of structural communication grids should be introduced in some subject areas, up to 10% of the credit coming from such questions

Extent of Common Formats ?

Different subjects have different requirements and there should be no common pattern of formats across subject disciplines

Each subject should come up with proposals that reflect the need of the subject best

Suggestions are offered later

There is a view that only externally set formal examinations can give reliable information. However, this is NOT supported by the evidence. Teachers know their students best and are often very highly skilled in undertaking less formal assessments. The practicalities of doing this will be discussed later in order the avoid cheating, undue pressure on teachers and to make the assessments fair and meaningful.

In using objective testing, there should be a move to a reduction in the use of multiple choice in that research raises many serious questions about the reliability of such questions. Partial knowledge formats are better while structural communication grids will help in some subject areas. Both formats are known to be more reliable and valid.

It is tempting to impose a fixed format on every subject with a specified proportion of objective questions, short answer questions and so on. However, this fails to recognise the very different needs of different subjects and some flexibility needs to be allowed.

The value of flexibility

Question Formats

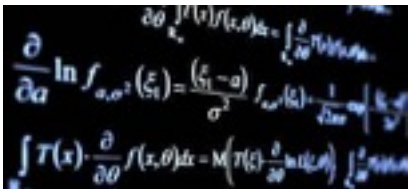
In the first part of the monograph, many formats for questions were presented. Here, an attempt is made to suggest the best formats that might suit different subjects areas.

Subject	Objective Formats	Non-Objective Formats	In-school Credit
Mathematics		Short answer: 50% Longer answer: 40%	Mathematics applications: 5% Project: 5%
Languages	Multiple choice or partial knowledge multiple choice: 5%	Short answer: 25% Essays or longer questions: 60%	Technical essay: 5% Communication skills: 5%
Social Subjects	Multiple choice or partial knowledge multiple choice: 10%	Short answer: 30% Essays or longer questions: 50%	Essay: 5% Project: 5%
Sciences	Multiple choice or partial knowledge multiple choice: 10% Structural Communication grids: 5%	Short answer: 50% Longer answer: 20% Open book: 5%	Laboratory work: 5% Group work units: 5%

The table above presents four very broad discipline areas and makes suggestions about possible ways assessment might be structured. It has to be stressed that these are suggestions but the information on the following page expands and illustrate show this might be carried out.

The meaning of the table is now expanded:

Mathematics: The subject does not lend itself to objective testing (which ends up testing the recall of trivial facts). Mathematics is a skills-based subject. Can the students undertake mathematical procedures and do they understand what they are doing? This is best tested using short and longer questions. During the course, students are given open-ended problems to solve while they also undertake a short project where they have to demonstrate that they understand the significance of some mathematical technique or idea as it applies in life. These are graded by classroom teachers and then samples are cross-marked by experienced and trained examiners.



Languages: Some key facts and information can be assessed using objective testing but most of the credit will come from more sustained answers. During the course, students are required to write a number of '*technical*' essays, essays on topics and themes which are relevant to *them*. They are marked for skills like clarity, communication, logical thought, and not for content. The best essay will give the credit, the process being discussed later (page 179-180). Samples are cross marked by experienced and trained examiners. Listening and speaking skills can be assessed during the course and agreed marks awarded by the class teacher. Statistical moderation can ensure standards: statistical moderation will be discussed later.

Language	%
Balochi	3%
Pashtu	14%
Punjabi	45%
Saraiki	11%
Sindhi	15%
Urdu	8%

Social Subjects: This is interpreted widely to include subjects like Islamic Studies, Pakistan Studies, History, Geography, Economics - all subjects that relate to people and their behaviour. Some information and understanding can be assessed using objective testing. The possible use of structural communication grids needs explored. However, assessment should depend heavily on questions, many of which will demonstrate sustained argument and understandings. During the course, students are required to write a number of essays on set themes (how this will work is discussed later - page 179-180) as well as undertake a short project. These should be marked for skills of understanding, sustained thought, clarity of ideas and so on. Samples are cross marked by experienced and trained examiners

Sciences: These subjects are highly conceptual and this is where structural communication grids show great promise (set at only 5% at the outset this may be increased while the multiple choice might be decreased). The written papers should contain at least one question where all the information is given and the students have to use that information to observe patterns, take deductions, show critical and scientific thought and so on.

Agreed Ground Rules

Assessment Moving *into* a New Future

Quality in Practice

Aims

This chapter outlines how the course objectives might be interpreted into paper structures, question formats along with the practical steps needed to ensure quality in the various processes

This section now considers the practical issues so that we can ensure that there is quality. Quality is when the grades awarded reflect fairly and validly the achievements of the learners set against the goals for the courses being studied.

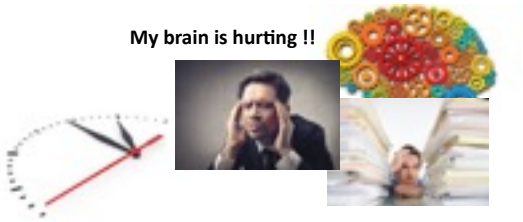
The Shredding Process

The process of examining goes through a series of logical steps, some of which are now shown:

Determine Course objectives	Using the seven suggested objectives or some modification of these, agree course objectives
Percentages credit for each course objective	Agree what percentage of the credit will be gained under each course objective
Decide on best question format	Decide which type of question can be employed to test the agree objectives
Create questions	Create questions to meet the examination specification, creating many more than required
Shred questions	In small teams (often 3), consider all the questions asking what they test, whether it is worth testing and are there any obvious flaws
Create paper	From the questions that survive the shredding process, develop a paper
Check paper	Analyse paper to see if it is testing against the agreed objectives taking into account the amount of credit for each objective

Decisions to be taken by the Examination Board

Developing a quality examination paper that genuinely tests what we want to test is a slow, demanding process. In some countries, well over 12 months are allowed for the process, the final 3 months being used for printing and distributing the papers. Considerable patience and creativity are required.



It is quite useful to have several people creating possible questions and then they sit down together (or it can be another group) to shred the questions. The essential process of shredding is:

For every question or item, ask:

What is the question testing?
Is it worth testing?
Are there any obvious flaws?

Work in a group of about three

The problem is that most questions or items merely test the recall of some information or procedure. Mostly, this is not worth testing. There is also a tendency to set questions on areas of the subject where it is easier to set questions, neglecting other areas. The most common flaws relate to working memory overload and ambiguity in the language. We shall look at the practical details of how shredding works later but there are general principles.

The best way to shred questions is to *think as the candidates* will think. The best people to shred questions are often experienced school teachers for they are best placed to think as their own students think. Students have a simple agenda:

The Learners Goal
Find a way to get a good answer legitimately,
with the minimum of effort !

Often, there is something in the question which leads them to the answers but this was never planned by the question setter. The real issue is validity: are the marks going to be gained for the skills we intended to measure ?

Never forget the problems of language. It is bad enough in the student's first language. To be tested in English, which is not the first language, makes it much, much more demanding. Here are some simple principles:

- Keep the question text brief
- Keep the sentences short, avoiding relative clauses
- Avoid using specific words that may be unfamiliar
- Diagrams may even cause confusion if they are too complex but, overall, diagrams usually can help
- Layout is everything and good layout can aid clarity of communication greatly

If a question does not test what you intend, then it is an INVALID measure.
To create a valid question is far more difficult than it might appear.

Examination Structure

We now reached the point where we have many questions which have been shredded and we use these questions to construct an examination paper. The questions is how do we do this?

Here is a way to start:

(1) Take each question, decide how many marks it is worth and then specify what the question is testing.

(2) Go back to the examination specifications and place the questions under each heading.

(3) Now build the paper so that the balance of skills to be tested reflects the examination specifications.

(4) Finally, check to see if content coverage is reasonable and reflects the curriculum.

This all seems a bit abstract! Let us illustrate this by considering some examples.

Consider an examination in, say a subject like geography. This is an example, not a prescription. You will need to adapt and modify to suit the decisions you take.

You have already decided that the examination format is:

Subject	Objective Formats	Non-Objective Formats	In-school Credit
Social Subjects	Multiple choice or partial knowledge multiple choice: 10%	Short answer: 30% Essays or longer questions: 50%	Essay: 5% Project: 5%

You are also working in the following agreed structure:

	Objectives		%
Evidence of student skills in relation to:			
1	Know	What they can recall or recognise	40%
2	Understand	Can they apply knowledge in novel situations?	20%
3	Do	Can they carry out key procedures correctly?	20%
4	Think	Evidence of creative, critical and scientific thinking?	5%
5	Evaluate	Can they weigh information for its accuracy, validity, usefulness?	5%
6	Communicate	Can they explain key ideas clearly?	5%
7	Relate	Can they appreciate the importance and impact in life	5%
		Total	**100%**

In practice, this means something like:

- ☑ 10 partial knowledge multiple choice questions
- ☑ 30 marks for short answer questions
- ☑ 50 marks for more extended writing

You are leaving the in-school assessment to give evidence of being able to appreciate the importance and impact of geography in life as well as some evidence of critical thinking and evaluating. We shall return to the practicalities of in-school assessment later.

Putting the Geography Examination Paper Together

Here are a series of steps you might wish to follow:

Step 1

Look at the partial knowledge multiple choice questions you have. Select the best 10 in your opinion and check if they reflect content coverage (avoid two questions assessing the same content or exactly the same skill).

Step 2

Now look at the short answer questions. These will largely measure knowledge, understanding or carrying out key procedures. Pick what seem the best and, again, check if they reflect content coverage (avoid two questions assessing the same content or exactly the same skill).

Step 3

Now look at the questions requiring extended writing. You might wish to have 5 questions at 10 marks each. You wish to assess understanding but, more importantly, the abilities of the students to think and evaluate as well as abilities to communicate with clarity and logic. Pick the best 5 questions and, as before, check if they reflect content coverage (avoid two questions assessing the same content or exactly the same skill).

It is often useful to have each question or item on a separate piece of paper. Use a large table and lay them out in order. Now place them in order as a bundle of papers.

Go and take a break !!
Come back at look at the possible paper tomorrow.

You will return with a fresh mind. Look at the possible paper. Ask some key questions:

- Overall, does it reflect the kind of skills you wish to test?
- Overall, does it reflect the course the students have been following?

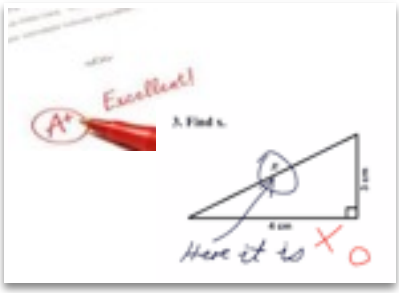
Make any adjustments you think might be needed and then analyse the paper again, using a table like the one shown on the next page. This gives a check if your paper is measuring the skills that you plan to measure.

Analysing the Paper Structure

		Knowing	Understanding	Doing	Thinking	Evaluating	Communicating	Relating
Partial Knowledge Multiple Choice	1							
	2							
	3							
	4							
	5							
	6							
	7							
	8							
	9							
	10							
Shorter Answer	1							
	2							
	3							
	4							
	5							
Extended Writing	1							
	2							
	3							
	4							
	5							
In-school credit	*Essay*						3	2
	Project					3		3
Total Marks								
Target Marks		40	20	20	5	5	5	5

This analyses the paper to see whether it does, in fact, reflect the objectives of the assessment. It is almost impossible to achieve a perfect match between what the paper seeks to reward and the target marks. However, such a table does give some evidence that the paper is reflecting course and assessment objectives. It is also important in that it places emphasis on reducing the the proportion of marks that are simply given for student recall information or procedures. If assessment is to move forward, the place of recall MUST be reduced.

At this stage, a marking brief is required. This show how every mark is to be allocated, indicating possible answers or criteria.



In developing this, possible ambiguities of language may still be found and can be adjusted. The art of a good marking brief can be summarised:

Once this analysis is completed, it is possible to adjust questions or to replace questions from the pool of questions that passed through the shredding process. At this point another major check is required:

A valid examination paper

We want a paper that rewards the skills we want to reward. At this stage, ask, say, three '*experts*' to look at the paper overall. The three '*experts*' can be drawn from:

(a) An experienced practising school teacher

(b) An experienced master examiner

(c) A subject specialist (eg. university lecturer in the subject)

Their remit is to consider whether the paper reflects the curriculum goals overall and the paper seems balanced and fair. In addition, they can make checks on the marking brief. Their reports can then be given back to the paper setter and final adjustments made.

Setting a Good Examination Paper

It is now obvious that setting a good examination paper is frustrating and very time-consuming. Going through all the procedures and taking the time is essential. The goal is to have an examination paper that is valid and fair. Creating good questions is not an easy task. Creating questions that merely test recall is easy but that is not going to serve our future students well. We want to reward understanding and the ability to apply their understandings in useful ways. We want to develop a population who are willing and able to question, evaluate and think. We want our future students to be equipped to make a contribution to society and make a positive difference throughout their lives. The controlling power of national assessment is probably by far the most critical factor in enabling this to be possible. The time invested into generating quality examination papers will be returned many times over in generating a population of learners who will grow and develop in exciting ways, bringing benefit to all around.

The desktop of the good paper-setter !!

Assessment Moving *into* a New Future

Examining Quality

Aims

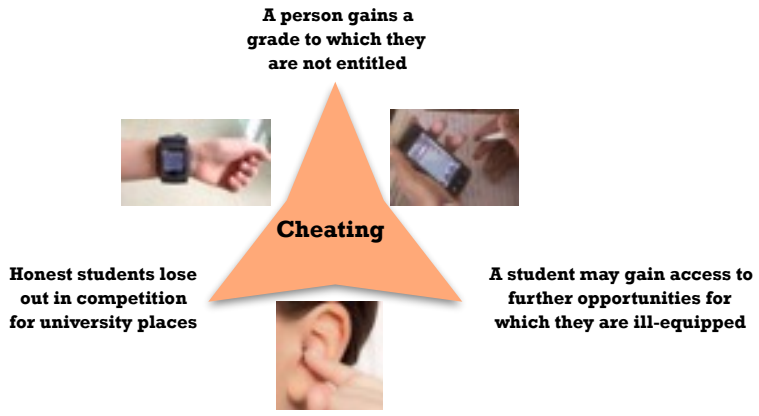
This chapter offers an overview of how to develop quality in relation to cheating, marking, the award of valid grades, handling data, and developing and using orders of merit

Examining Procedures

Examination papers have to be distributed and stored in such a way that security is maintained throughout. The examinations then take place at the appointed time on the appointed day. In many countries, those who supervise the examinations (invigilators) are drawn from *outside* the education system - often those who have retired from various professional jobs, including retired teachers. It is important that no one can enter the examination area during the time examinations are taking place. It is also important that the candidates do not indulge in any kind of cheating.

The problem is that modern electronic equipment can make cheating far too easy. Universities now take great care to check on plagiarism, a problem that has been largely created by the internet. However, students today can access more or less any information from iPads, and even from the iWatch ! Mobile smart phones give access to almost anything while specific information can be stored on a phone. This makes the integrity of examinations a very difficult issue to deal with and there have to be very clear '*rules*' set by the Board and followed exactly by every examination centre in order to minimise cheating.

Cheating has many outcomes that are most serious. Here are a three:



However, there is a longer-term much more serious issue. The candidates who cheats know they have cheated. Their integrity has been compromised and that may last a lifetime. They will never know what they can actually do, without help from outside. They move into life with an attitude that success can be gained at the expense of others. Society suffers. Their very integrity has been fatally compromised. They are damaged people.

However, cheating can be considered from another perspective. Most cheating depends on the fact that answers to examination questions merely reflect memorised knowledge. If the emphasis is moved steadily away from rewarding the recall of information and procedures, then cheating is not so easy. If questions supply the information and require the candidates to use and apply that information, showing that they understand, then cheating becomes much more difficult to undertake. This is another incentive to change the focus of formal assessment.

Nonetheless, there are deep cultural factors at work. Achieving high grades - no matter what the cost - is seen to be the key to access to better jobs and, therefore, better remuneration. If a society rewards status, wealth and the supposed security that these may bring, then there is an underlying drive to achieve high grades, and any way to achieve these is acceptable. It is important to recognise that, while it is important for any examination board to take all the steps it can to minimise cheating, there are cultural factors that may make the complete elimination of cheating an impossibility.

Getting the Marking Right

Marking is perhaps one of the most boring tasks that teachers at any level have to undertake. Setting examination papers can be creative, if demanding. Marking script after script can be soul-destroying and, as boredom sets in, standards may slip.

The joy of marking

Here are some simple ways to reduce some of the problems:

Who marks:	In many countries, practising teachers are hired as markers. They know the curriculum and the way to understand what candidates may write. However, it is important that they never take any scripts into their own school, for security reasons.
Marking briefs:	If there is a clear marking scheme, laying down exactly what gains the credit for each mark, this helps the markers to maintain similar standards.
Markers Meetings:	In some countries, about a week after the actual examination, markers meet in teams to agree the final details of the marking scheme. Where this is possible, it certainly helps consistency across different markers.
Checking marking:	Inevitably, some markers will mark slightly more harshly than others. To ensure standards, samples of scripts need to be re-marked centrally and the entire marks from any marker raised or lowered slightly in the light of this.
Marking Inconsistency:	This is the most difficult problem and, in some countries, if inconsistency is found, all the papers of that marker are re-marked and the marker is removed from the register of markers permanently.
Total Checking:	It is sensible for teams of staff to check the totalling in all papers and enter the data into spreadsheets in computers. This allows for analysis of questions afterwards, a process that can be very informative for any examination board.

Data Handling

It has to be remembered that examination scores are not very accurate !! There are many sources of error. Here are a few:

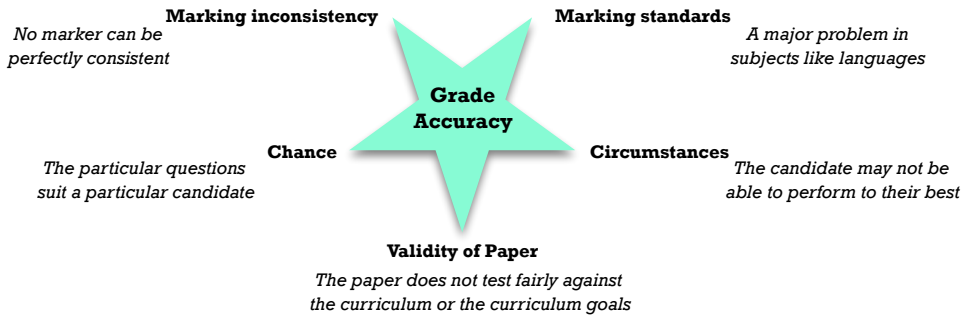


Figure 17.1 Grade Accuracy

The best way to improve the accuracy of examinations is NEVER depend on one paper. It is even better to have candidates sitting two papers, separated by an hour or two. However, this is one of the strongest arguments for using school assessments - it gives another assessment situation. Overall, there is one good way to assist accuracy and this involves using school predictions.

An examination is a test of student abilities on one day, under one particular set of circumstances, using one measuring tool (examination paper) that is far from perfect

Using School Predictions

Each schools is asked to submit an order of merit for all candidates who will sit each subject, indicating the grade that the school expects. Teachers are incredibly good at placing their own students in a rough order of merit. Here is a picture of what might be submitted for, say, 20 candidates in one school who are sitting, say, a chemistry examination, using a five point scale:

Subject Chemistry 2017		Level HSSC II	
Candidates			Predicted Grade
Order	Name	Number	
1			A
2			A
3			B
4=			B
4=			B
6			B
7			C
8			C
9			C
10			C
11			C
12			C
13			C
14=			C
14=			D
16			D
17			D
18			E
19			F
20			F

The table can be interpreted:

- The class teacher completes the table and is offering a prediction of the grades expected based on class teaching and class assessments.
- For the moment, we are assuming passes at grades A, B, C, D, and E, with F representing a 'Fail'. However, the procedure works well using any marking system.
- Where two candidates have performed equally well in school tests, they are awarded the same place in the order.
- This '*Order of Merit*' can be extremely helpful in ensuring that all candidates are treated fairly. This is carried out in the following way.

Using an Order of Merit

If a candidate does not achieve the expected Grade in the formal Board Examinations, the school can appeal. However, to stop frivolous appeals, the following conditions are applied:

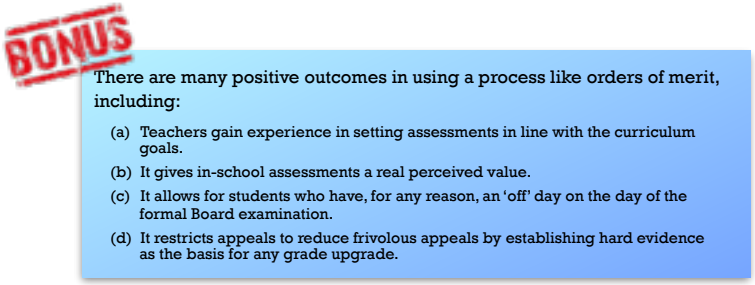
- The candidate *must* have scored at least one Grade *lower* than expected (or a minimum number of marks on a percentage scale: maybe a minimum of 10%)
- The order of merit generated by the school *must* match closely the order of merit in the formal Board Examinations.
- The appeal *must* be supported by submission of the school examination papers and scripts for the candidate and the candidates immediately above and below the candidate in the school order of merit.
- Only schools can appeal, parents or candidate into being allowing to appeal.

The appeal for an upgrade will *only* be granted if:

(a) The school order of merit is similar to the Board order of merit, this being the evidence that the school is assessing the same skills as the Board.

(b) The submitted evidence of examination and papers and scripts shows the school assessment to be robust.

This kind of system is operated in many countries and works quite well.



After the papers are marked, the data sorted out, the grades awarded and the outcomes of the entire examination process are made known to the schools and candidates, it might be thought that we can relax just a little. However, it is very often at this point that more problems arise. We have to recognise that examination grades not only offer some kind of recognition of the achievements of the candidates but they also are used by the media and politicians for other purposes.

- *As the number of candidates increases, do we allow the same proportion to gain a pass?*
- *As educational leaders and politicians seek to demonstrate improvements, do we allow more to pass to satisfy them?*
- *If we pass, say, 70.2% this year when we passed 69.9% last year, the media will claim greater success for schools OR will claim that standards have fallen. Neither might be true !*

We have to recognise that the data for any examination system will abused and misused by wider society. There is a great need to educate that wider public, especially those with power and influence. Sometimes, the success as well as aspirations of the learners are lost in the arguments?

We can never win !

Assessment Moving *into* a New Future

Creating Questions

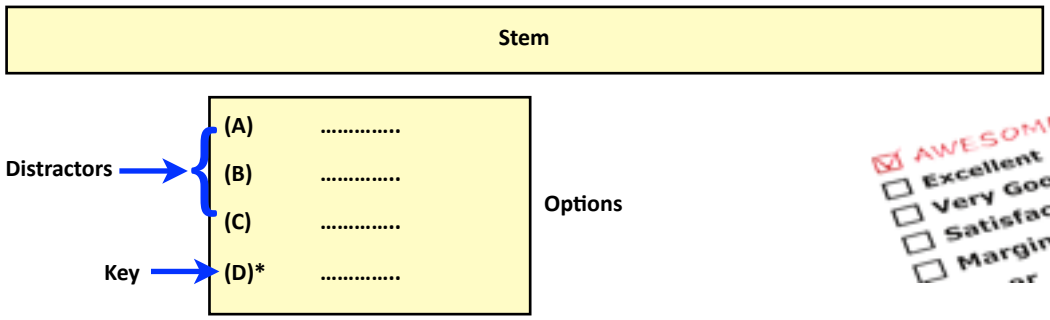
Aims

This offers some key advice in the development of quality questions in three broad areas: multiple choice questions, structural communication grids, and open-ended questions

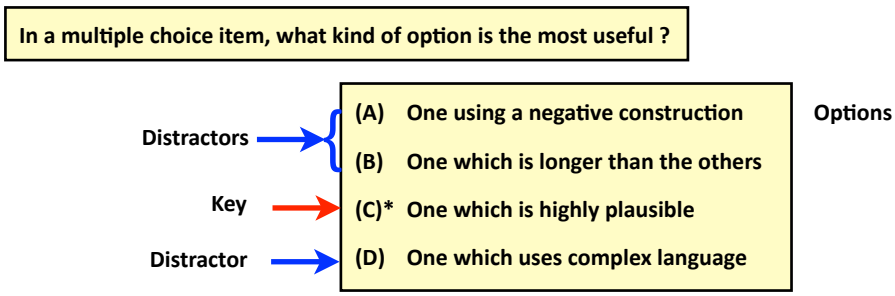
Creating Multiple Choice Items[134]

Reminder

The typical multiple choice item has the following format:



Example



[134] There is a brilliant monograph which is free online, written by a world expert in assessment. Although written for a different purpose and audience, it is full of useful practical guidance: Johnstone, A.H. (2003) Effective Practice in Objective Assessment, available at: https://www.heacademy.ac.uk/resource/effective-practice-objective-assessment

For fun, try this one……

Are multiple choice examinations an accurate measure of one's knowledge?

(A) Yes
(B) A and C
(C) A and B
(D) All of the above

Confused ??

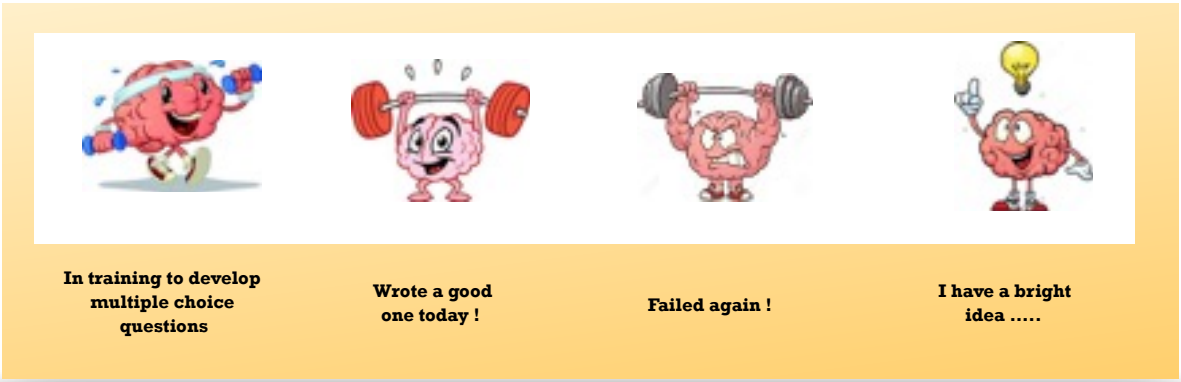
The Creation Process

Here is one good way to create, refine and edit multiple choice questions (or, better, partial knowledge multiple choice questions).

(a)	**Conception**	Always watch for ideas The best ideas often arise from actually teaching Note any potential ideas Return later to develop them into questions
(b)	**Distractors**	These must be plausible Observing the real misunderstandings that learners make in the classroom often give the pointers
(c)	**Creation**	Write out the question in full Leave it for a few days Return to it for a second consideration
(d)	**Documentation**	Note: the key the syllabus area the skill intended to be tested
(e)	**Realism**	Few can create more then a very few questions at one sitting Only 1 item of every 6 will be good enough to use in a national examination

Just occasionally, we can find someone who is capable of generating multiple choice questions quickly. However, this is very rare. For most of us, generating good ideas is hard-going. The key is to develop a mind-set. In this way, as we teach we see ideas and begin to observe the kinds of confusions that students often have. This can lead to good multiple choice questions (better partial multiple choice questions).

Because good questions are so difficult to create, it becomes important to be able to use such questions many times. This raises all kinds of security issues but the idea of a question bank has been taken up in several countries while, at university level, such banks do exist, often being built up by harnessing the skills of lecturers across many universities. The aim is to reduce work and take pressure off teachers.



Question Faults

It is very difficult to create multiple choice questions which are really good. Here are the most common faults seen in multiple choice items:

(a)	Superficiality	Items that test what is easiest to test (but not necessarily important) or merely test recall of relatively trivial items of information.
(b)	Superficial clues	There is something in the stem or options that gives the answer away.
(c)	Option variability	Options should be of similar format and length - none should stand out as markedly different.
(d)	Negatives words	Minimise use of negatives (or negatives implied in words like *'never', 'unlikely', 'smallest', 'least', 'false*') and put such words (like NOT) in capitals.
(e)	Double negatives	Never employ double negatives (like '*not untrue*').
(f)	Language	Keep the stem short and the language simple, avoid relative clauses (all critical to be fair to those whose English is less robust).
(g)	Working Memory	Watch for any demand to hold too many ideas or words in the mind simultaneously.
(h)	Stem clarity	Ensure that what is being asked of students is clear, often best to set in form of a question rather than a statement.
(i)	Option English	All options should be in the same language form and follow logically and grammatically from the stem.
(j)	Layout	Avoid items written in line. Much better to layout as illustrated below.
(k)	Language redundancy	Look for related words in options and transfer to stem to minimise words used.
(l)	Test editing	Make sure that the stem of one item does not provide the answer to another item.

Layout is everything

The green version is much easier to follow compared to the purple version:

What is the oxidation state for Manganese in the following compound?

K_2MnO_4

(A) +2
(B) +3
(C) +4
(D) +6

What is the oxidation state for Manganese in the following compound: K_2MnO_4 ?

(A) +2; (B) +3; (C) +4; (D) +6

The aim is always to enable the candidates to perform to their best and reveal their true skills and abilities. Poor layout can confuse and make things much harder.

The Shredding Process

Working with two colleagues, look at each question and ask three questions:

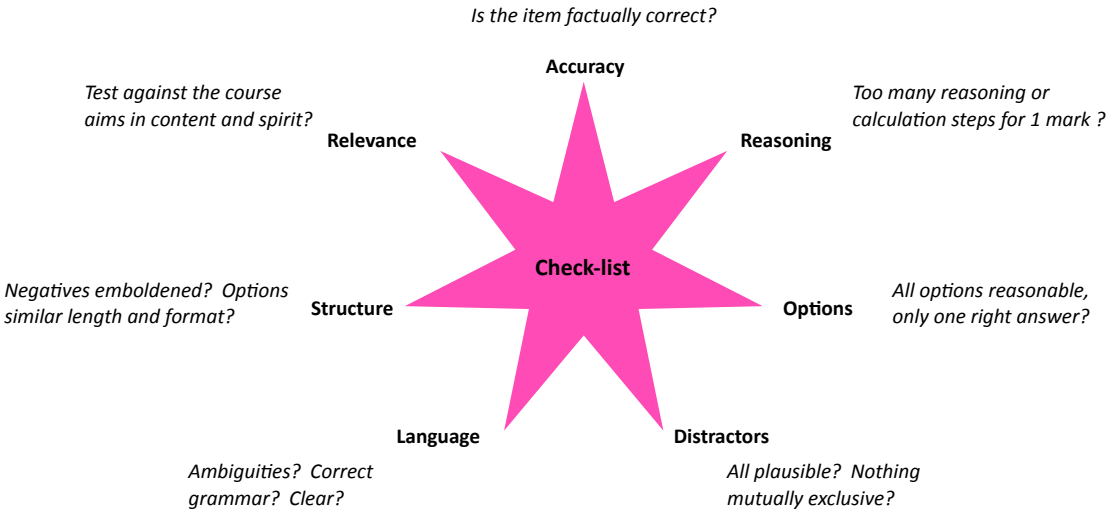
(1) What is the item testing

(2) Is it worth testing

(3) Are there any obvious faults?

The shredding process may find ambiguities, inaccuracies, technical points and faults that you, as setter, never saw. Here is an approach that works:

The check-list:

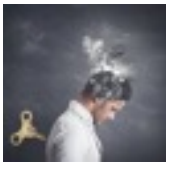


Question Banking

It may useful for the Board to develop a bank of items. Items can then be drawn randomly from this bank for the examination paper each year. It has been found that any bank must contain at least six times the number of items that are used each year. Each year, papers must be withdrawn from every candidate after the examination to ensure confidentiality. If it is not possible to guarantee security, there is no alternative but to develop fresh questions each year but this is expensive in terms of time and money.

Item Setters

The best way to set questions is to depend on practising teachers, working alongside experienced setters. For example, teachers with adequate teaching experience can be invited to submit 5 items. They are given a small fee for this and a '*bonus*' for each item that passes the shredding process. There are huge advantages in developing this approach. Setting good questions grows with experience while, for the teachers, expertise is being developed continuously and this can bring considerable benefits back into schools.




The experienced multiple choice item setter

Item Analysis

After an item has been used in a national examination paper, the following data can be gathered and this provides further evidence of the quality of the item. In a sense, this is the *'ultimate shredding*'.

Facility Value:	The proportion choosing the correct answer - if 65% select the right response, the facility value of 0.65
Functioning distractors:	Every distractor should attract at least 5% of the responses
Discrimination Index:	The correlation (the point biserial coefficient) shows the extent to which those who chose the right answer for each specific question were also those who did best in the multiple choice test overall - must be positive and above 0.2

Good items can be *'cloned*' for future use by retaining the same format and just modifying the content. This is particularly useful in some areas of the sciences and the social subjects.

Test Specification

Consider an examination paper which contains 10 multiple choice items. Here is a way to set criteria for acceptability - this is not rigid but is an exemplar of the appropriate standards.

Criterion	Required Standard
Item Value	Make sure that each item is *'worth'* one mark - a matter of agreed professional judgment
Average facility value	Average mark on test is 55% (Average FV = 0.55)
Acceptable facility values	Every items should fall in the range of about 0.3 to 0.8
Distractor quality	Every distractor attracts at least 5%
Discrimination index	Not less than 0.2
Goals measured	No more than 50% for recall
Items	Reflect curriculum emphases

Creating Structural Communication Grid Questions

Reminder

The typical structural communication grid question has the following format:

A	**B**	**C**
D	**E**	**F**
G	**H**	**I**

Select all the box(es) where
(a) .....
(b) .....
(c) .....
(d) .....

- Text, diagrams, pictures, numbers …. can all be used.
- You can ask many questions using one grid, examining many aspects of some concept or theme.

Here is an example:

1 Ammonia	**2** Methane	**3** Nitrogen dioxide
4 Hydrogen	**5** Carbon dioxide	**6** Argon
7 Neon	**8** Carbon monoxide	**9** Nitrogen

Look at the nine gases in the table above.
Select all the box(es) where there are gases which are,-

(a) Major greenhouse gases
(b) Produced in car and truck engines
(c) Produced on farms
(d) Able to burn in air
(e) Able to dissolve in human lungs causing possible health problems

Creation

Here is some guidance:

(a) Think of an area of the curriculum where there are many possibilities (easy in subjects like the sciences and the social subjects).

(b) Develop a three-by-three grid to contain the array of possible answers.

Three by three is useful for ages 16-18
However, 2 by 3 or 4 by 4 are possible

(c) Develop at least three (up to 6) questions which can be related to the grid.

(d) Write down the answers for each question: any question can have 1, 2, 3 right up to all 9 answers but try to avoid just 1 answer.

(e) Re-adjust the grid and questions to refine the question as necessary.

(f) The entire question can be marked out a many marks as you wish, depending on what you consider the demand level of the questions. It can then be scaled to suit the examination paper.

Marking

To assist human marking and make it fast, here is a way to mark the question. We shall use the question above (which explores aspects of modern day pollution) to illustrate:

1 Ammonia	**2** Methane	**3** Nitrogen dioxide
4 Hydrogen	**5** Carbon dioxide	**6** Argon
7 Neon	**8** Carbon monoxide	**9** Nitrogen

Look at the nine gases in the table above.
Select all the box(es) where there are gases which are,-

(a) Major greenhouse gases
(b) Produced in car and truck engines
(c) Produced on farms
(d) Able to burn in air
(e) Able to dissolve in human lungs causing possible health problems

What the candidate wrote

			Marking		
	Write in number of the boxes chosen	Correct Choices	Number correct	Number Wrong	Score
(a)	1,2,5,8	1, 2, 5	3	1	0.83
(b)	3,5,8	3, 5, 8	3	0	1.00
(c)	1,2	1, 2, 5	2	0	0.67
(d)	2,4,7,8	2, 4, 8	3	1	0.83
(e)	3,8	1, 3	1	1	0.36
				Total Score out of 10	7

The marker merely types in the number of correct and number of wrong answers chosen (the white boxes).
The spreadsheet then instantly calculates the score for each question and the total score out of 10 (for example).

The formula used for marking is shown on page 113. However, the spreadsheet does all the work for you. Programming the spreadsheet will take only a few minutes for anyone familiar with spreadsheets. The spreadsheet can be made available to examiners or the scores arising from various combinations of right and wrong answers for each question can be supplied to markers. The entire process, although initially unfamiliar, is very straightforward.

Question Analysis

If the response data for all candidates (or a random sample of at least 300) are entered into a spreadsheet, the entire set of questions can be analysed to show where there are patterns of gaps in understanding and areas of confusion. This is very powerful and, if the information is given back to schools, it can guide future teaching.

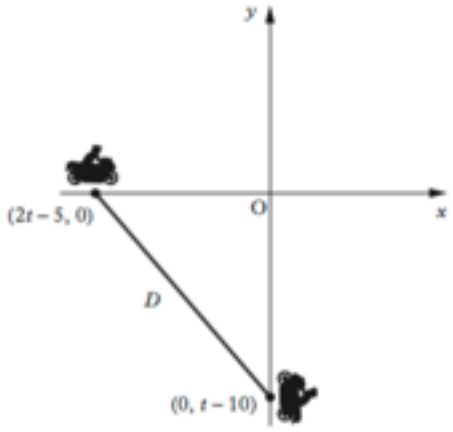
Creating Open Questions

These can range from a question with several linked parts to a question that is essentially unstructured.

Here are some examples.

Examples

A stunt performed by two members of a motorcycle display team requires them to travel, at speed, at right angles to each other across the arena.



The positions of the motorcyclists, relative to suitable axes, t seconds after the stunt begins, are $(2t - 5, 0)$ and $(0, t - 10)$.

(*a*) Show that, at any given moment, the distance, D, between them is given by

$$D = \sqrt{5t^2 - 40t + 125}.$$ **2**

(*b*) Determine whether the distance between the motorcyclists is increasing or decreasing 5 seconds after the start of the stunt. **4**

Here are other examples of more open questions, with no division into parts:

> ***You are asked to advise the Government on future policy about electricity generation. Discuss the relative merits of coal-fired power stations, hydro-electric power stations and the generation of electricity using solar power.*** **(10)**

In the context of having studied the problem associated with the use of vaccines to protect young children against pertussis (whooping cough), the following question was used:

> ***Design a poster to be used in doctor's surgeries to advise parents on the what to do about having their child immunised. The poster should show the key information in an accessible way.*** **(5)**

Following the study of the role of Alfred Nobel in Norway in his research on explosives in the 19th century, the following question was used:

> ***Alfred Nobel researched explosives and developed the safe use of dynamite. Discuss the strange fact that he left money to fund the Nobel Peace Prize, still awarded today.*** **(10)**

Having read CP Snow's essay entitled '*The Two Cultures*', a question might be:

> ***Summarise the main arguments put forward in CP Snow's essay entitled 'The Two Cultures' and show the extent to which you agree with him.*** **(10)**

Creation

Creating such questions requires a very detailed understanding of the curriculum and the goals of the curriculum. In addition, getting the standards right depends heavily on experience and this is where experienced classroom teachers have an enormous advantage.

Here is a way forward:

(a) Based on detailed knowledge of curriculum goals, select an area of importance where a question can be developed.

(b) Develop ideas and possible data for the question.

(c) Keep the length of the words you use to a reasonable level and avoid long text for questions of low value.

(d) Leave the question and return to it later for a fresh look, refining as necessary.

(e) Ensure that the instructions are clear and that there are no ambiguities

(f) Mark allocation can be related on the amount of work required to produce answers and the relative importance of what is being asked.

Marking

This requires very careful analysis and then the taking of decisions about what is being tested. The candidates need to know what such questions are testing.

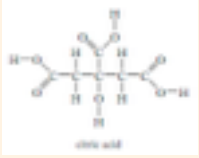
Structured questions

Many question (from about 6 marks up to 20 marks) are structured into parts. It is quite easy to see what is being tested. The marking brief must specify what is to receive credit, in ways precise enough to enable markers to mark consistently.

For example:

Sherbet is a sweet powder that fizzes on the tongue.

(a) A sherbet contains citric acid.



(i) Name the two types of functional group in the citric acid molecule.

(ii) Explain why citric acid is very soluble in water. (2)

(b) The sherbet also contains sodium hydrogencarbonate. It reacts with citric acid when water is present.

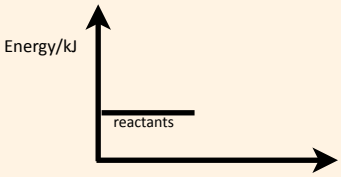
$C_6H_8O_7$ + $3NaHCO_3$ → $C_6H_5O_7Na_3$ + $3H_2O$ + $3CO_2$

(citric acid)

Explain why the reaction will only take place when water is present. (1)

(c) When the sherbet fizzes on the tongue a cold sensation is felt; the reaction is endothermic.

Complete the potential energy diagram to show the energy pathway for the reaction. (1)



(d) A sherbet is made by mixing 15 g of sodium hydrogencarbonate with excess citric acid.

$C_6H_8O_7$ + $3NaHCO_3$ → $C_6H_5O_7Na_3$ + $3H_2O$ + $3CO_2$

mass of 1 mole = 84 g

Calculate the maximum volume of carbon dioxide that would be released from this sherbet. (2)

(Take the volume of 1 mole of carbon dioxide to be 24 litres)

The danger is to award marks for recall. Looking at the above chemistry question, there has been a real attempt to move beyond recall. Citric acid is totally unfamiliar to the candidates. They have to look for structural features that might explain solubility and they are directed to functional groups as a starting point. They have to grasp that molecules (strictly species) must collide to react for part (b). Part (c) requires them to interpret the concept of endothermicity into an energy diagram while the final part is a familiar calculation although using unfamiliar materials. In passing, this question was used in a national examination. Marking is fairly straightforward.

The greater difficulties come when marking longer questions which have no structure - marking essays.

Essay Marking

Look at the question:

> ***Summarise the main arguments put forward in CP Snow's essay entitled 'The Two Cultures' and show the extent to which you agree with him.*** **(10)**

Suppose marks were awarded as follows:

Evidence that ….	**Marks**
The candidate has read the essay	2
The candidate has understood the key point CP Snow was making	2
The candidate has expressed a personal viewpoint	2
The candidate has justified their viewpoint	2
The candidate has written clearly, logically and correctly	2
Total	**10**

This focusses on specific skills that might be important in the context of the course being taught. 2 marks might be awarded for an adequate response, 1 for partial, 0 for no evidence.

In looking at essay marking, the essential point is to specify the specific skills that are to be rewarded. It may be necessary to tell the candidates what is expected. For example, the question might state that *'Marks will be awarded for clarity of argument and for clear and logical writing*'. Pages 117 and 118 offer examples of the kinds of skills that more extended writing can assess.

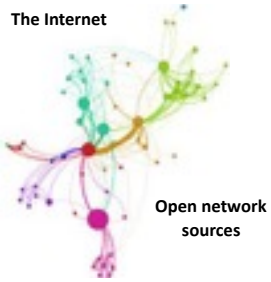
Gathering Ideas

Good assessment arises naturally from experienced teaching. As we teach, we start to gain more and more insight into the ways the learners are developing ideas, the kinds of confusions that arise, the kinds of issues that are important. This experience can feed into the development of quality assessment questions and tasks. There are two key requirements:

Simply knowing and understanding the relevant subject discipline is not enough. Knowing the learner and how the learner reaches understanding is also required. Indeed, the two requirements each feed into the other, enriching both. In a sense, teachers are learning *with* their students as, together, they seek to gain insights and understandings that are meaningful and useful.

However, there is another way of developing good ideas for assessment questions and tasks. We can look at the ideas of others. In one workshop on assessment, the participants were asked to shred questions derived from two sources: one set of questions came from test papers used in international comparisons, the other from schools. The questions were mixed randomly but the workshop participants, with unerring accuracy, criticised the questions from international comparisons while applauding those that had bene taken from school examination papers. This illustrates an important point: not all sources of questions are equally good!

The internet gives access to endless information, insights and examples. While not suggesting that everything is in any perfect, the following site is recommended in that interesting examples of questions can be found[135]. In this, it is important to stress that searching the internet is to provide ideas. It is not sensible simply to copy questions devised for one curriculum in one country and then use them in another curriculum in another country. Curricula vary, not simply in terms of content, but emphasis, setting, and assessment ethos. We can take advantage of the internet to gain ideas, not to provide neat answers!



There is yet another approach. In most subject areas, there are good magazines and journals available. In the same way as medical doctors keep up to date on new procedures, drugs and understandings by means of high quality journals and specialist magazines, other subject disciplines have generated similar publications. Many are produced by learned societies or organisations that specialise in support for practising teachers. These often give all kinds of new ideas which can lead to excellent assessment questions and tasks.

[135] http://www.sqa.org.uk/pastpapers/findpastpaper.htm
Select the subject you wish and we suggest you select the Higher Grade as the qualification level

Assessment Moving *into* a New Future

Marking and Marks

Aims

This chapter looks at some of the practicalities of marking, how to interpret the data, presenting the data for those who need it as well as the way data can be used to bring future benefits.

Introduction

One of the most tiring and demanding tasks of all is the marking of examination scripts. It is repetitive, boring, and sometimes creates disillusionment as we see what our students did *not* manage to achieve! It is easy to lose concentration and for inaccuracies to creep in. However, the marks we give and the grades that are awarded are of vital importance to the learners for they may determine their future opportunities, they may close entry to colleges or universities or they may change career possibilities.

Marking Practicalities

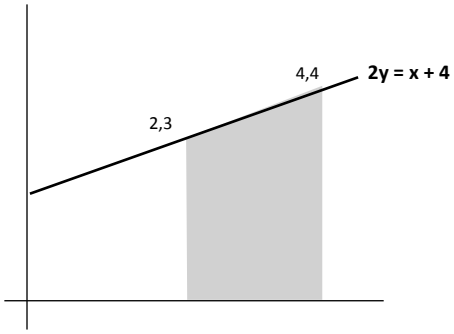
Here are some practical hints on the best way forward, especially if we are marking scripts for the examination board:

- Never try to mark for too long at one sitting.
- Take a short break to walk about, drink a cup of coffee or do something very different.
- Never mark when you are exhausted.
- Follow the agreed marking brief precisely.
- Total the marks for each long question and then check your total.
- Total the marks for the entire paper and then check your total.
- When all the marking is completed, re-mark the first paper you marked to check standards have not changed with time.

Following Marking Schemes

Examination setters, in collaboration with others, must develop clear marking schemes. The aim is to make them precise, unambiguous but not so prescriptive that it removes professional judgements. A practical way forward is to indicate how every mark is allocated. It is essential that markers follow marking briefs exactly.

In a subject like mathematics, suppose the students are given the following diagram and asked to find the area that is shaded.



It is possible to allocate 1 mark for dividing the area into a triangle and a rectangle, 1 mark for the area of the rectangle, and 2 marks for the area of the triangle, 1 mark of the answer. That is very precise, giving a mark out of 5.

However, if we ask a question in a social studies paper like:

> *Outline two voting systems that use proportional representation and discuss their advantages and disadvantages when compared to a first-past-the-post system of voting.*

We might have to award marks as follows:

Description of at least two proportional systems:	4 marks total
For each system outlined	
If totally clear - 2 marks	
If partiality clear - 1 marks	
For evidence that the candidate understands a first-past-the-post system of voting	1 mark total
For clear specification of relative merits:	4 marks total
For each system outlined	
2 advantages stated - 2 marks	
1 advantage stated - 1 mark	
For one disadvantage of all proportional systems - 1 mark	1 mark total
Evidence of clarity of critical thought	2 marks total
2 marks for weighing the evidence of advantages/disadvantages with care	
1 mark for some attempt to weigh evidence	
Question total mark	**12 marks**

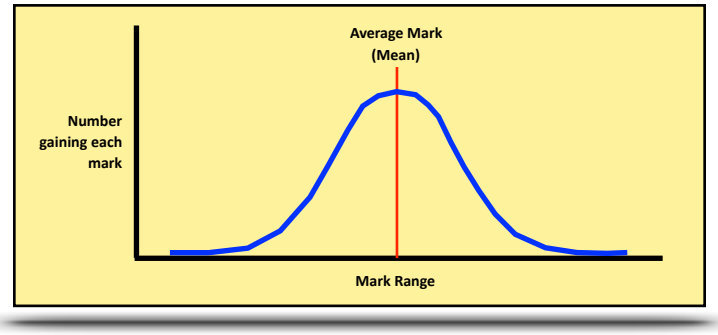
This illustrates how a marking brief in some subjects can look very different when compared to a marking brief in another subject. It is essential that the marks are allocated for the skills laid down in the course objectives. it is also important to try to allocate marks in such a way that the value of a mark in one question is approximately similar to that of other questions.

Marks and Grades

The performance of our students in their examinations needs to be communicated to the students and to many other groups. In deciding how to communicate, the key thing is to think of what will convey greatest meaning to most of the groups.

For any kind of national assessment, where there are large numbers, the most common approach is to use the *entire marks of ALL candidates* to determine the meaning of specific marks for an individual student. This also allows us to ensure standards remain constant.

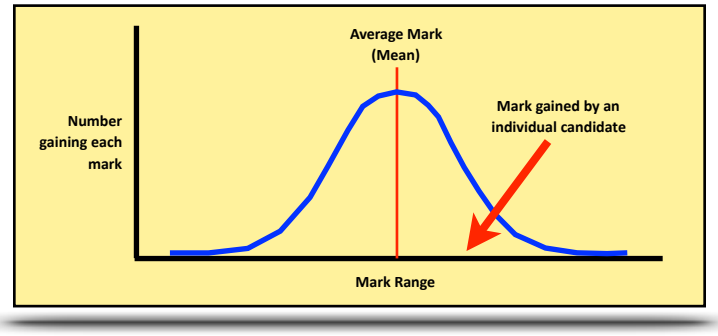
On page 120-121, we discussed standardisation and this is the best way forward (provided the number of candidates is not very small). For all the candidates put together, you will always obtain a pattern of marks like:



Remember: the actual make gained means ABSOLUTELY NOTHING. High marks will arise from easy examinations and low marks from more demanding examinations. Standardisation (perhaps using a mean of 60, standard deviation of 12) is essential. On pages 126-131, the process of standardisation is explained in detail. What it does is that it gives marks ***meaning*** and the meaning is the same in ***every*** examination in ***every*** subject.

Standardisation gives marks meaning and the meaning is the same in every examination in every subject

Then, the mark gained by any individual candidate can now be seen in relation to this total spread of marks:



There are two key systems - using standardised marks or using grades:

Standardised Marks

If we use a mean of 60 and a standard deviation of 12, it means that about 70% of our candidates will gain a standardised mark of 50% or over.

Suppose that our candidate gains a standardised mark of 70%. This means that the candidate is just within the top 20% of all the candidates. If she gains 70% in Physics and 70% in Urdu, it means that she has performed ***exactly equally*** well in both. This gives marks meaning. If we do not standardise the marks, then a 70% in Physics will almost certainly not hold the same meaning as a 70% in Urdu. Remember, the aim is provide the ***greatest*** meaning for the candidates and other users (parents, potential employers, university entrance controllers).

The mark now has meaning

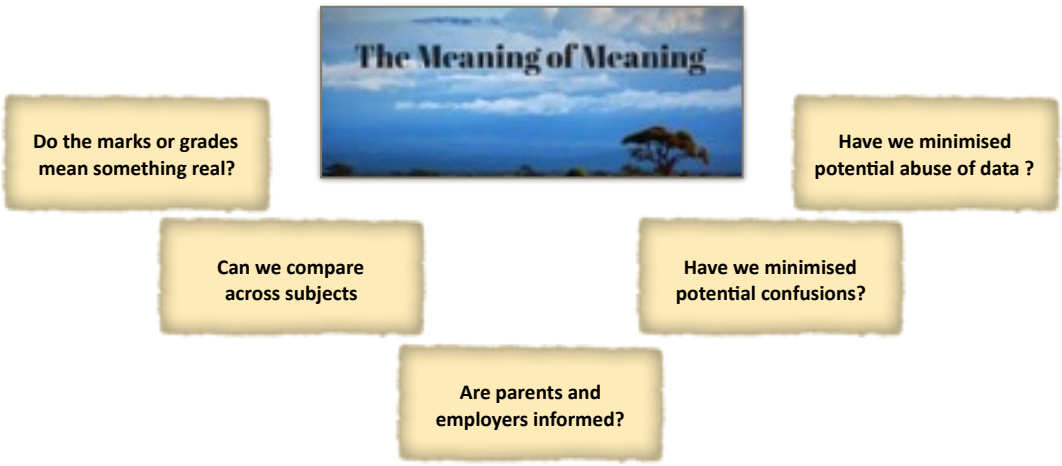
It also means that marks in different subjects can be added together to a total score overall. This has some advantages - maybe for university entrance - but it is slightly risky in that specific subjects areas of study at university level require very strong abilities in specific subject area at school level, while other subject areas are less important.

Grades

In many ways, converting the standardised marks into grades is even better in terms of parents, the wider public and the media. It gives clear understanding of meaning. On page 138, we suggested that 70% of the candidates might be given a pass. Here is another way forward, this time with 80% of the candidates being given a pass:

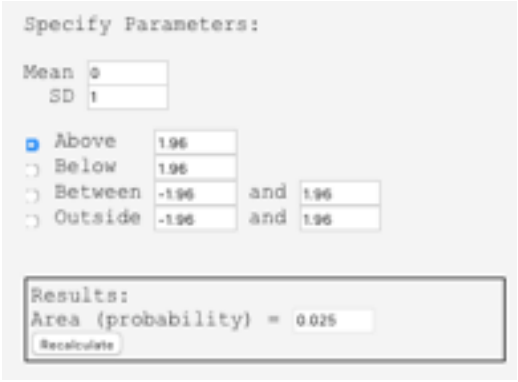
Award of 'A' pass	5%
Award of 'B' pass	10%
Award of 'C 'pass	15%
Award of 'D' pass	30%
Award of 'E' pass	20%
Award of 'F' fail	20%

In all of this, the aim is make marks mean something for people. We want to minimise confusion. We want to hinder the abuse of examination board data. We want to inform candidates, their parents and other key groups clearly and unambiguously about the success of the students.



Standardised Marks and Grade Awards

We can convert standardised marks into grades (like A,B,C etc) using a simple calculator which is online[136]. Under the normal distribution curve on the screen, you will find a layout like this:

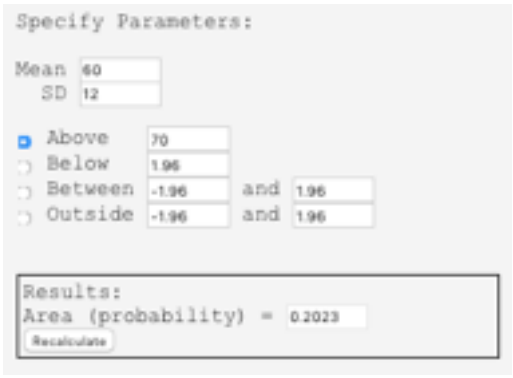


Change the mean to 60 and the Standard Deviation (SD) to 12.

In the box after the word '*Above*', type in 70.

Press the box '*Recalculate*'.

You will get a layout like:



Look at the box near the foot after the words '*Area (probability)*'. The value here is 0.2023.

This means that 20.23% of the candidates score a mark from 70% upwards.

You can type in various values into the box after the word 'Above'.

Each time press '*Recalculate*' and look at the percentage of the students who scored upwards from that mark upwards.

From this you can calculate the mark needed to be awarded various grades:

	Grade	% gaining grade[2]	Approximate Mark range
Pass	A	5	80 and above
	B	10	73-79
	C	15	67-72
	D	30	57-66
	E	20	50-56
Fail	F	20	less than 50

[136] http://onlinestatbook.com/2/calculators/normal_dist.html

[137] The percentage gaining each grade is a Board decision - this is given as an illustration only.

Communicating

The central goal is clarity and the avoidance of ambiguity. Both can be achieved using standardardised marks or the lettered grades derived from these. This is what to avoid:

I do not understand what it means .....

Thus, the certificate given to the students requires the minimum of information. Here are exemplars of the key information:

Federal Board of Intermediate and Secondary Education

Candidate Name:
Examination Level:
Year:
Grades Awarded:

Urdu	75
English	73
Mathematics	84
Social Subjects	69
Chemistry	82
Physics	86
Biology	76

OR

Federal Board of Intermediate and Secondary Education

Candidate Name:
Examination Level:
Year:
Grades Awarded:

Urdu	B
English	B
Mathematics	A
Social Subjects	C
Chemistry	A
Physics	A
Biology	B

It is also important to provide summaries of information, looking at the entire population. Given that marks will be standardised and the number of passes at various grades will vary very little from year to year, the important aspect is to give verbal impressions from examiners which will inform schools and teachers for the future. For example, here is an imaginary examiners report for, say, Chemistry:

Examiner's Report: Subject: Chemistry

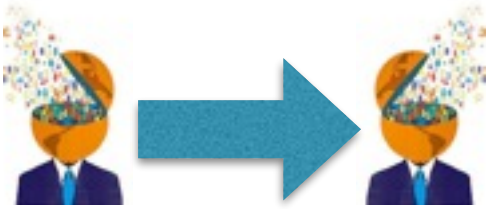
Students tended to perform well in questions that tested in reaction kinetics but were not so confident with thermodynamics where, at times, there was a lack of clarity that they fully understood some of the key concepts. In organic chemistry, students performed well in routine questions but lacked depth of understanding when considering areas which involved stereochemical insights. Perceptions of three-dimensionality were poor. Redox questions were handled poorly but metal reactivity, and the chemistry of main group elements was well understood. Students performed particularly well in the social implications arising from developments in chemistry.

This kind of report would be compiled by the chief examiner in chemistry and would be based on reports from individual examiners as well an analysis of the examination performance data. Such a report is very useful for teachers and schools, enabling teachers to see areas where students are finding difficulties so that steps can be taken to offer help in future.

A Deeper Thought

There is anther important outcome that might arise. These reports provide an endless supply of potential research projects for university education schools and faculties. Such projects (ranging from a short Master project right up to a full PhD) could focus on areas where students find difficulties, seeking to identify the source of the problems and then develop new approaches to offer positive assistance for the future. It is this kind of project which then offers the teaching profession and, more importantly, future students potentially very useful findings that might move Pakistan education forward positively[138].

The data gathered by the Examination Board over several years provides quite remarkable insights into learning in schools in Pakistan. Analysing the data can reveal areas of student successes and confusions, differences between different subjects in their emphases, the developing range of skills being tested, the predictive value of examination data and long-term curriculum trends. The data are also a gold-mine for statisticians !



An endless pile of data ? **An endless pile of useful data**

With the modern use of spreadsheets and powerful computing resources that are readily available, research studies will offer major insights that can guide education forward and enrich the learning experiences for future generations.

'Thinking is hard work - that's why so few do it'
Albert Einstein

[138] An interesting example of what is possible can be found by tracking through the research publications of Professor Alex H Johnstone who, over four decades from about 1968, supervised Masters and PhD students who undertook work of immense value and insight. The internet now gives access to this.

Assessment Moving *into* a New Future

Some Practical Issues

Aims

This chapter seeks to summarise ways to approach some of the key issues that are not often to easy answers and to consider tested ways to carry out practical procedures

The Overall Process

The overall process can be seen in figure 18.1 and the process is amplified under the five broad areas:

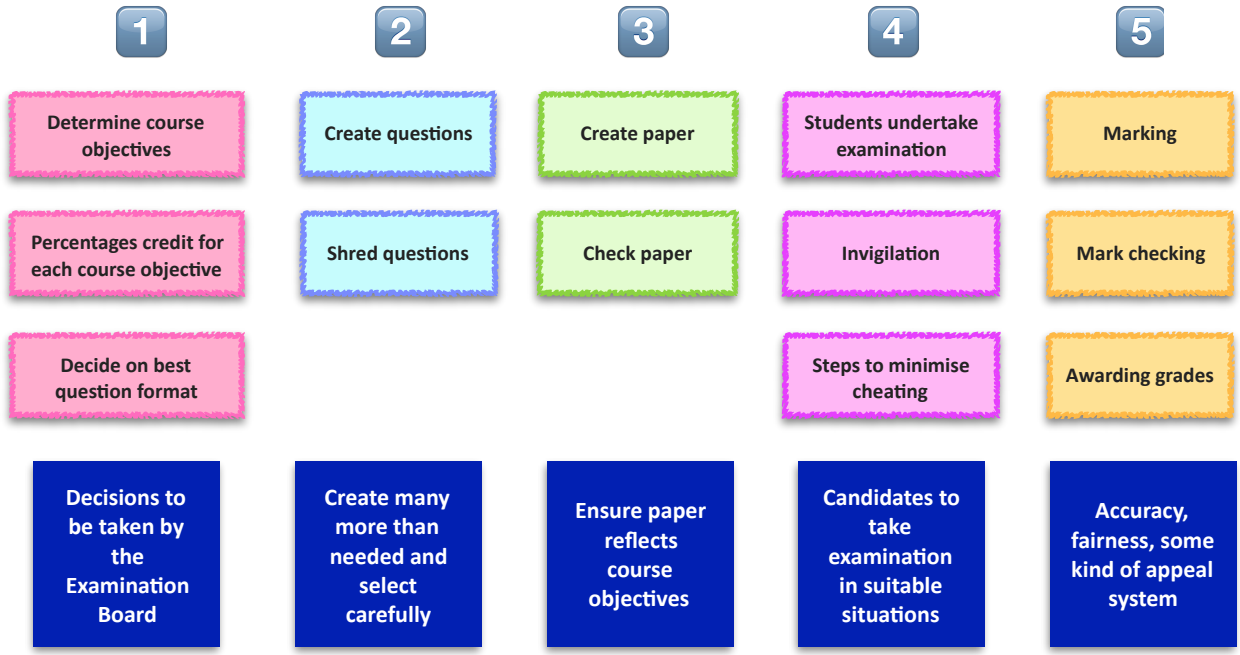


Figure 18.1 The Overall Process

1 Chapter 15 outlined some of the key decisions that need to be taken at Examination Board level. It is important to stress that genuine consultation must be integral to the entire process. Teachers require to be able to work *within* the system while universities and employers need to use the outcomes *from* the system. However, the Board must offer decisive leadership, enabling every group to move forward into the best ways for the future, and not remain wedded to the practices of the past.

2 Chapter 18 offered practical advice about question creation. This considered multiple choice and the improved variant, partial multiple choice, items as well as structural communication grids and their power to test for conceptual understanding. The development of open-ended questions was also outlined, as well as the important process of question shredding. We shall discuss in-school assessment later in more detail in this chapter.

3 The way an examination paper might be put together was discussed, using the subject geography as an example (see pages 158-161). The essential aspect of this process is the checking that is needed to ensure that the paper, as a whole, is testing against the balance of objectives that reflect the curriculum and the Board policy.

4 There are many steps involved in enabling the candidates to sit the examinations under fair conditions. Papers have to be distributed in advance, without compromising security. Invigilators need appointed and trained as necessary. The locations for the examinations need to be set up while there need to clear 'rules' about what a candidate can take into the examination area, to minimise cheating.

5 Marking is a tiring and repetitive task but requires constant attention to ensure standards are maintained. The key role of a good marking brief was stressed on pages 99-100. Nonetheless, checks need to be made at every stage to ensure that standards are maintained throughout. Decisions need to be taken about whether to use marks or grades and the importance of mark standardisation across all subjects was stressed: how to carry this out was outlined on pages 120-121, 129-131.

In-School Assessment

All the research evidence shows very clearly that teachers are highly capable of assessing their own students. Indeed, the best national assessment procedures rely heavily on teacher assessment while combining teacher assessment with formal end-of-course examinations is often found to give the most reliable outcomes. Depending on a one-off assessment (such as an examination paper at the end of the course) is not a very reliable way to award credit to learners. Using multiple assessments immediately increases the reliability very markedly. In the light of all the very clear evidence, it was recommended that part of every subject examination should come from internal assessment. Inevitably, there is a natural hesitation in introducing such a policy and it was suggested that, at the outset, only 10% of the credit for any course should be based on internal school-based assessments. How this might work was exemplified on page 154 and the table is repeated here:

Subject	Objective Formats	Non-Objective Formats	In-school Credit
Mathematics		Short answer: 50% Longer answer: 40%	Mathematics applications: 5% Project: 5%
Languages	Multiple choice or partial knowledge multiple choice: 5%	Short answer: 25% Essays or longer questions: 60%	Technical essay: 5% Communication skills: 5%
Social Subjects	Multiple choice or partial knowledge multiple choice: 10%	Short answer: 30% Essays or longer questions: 50%	Essay: 5% Project: 5%
Sciences	Multiple choice or partial knowledge multiple choice: 10% Structural Communication grids: 5%	Short answer: 50% Longer answer: 20% Open book: 5%	Laboratory work: 5% Group work units: 5%

The final column is the area which now needs discussion. What is envisaged is now amplified.

Formats for in-School Assessment

In subjects like social subjects and languages, it was suggested that an essay would give 5% of the total credit. Over the course, the students might be required to write several essays on topics and themes relevant to the course. The themes might be laid down for social subjects but might be more open to student choice in the languages where the topics could be chosen by the students (with approval for teachers) and be on themes of direct interest to individual students, perhaps reflecting their preferred career direction.

In both cases, a portfolio would be developed and this might contain 3 or 4 essays written by the student. The work could be carried out in class to avoid any help from outsiders. The student would then select the essay they think is the best and this would be formally marked to give the credit (5% of the course only). The essay might offer evidence of skills like ability to communicate clearly, ability to conduct a sustained argument, evidence of logical thought, evidence of creative thought (languages maybe), ability to see social significance (social subjects) and not for content. Many of these skills are more difficult to assess in formal end-of-course examinations.

Projects and dissertations can also be employed. Here students, working on their own, carry out the reading on some theme of interest to them. There is always the possibility that help could be provided by adult friends or parents. This might be avoided by draft write-ups being compiled in class time, with students allowed to bring in any notes or resources that they wish. The project reports (or dissertations) would be short but offer evidence of scope of reading, ability to sift information and argument, evidence of critical thought, valid selection of information, coherence of thought, clarity of writing. Pages 117-118 and 176 offer suggestions about criteria to govern the way such work is marked.

In mathematics, the internal assessment is based on mathematics applications and a project. For the first, the students might be invited to take a topic or theme and show, based on their outside reading, something of the significance of this mathematical insight on societies and the way the world is described. For example, the use of letters for counting (the normal practice in the Mediterranean and Middle Eastern cultures) and the impact of the use of numerals (imported from the Indian sub-continent to the Middle East and then to the West), the invention of the ideas of zero, infinity, logarithms and exponentiation, imaginary numbers, pi, trigonometry and so on. The project might be to show how mathematical insights have been applied in life in solving problems of importance in ordinary living: for example, the way mathematical ideas underpin modern medicine, kinetics, astronomy, social statistics and so on. This would involve writing a report and this could be carried out in class time, with students allowed to bring in any notes or resources that they wish.

Laboratory work underpins the sciences but actually conducting experiments is not an important skill in that very few students will ever use any experimental skills later in life. Here is the ideal place for a '*duly-performed*' assessment. For example, students might be required to have completed 10 experiments (or whatever number is thought appropriate) and submit very brief reports. These reports are not marked but it is noted if an adequate report is submitted. Where the student completes all 10, with adequate reports, they are given the credit. Where less than the 10 were completed satisfactorily, the marks gained would be reduced. Where resources are limited, a very useful idea was developed in Pakistan several years ago that can assist[139].

[139]Reid, N and Shah, I. (2010) The Idea of the Paper Laboratory, *Journal of Science Education*, 11(10), 8-12.

The other suggestion related to what are known as '*group-work units*'. Sets of these now exist[140] but more need to be developed. Here the requirement is that, over the course, the student has to complete, say, 2, 3 or 4 of such units. The satisfactory completion of the specified number brings the credit.

It is known from research that the participation in such units encourages the development of very important skills (like critical thinking, team-working, summarising ideas, taking decisions based on evidence, compromising, communicating clearly), all such skills being highly valued in modern society[141].

There are important issues that need addressed in undertaking in-school assessment like those described above.

(a) ***Cheating***

The general principles have been discussed on pages 162-3. However, there is the great fear of cheating occurring in introducing in-school assessments. It can arise in several ways. Firstly students can seek help from adults and parents. Secondly, parents can apply unfair pressures on teachers to award higher grades. Thirdly, teachers can give help in order to improve school examination results or even show favouritism towards specific students.

These fears can be addressed. The total proportion of credit for the national examinations is only 10%. However, these assessments are, by their very nature, placing a value on very important education outcomes and student skills which formal end-of-course examinations cannot reward easily. By arranging that the actual essay or report writing is conducted *within* the schools, outside help is minimised. Students can bring in whatever resources they wish but this gives the teacher an opportunity to observe a student who is merely copying out notes written by an adult in advance.

Schools could set clear guidelines (following general principles laid down by the Board). For example, any evidence of unfair outside help would mean the student was awarded a zero for that element of the assessment. Any parent placing pressure on a teacher would immediately mean that the student's maximum mark was reduced to 50%. Any evidence of plagiarism (easy to spot by an experienced teacher and there is useful software to confirm) would mean that the maximum mark was reduced to 50%…. and so on.

Finally, if school examinations results are never used to grade schools and their teachers, then there is no temptation to manipulate grades to raise reputations. The folly of using school examination data in this way has been emphasised in this monograph (see chapter 4), the research evidence showing incontrovertibly that such approaches tend to lower school standards rather than raise them

---

[140] A set of 16 units in chemistry was developed in the 1970s, several units exist in physics while there is a range of units for biology: details can be obtained from one of the authors: dr_n@btinternet.com
There is also a set of short units in a textbook: Johnstone, A.H., Morrison, T.I. and Reid, N (1980) *Chemistry about us*, London, Heinemann. Old copies of this out-of-print book are readily available.

[141] The evidence is scattered widely in the literature and some is old. Much derives for psychological understandings related to attitude development:
Reid, N. (1976) Simulations, Games and Case Studies, *Education in Chemistry,* 13, 82-83.
Reid, N. (1982) The Time Dimension in Chemistry, *Education in Chemistry,* 19(6), 166-168.
Clarkeburn, H., Downie, R., and Reid, N. (2000) Teaching Biology Students Transferable Skills, *Journal Biology Education,* 34(3), 133-137.
Reid, N. and Yang, M-J. (2002) Open-ended Problem Solving in School Chemistry: A Preliminary Investigation, *International Journal Science Education,* 24(12), 1313-1332.
Chu, Y-C and Reid, N. (2012) Genetics at school level: addressing the difficulties, *Research in Science and Technological Education,* 31(1), 1-25.

(b) Standards

The central issue is how we ensure that teachers make in-school assessment to the right standards. Two influences are important. In developed countries, where in-school assessments as part of formal national assessment are common, the finding is that teachers often tend to mark *very strictly*, setting standards that are close to what is needed but sometimes even more demanding. In Pakistan, by contrast, what little evidence exists suggests the reverse, with teachers setting their own standards below that which is required. It appears that this is caused simply by lack of experience.

There are numerous ways to established correct standards and here are some practical suggestions:

It is possible to ensure standards statistically. On the basis of the overall performance of students in a school over the previous 3 years (to allow for year-by-year variations), schools are permitted to award a given proportion of marks. Thus, for an assessment out of 10 in a specific subject, a school might be allowed to award 5% at 9 or 10, 45% at 7 or 8, 40% at 5 or 6, and the remainder less than 5. The allocations would be reviewed every few years. This is rigorous but it does require the Board to analyse overall school results to lay down the awards permitted. However, with modern technologies (even simple spreadsheets), this is not onerous, once set up.

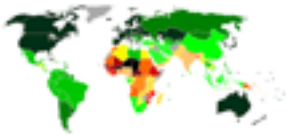
One way that is widely used in many countries is the moderation of marks by teachers from one school looking at samples of essays, projects or reports from another school and suggesting upgrades or downgrades. This is done subject by subject. Teachers (known as moderators when fulfilling this task) would need trained for this but this has been found to be undemanding. Once a cadre of teachers is trained, repeated training is not necessary and there is a huge benefit for the teachers in gaining valuable insights into what is good education. Where serious disagreement exists between the moderator and the school, the matter can be referred to the Board. However, experience elsewhere indicates that this system works extremely well and the time demand is not excessive.

Another way is for sample essays, projects or reports being submitted to the Board for re-marking. Where serious discrepancies are found between the school marking and the Board marking, all the essays, projects or reports from that school in that subject can be re-marked. For '*duly performed*' assessments, the school would be required to give evidence that the work had actually been completed, the Board checking this randomly across schools for time to time.

All of these approaches could be made to work fairly easily and most have been widely used elsewhere in the world. Studies often show the deep commitment of teachers to do the best for their students. Given support, affirmation in wider society and the awareness that the '*system*' is one of integrity, teachers will move forward and the education provision in the country will naturally be enriched. All the research shows that these are the best ways forward while the writing of endless policy documents by those outside schools and the imposition of regulations will usually have no positive impact. The key words are *integrity* and *affirmation*.

Never apologise for setting high standards
Given support, people will rise to them

End piece



The overall aim of this monograph has been to bring together the enormous range of worldwide evidence from carefully conducted research related to assessment in order to establish the key principles for moving forward in Pakistan. This has then been interpreted into the HSSC system of examinations to show how, without enormous upheavals, the current system can be adapted and adjusted to move assessment forward. The end goal is to develop a national assessment system that matches the best in the world. The best systems combine width of study with rigour. The HSSC system has the essential structure to be adapted easily to fulfil such goals.

The analysis has aimed to take into account the specific issues and challenges faced by Pakistan. It is recognised that any idea of an overnight transformation is simply asking the impossible. However, what has been outlined gives a clear agenda for the next five years, with the recognition that adaptation and development are on-going. What has been suggested is known to be practical and workable.

Endless lecturing and endless writing

The central problem with the current examination system is that it rewards recall of memorised information and skills, ignoring the wide range of much more important educational outcomes. These wider outcomes have been translated into seven areas and it has been shown how assessment in each of these areas can be undertaken successfully, based on worldwide evidence. The monograph does not address the issues of updating and modernising current subject curricula, especially the need to prune out some of the content that is no longer relevant in a modern society. However, by focussing the assessment on to the skills which are more important, these more irrelevant areas may fade in importance over time. The monograph also does not address decisions about curriculum directions in schools (eg arts, sciences) nor whether the national assessment should adopted a Baccalaureate structure. There are strong arguments for a school curriculum that minimises any element of specialisation while some of the most rigorous world systems are, indeed, Baccalaureate in nature.

The expanding brain

The monograph starts with a strong view that experienced classroom teachers are central to the entire process and research evidence supports this strongly. Thus, in introducing developments, experienced practising teachers MUST be involved and the entire teaching profession (of those teaching at the relevant senior stages) must be consulted and their input taken very seriously. Developments will *only* work successfully if the practitioners are consulted, supported and affirmed. Again, much evidence supports this. There is a wealth untapped talent to be considered in schools.

One of the features in the proposed way forward is that, for the moment, national assessment must depend on external end-of-course examinations for 90% of the credit awarded. However, given the overwhelming evidence that in-school teacher assessments can be highly reliable, the introduction of 10% of the credit for in-school assessment is a first stage in development. This gives a wonderful opportunity for teacher development and for a general enrichment of the school provision while allowing important skills that cannot be assessed easily in end-of course written papers to be rewarded effectively. With only 10% of the credit being awarded in this way, there is ample opportunity to refine the system without any student being treated unfairly.

Research evidence does not support the idea that improving education holds the key to generate economic progress, national progress or social happiness, simply because other factors in any society tend to be much more powerful. However, in the world today, developing education is a central demand in that Pakistan society at every level has to compete on a world stage. While those who have enjoyed the best educational provision in the country have moved to high levels, the benefits of education have not been spread evenly or equally across society, leaving some at a major disadvantage.

A good examination system can make a large contribution here. By focussing on the skills that are important for the 21st century and ensuring the highest standards, the population can compete on the world stage. However, it is also important that the benefits of education are equally available to every young person, irrespective of parental wealth, geographical location, gender or religion.

Thus, much of the future does depend in some measure on education. In this, the nature and role of assessment plays a critical role. The backwash of national assessment to influence curricula, teaching and the developing emphases of education cannot be underestimated. In simple terms:

Assessment
has a dominant influence
on
an entire education system

The opportunity for an Examination Board to make a powerful and positive impact are enormous. The challenge in doing this is considerable.

The future
starts
here

Reference List

AfiL, (2004) *Assessment is for Learning Information Sheet*. http://www.scotland.gov.uk/Resource/Doc/69582/0017827.pdf

Al-Ahmadi, F. and Reid, N (2011) Scientific Thinking - What is it and can it be measured? *Journal of Science Education*, 11(2), 53-59.

Al-Ahmadi, F. and Reid, N (2012) Scientific Thinking - *Can it be Taught? Journal of Science Education*, 13(1), 18-23.

Al-Ahmadi, F.M.A. (2008) *The Development of Scientific Thinking with Senior School Physics Students,* PhD Thesis, University of Glasgow, Glasgow. [http://theses.gla.ac.uk/241/]

Al-Osaimi, K.H. (2013) *The development of critical thinking skills in the sciences*, PhD Thesis, University of Dundee, Dundee. [http://discovery.dundee.ac.uk/portal/en/theses/the-development-of-critical-thinking-skillsin-the-sciences%28f8eb450c-0914-4e1b-aa1c-65dd9d02cfba%29.html]

Al-Osaimi, K.H., Reid, N., Rodrigues, S. (2014a) Critical Thinking - Can it be Developed? *Journal of Science Education*, 15(2), 57-75.

Al-Osaimi, K.H., Reid, N., Rodrigues, S. (2014b) Critical Thinking - Can it be Measured? *Journal of Science Education*, 15(1), 30-36.

Albaz, K. S. (2001) Effectiveness of the Use of Marzano's Model for Learning Dimensions in Teaching Chemistry on Achievement and Compound Thinking and Aptitude towards the Subject for the 1st Secondary Class in Bahrain, *Procedure of the 5th Scientific Conference*, Vol. 2, Egyptian Education Association, Cairo: Ain Shams University, pp. 413-448

Ali, A. A. (2008) *Perceptions, Difficulties and Working Memory Capacity Related to Mathematics Performance,* MSc Thesis, University of Glasgow, Glasgow. [http://theses.gla.ac.uk/441/]

Ali, A.A. and Reid, N. (2012) Understanding Mathematics Some Key Factors, *European Journal of Educational Research*, 1(3), 283-299.

Almadani, K., Reid, N. and Rodrigues, S. (2012) What examinations test, *Problems of Education in the 21st century*, 1(1), 6-19.

Ambusaidi, A.K.A. (2000) *An investigation into fixed response questions in science at secondary and tertiary levels*, PhD Thesis, University of Glasgow, Glasgow.

Anastasi, A. (1988). *Psychological Testing.* New York, New York: MacMillan Publishing Company.

Are school league tables good or bad? UKEssays.com [http://www.ukessays.com/essays/education/are-school-league-tables-good-or-bad-education-essay.php [For a very brief summary of key landmarks and comments]

Ashcraft, M. H. (1994) *Human memory and cognition*. New York, Harper Collins College.

Assessment Reform Group, (1999) *Assessment for learning*: Beyond the Black Box. University of Cambridge.

ASTI Briefing (undated) *League Tables*, Association of Secondary Teachers Ireland, Dublin. [http://www.asti.ie/about-asti/policy/league-tables/]

Atkinson, R. and Shiffrin, R. (1971) The Control of Short-Term Memory. *Scientific American*, 225, 82-90.

Ausubel, D.P. and Robinson, F. G. (1969). *School Learning: an Introduction to Educational Psychology*. Holt, Rinehart and Winston, New York.

Ausubel, D.P., Novak, J. D. and Hanesian, H. (1978) *Educational psychology: a cognitive view*. 2nd Ed. Rinehart and Winston, New York.

Baddeley, A.D. (1986) *A Working Memory*, London, Oxford University Press.

Baddeley, A.D. (1994) The Magical Number Seven: Still Magic after all these Years? *Psychological Bulletin*, 101(2), 353-356.

Baddeley, A.D. (2002) Is Working Memory Still Working, *European Psychologist,* 7(2), 85-97.

Baddeley, A.D., & Logie, R. (1999) Working memory: The multiple component model, in Miyake, A., & Shah, P. (Eds.), *Models of Working Memory* (pp. 28- 61) (New York: Cambridge University Press)

Bahar, M. (1999) Investigation of Biology Students' Cognitive Structure through Word Associated Tests, Mind Maps, and Structural Communication Grids, PhD Thesis, University of Glasgow.

Bahar, M. and Hansell, M. H. (2000) The relationship between some psychological factors and their effect on the performance of grid questions and word association tests *Educational Psychology*, 20(3), 349-367.

Bennett, S. N. (2004) Assessment in Chemistry and the Role of Examinations, *University Chemistry Education*, 8(2), 52-57.

Biggs J.B. (2002) *Aligning teaching and assessment to curriculum objectives*. LTSN Imaginative Curriculum Guide IC022. [available on line, free, from: jisctechdis.ac.uk].

Biggs, J. (1999) *Teaching for Quality Learning at University Buckingham UK* : SRHE and Open University Press.

Black, P., and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5 (1), 7–74.

Blakemore, C. (1988) *The Mind Machine*, London: BBC Publication, Page 46.

Bloom B. S. (1956). *Taxonomy of Educational Objectives,* Handbook I: The Cognitive Domain. New York: David McKay Co Inc.

Bloom B.S. Bertram B. M, and. Krathwoh, D.R. (1964). *Taxonomy of Educational Objectives* (two vols: The Affective Domain & The Cognitive Domain). New York. David McKay.

Bostock, J. (2004) *Motivation and electronic assessment*. In: Irons, A. and Alexander, S. (Eds.) Effective Learning and Teaching in Computing (86-99), Routledge Falmer: London.

Boud, D. (1995). *Assessment and learning: Contradictory or complementary?* In P. Knight (Ed.), Assessment for Learning in Higher Education, London: Kogan Page.35.

Boud, D. (2007) *Reframing Assessment as if Learning Were Important*, In Rethinking Assessment in Higher Education: Learning for the Longer Term, edited by David Boud and Nancy Falchikov, 14–26, Abingdon: Routledge.

Broadfoot, P., and Black, P. (2004) Redefining Assessment? The First 10 Years of Assessment in Education, *Assessment in Education*, 11(1), 7–27.

Brooks, V. (1993) The resurgence of external examine in Britain: a historical review, *British Journal of Educational Studies,* 4(1), 59-72.

Brown, S., Rust, C. and Gibbs, G. (2003) *Strategies for diversifying assessment.* Oxford: OCSD.

Bruning, R.H., Schraw, G.J. & Ronning, R.R. (1995) *Cognitive Psychology and Instruction*, Englewood, USA, Prentice-Hall Inc.

Burton, L and Haines, C. (1997) Innovative in teaching and assessing mathematics at university level, *Teaching in Higher Education,* 2(3), 273-293.

Capper, J. (1996) *Testing to Learn - Learning to Test*, Washington DC, International Reading Association and the academy for Educational Development.

Chaiken, S. H., & Eagly, A. (1993) *The Psychology of Attitudes*, Harcourt Brace College Publishers, New York.

Chandi, S.S. (2008) *Using Systems-Thinking as a Teaching and Learning Tool in Biology Education* (PhD thesis), University of Strathclyde and University of Glasgow.

Chandi, S.S., Reid, N., McWilliam. A., and Gray, D. (2009) Exploring the usefulness of a systems-based model from the perspective of biology students, *Scottish Educational Studies*, 41(1), 68-85.

Chen, W-C. and Whitehead, R. (2009) Understanding physics in relation to working memory, *Research in Science and Technological Education*, 27(2), 151–160.

Chu, Y-C and Reid, N. (2012) Genetics at school level: addressing the difficulties, *Research in Science and Technological Education,* 31(1), 1-25.

Clarkeburn, H., Downie, R., and Reid, N. (2000) Teaching Biology Students Transferable Skills, *Journal Biology Education,* 34(3), 133-137.

Cooper, J. M. (1999) *The teacher as a decision-maker,* in Classroom Teaching Skills. (6th Ed.) James M. Cooper (editor) pp. 1-19, Boston: Houghton-Mifflin.

Croxford, L. (1999) *League Tables - who needs them?* Centre for Educational Sociology, Edinburgh.

Currie, R. (1986) 'Objective testing' in Lloyd - Jones, R., Bray, E., Jonson,G. and Crrie, R (eds,) *Assessmest from Principles to Action*, London, Macmillan Education Ltd., 47-54.

Danili E. and Reid N., (2005) Assessment formats: do they make a difference? *Chemistry Education Research and Practice*, 6, 204-212.

Dietel, R. J., Herman, J. L., & Knuth, R. A. (1991). *What does research say about assessment?* NCREL, Oak Brook. Available online: http://www.ncrel.org/sdrs/areas/stw_esys/4assess.htm

Duncan, K. D. (1974) *Analytical techniques in training design*. In Edwards, E. and Lees, F. P. (Ed) The Human Operator in Process Control, London, UK: Taylor and Francis, 283-319.

Earl, L., and Cousins, J. B. (1996). *Classroom Assessment*. Canada: Ontario Public School Teachers' Federation.

Egan, K. (1972) Structural communication - A new contribution to pedagogy, *Programmed Learning and Educational Technology*, 1, 63-78.

Eggen, P. and Kauchack, D. 2004. *Educational psychology: windows on classrooms*. 6th ed. New Jersey: Pearson.

Friel, S and Johnstone, A.H. (1978a) A Review of the Theory of Objective Testing, *School Science Review*, 59, 33-8.

Friel, S and Johnstone, A.H. (1978b) Scoring Systems which allow for Partial Knowledge, *Journal of Chemical Education*, 55, 717-9.

Friel, S and Johnstone, A.H. (1979a) Seconds Thoughts in Multiple Choice Tests in Science, *Journal of Chemical Education*, 56, 326.

Friel, S and Johnstone, A.H. (1979b) Does the Position of the Answer in a Multiple Choice Test Matter?, *Education in Chemistry,* 16(6), 175.

Friel, S. and Johnstone, A.H. (1988) Making Test Scores Yield More Information, *Education in Chemistry,* 25 (2), 46-49.

Gagné, R. M. (1964) *Problem solving*, in A.W. Melton (Ed.), Categories of Human Learning. New York: Academic Press.

Gagné, R. M. (1974) *Principles of Instructional Design*, Holt, Rinehart & Winston, Inc. USA.

Gagné, R. M. (1977) *The Conditions of Learning*, 3rd ed. New York: Holt, Rinehart and Winston, 1977.

Gardner, J., Harlen, W., Hayward, L., Stobart, G., and Montgomery, M. (2010) *Developing Teacher Assessment*, Maidenhead, Open University Press.

Griffin, A. H. (2001) Thinking in education yesterday, today and tomorrow. *Education,* 106(3), 268-280.

Habib, M. A. (1996) *Thinking: Theoretical Foundation & Strategy.* Cairo: Arabic Renaissance Library.

Habib, M. A. (1997) Self-Reflection and Innovative Features Associated with Multi-dimensional Thinking for Undergraduate Students, *Journal of Psychology*, Nos. 40-41.

Hargreaves, A., Earl, L. and Schmidt, M. (2002) Perspectives on Alternative Assessment Reform, *American Educational Research Journal,* 39(1), 69–95

Hargreaves, D. (1992) *Report of the Committee on the Curriculum and Organization of ILEA Secondary Schools.* London; ILEA

Hargreaves, D. (2004) *Learning for Life: The foundations of lifelong learning.* Bristol: The Policy Press

Harvey, L., and Green, D. (1994). *Employer satisfaction: summary.* Innovation in Higher Education Unit, University of Central England: Quality in Higher Education Project.

Hassan, A.K., Hill, R.A. and Reid, N. (2004) Ideas Underpinning Success in an Introductory Course in Organic Chemistry, *University Chemistry Education*, 8, 40-51.

Hayes, D., Mills, M., Christie, P. and Lingered, B. (2006) Teachers and Schooling Making a Difference, Sydney: Allen & Unwin.

Hayes, J. R. (1981). *The complete problem solver*, Philadelphia: The Franklin Institute Press.

HeraldScotland (Dec, 2011) *School league tables show extent of competition, not quality of education,* HeraldScotland, Glasgow

Hindal, H., Reid, N. and Whitehead, R. (2013) A Fresh look at High Ability, *International Journal of Instruction*. 6(1), 59-76.

Hornby, W. (2003) Assessing using grade-related criteria: a single currency for universities? *Assessment and Evaluation in Higher Education*, 28, 435 - 454.

Huba, M. E. and Freed, J. E. (2000) *Learner centred assessment on college campuses – shifting the focus from teaching to learning*. MA: Allyn and Bacon.

Jerwan, F. (2002) *Teaching Thinking: Conception and Application.* Amman, Jordan: Dar Alfiker.

Johnstone and Ambusaidi (2001) Fixed response questions with a difference, *Chemistry Education Research and Practice*, 2(3), 313-327.

Johnstone, A. H. (1988) *Methods of assessment using grids*. Lab Talk, 10, 4-6.

Johnstone, A. H. (2003) *LTSN Physical sciences practice guide: Effective practice in objective assessment*. Hull: LTSN. This monograph is online: https://www.heacademy.ac.uk/resource/effective-practice-objective-assessment

Johnstone, A. H., Bahar, M. and Hansell, M. H. (2000) Structural communication grids: A valuable assessment and diagnostic tool for science teachers. *Journal of Biology Education,* 34(2), 87-89.

Johnstone, A.H. (1972) *An evaluation of new Scottish chemistry syllabuses*, PhD Thesis, University of Glasgow, Glasgow.

Johnstone, A.H. (1981) Chemical Education, Facts, Findings and Consequences, (The Nyholm Lecture), *Chemistry in Britain,* 17(3), 130-135.

Johnstone, A.H. (1984) New Stars for the Teacher to Steer by?, *Journal of Chemical Education,* 61(10), 847-9.

Johnstone, A.H. (1991) Why is Science Difficult to Learn? Things are Seldom What They Seem, *Journal of Computer Assisted Learning*, 7, 75-83.

Johnstone, A.H. (1993) In, *Creative Problems Solving in Chemistry,* Royal Society of Chemistry, London.

Johnstone, A.H. (1993) The Development of Chemistry Teaching - A Changing Response to Changing Demand, *Journal of Chemical Education*, Vol. 70, No. 9, pp. 701-705.

Johnstone, A.H. (1997) Chemistry Teaching, Science or Alchemy?, *Journal of Chemical Education,* 74(3) 262-268

Johnstone, A.H. and El-Banna, H. (1986) Capacities, Demands and Processes: a Predictive Model for Science Education, *Education in Chemistry*, 23(3), 80-84.

Johnstone, A.H. and El-Banna, H. (1989) Understanding Learning Difficulties - A Predictive Research Model, *Studies in Higher Education*, 14(2), 59-68.

Johnstone, A.H. and Kellett, N.C. (1980) Learning Difficulties in School Science - Towards a Working Hypothesis, *European Journal of Science Education*, 2(2), 175-81.

Johnstone, A.H. and Mughol, A.R. (1979) Testing for Understanding, *School Science Review*, 61, 174-50.

Jung, E-S., and Reid, N. (2009) Working Memory and Attitudes, *Research in Science and Technological Education,* 27(2), 205-224.

Killington, S.A. and Mitchell, A.C. (1980) Designing an assessment system Using art activities in the assessment of science in the primary school, *School Science Review*, 73(264), 33-47.

Kirschner, P.A., Sweller, J. & Clark, R.E. (2006) Why Minimal Guidance during Instruction Does not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-based, Experiential, and Inquiry-Based Learning, *Educational Psychologist*, 41(2), 75-86.

Kulieke, M., Bakker, J., Collins, C., Fennimore, T., Fine, C., Herman, J., Jones, B.F., Raack, L., & Tinzmann, M.B. (1990). *Why should assessment be based on a vision of learning?* [online document] NCREL, Oak Brook: IL. Available online: http://www.ncrel.org/sdrs/areas/rpl_esys/assess.htm

Lahey, B. B. (2004) *Psychology: an introductio*n, 8th ed. Boston: McGraw Hill.

Learning Landscapes (2009) Centre for Open Learning of Mathematics, Science, Computing and Technology. Open University, Miltm Keyes [downloadable: http://www.open.ac.uk/opencetl/resources/colmsct-resources/learning-landscapes-2009}

Leonard, L., Macteer. E. and Wilson, R. (1996) *Report of objective testing study group,* Glasgow, Institute of Biomedical and Life Sciences, University of Glasgow.

Lloyd - Jones, R. (1986) An overview of assessment, in Lloyd - Jones, R., Bray, E., Jonson, G. and Currie, R. (eds.) *Assessment from Principles to Action*, London, Macmilon Education Lts., 1-13.

Lorsbach, A.W., Tobin, K., Brisocove, C. and Ulerick, S. (1992) An interpretation of assessment methods on middle school science, *International Journal of Science Education*, 14(3), 305-317

Maxwell, G. S. (2004) *Progressive assessment for learning and certification:* some lessons from school-based assessment in Queensland. Paper presented at the third Conference of the Association of Commonwealth Examination and Assessment Boards, March 2004, Nadi, Fiji.

McGrath, M. and Ingham, A. (1992) Using art activities in the assessment of science in the primary school, *School Science Review*, 73(264), 33-47.

Meyer, H. D., & Benavot, A. (2013). *PISA, Power and Policy: the emergence of global educational governance,* Oxford, Symposium Books.

Miller, G. (1956) The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information, *The Psychological Review*, 63(2), 81-97.

Miller, G. (1990) The assessment of clinical skills/competence/performance, *Academic Medicine,* 65(9), 63-67.

Muir, R. (2013) *The impact of league table reform on vocational education in schools,* IPPR, London.

Nitco, A. and Brookhart, S. (2011) *Educational assessment of student.* (Sixth Edition). Boston: Pearson Education, Inc., p. 507

Onwumere, O. (2009). *Difficulties in Understanding Mathematics: An Approach Related to Working Memory and Field Dependency.* PhD thesis submitted to University of Glasgow, UK.

Oraif, F. (2007) *An exploration of confidence related to formal learning in Saudi Arabia*. PhD thesis, University of Glasgow, Glasgow.

Pidikiti, N.P. (2006) *Performance of secondary students in India related to working memory with reference to some learning styles*, MSc Thesis, University of Glasgow, Glasgow.

Pohl, M. (2000) *Learning to Think, Thinking to Learn: Models and strategies to Develop a Classroom Culture of Thinking*. Cheltenham, Vic: Hawker Brownlow.

Ramsden, P. (2003) *Learning to teach in higher education*. London & New York: Routledge Falmer

Reid, N and Serumola, L. (2006) Scientific Enquiry: The Nature and Place of Experimentation: a review, *Journal of Science Education*, 7(1), 1-15.

Reid, N and Serumola, L. (2007) Scientific Enquiry: The Nature and Place of Experimentation: some recent evidence, *Journal of Science Education*, 7(2), 88-94.

Reid, N and Shah, I. (2010) The Idea of the Paper Laboratory, *Journal of Science Education*, 11(10), 8-12.

Reid, N. (1976) Simulations, Games and Case Studies, *Education in Chemistry,* 13, 82-83.

Reid, N. (2003) *Getting Started in Pedagogical Research in Higher Education*, LTSN Physical Science, Higher Education Academy, Hull, ISBN 1-903815-07-X, [online: http://www-new2.heacademy.ac.uk/assets/ps/documents/practice_guides/practice_guides/ps0076_getting_sarted_in_pedagogic_research_in_the_physical_sciences_aug_2004.pdf]

Reid, N. (2008) A scientific approach to the teaching of chemistry (The Royal Society of Chemistry Nyholm Lecture, 2006–2007), *Chemistry Education Research and Practice*, 9(1), 51–19.

Reid, N. (2008) *Landmarks in Assessment*, Unpublished monograph, Centre for Science Education, University of Glasgow, Glasgow

Reid, N. (2009) Working Memory and Science Education, *Research in Science and Technological Education,* 27(2), 245-250.

Reid, N. and Yang, M-J. (2002) Open-ended Problem Solving in School Chemistry: A Preliminary Investigation, *International Journal Science Education*, 24(12), 1313-1332.

Reid, N. and Yang, M-J. (2002) The Solving of Problems in Chemistry: The More Open-ended Problems, *Research in Science and Technological Educatio*n, 20(1), 83-98.

Reid, N. (2009) Working Memory and Science Education, *Research in Science and Technological Education,* 27(2), 245-250.

Robinson, K. (2011) *Out of Our Minds: Learning to be Creative*, Capstone Publishing Ltd, Chichester. (He makes some very interesting observations of the effects in parts of the US).

Rowntree, D. (1977) *Assessing Students: How Shall We Know Them ?* (1) London: Harper and Row.

Saadah, G. (2003) *Teaching Thinking Skills*. Amman, Jordan: Dar Al Shorok.

Scottish Examination Board (1962a) *Curriculum Papers 490, Alternative Physics for Ordinary and Higher Grade*. Scottish Examination Board: Dalkeith, Edinburgh.

Scottish Examination Board (1962b) Curriculum Papers 512, Alternative Chemistry for Ordinary and Higher Grade. Scottish Examination Board: Dalkeith, Edinburgh.

Scottish Examinations Authority (1997): *www.sqa.org.uk/*

Slavin, R.E. (2000) E*ducational Psychology* (6th Edn), Boston, USA, Allyn & Bacon.

Smith, G.F. (2002) Thinking skills: the question of generality. *Journal of Curriculum Studies,* Vol. 24, No. 6, pp. 659-678.

Snowman, Macon and Biehler, (2009) *Psychology applied to teaching* (13th Edition). Belmont, CA: Wadsworth, Centage Learning.

Soliman, I. (1999) *Assessing Student Learning.* Introduction to University Teaching Series, Teaching and Learning Centre, University of New England.

Spencer. E. (1981) Inter-Marker Unreliability in SCE O Grade English Composition. Is Improvement Possible? *Scottish Educational Review*, 13(1), May 1981.

Spencer. E., (1981) Inter-Marker Unreliability in SCE O Grade English Composition. Is Improvement Possible? *Scottish Educational Review*, 13(1), 41-55.

St Clair-Thompson and Botten, C, (2009) Working memory and science education: Exploring the compatibility of theoretical approaches, *Research in Science and Technological Education*, 27(2), 139-150.

Stiggins, R. J., and Conklin, N. F. (1992) *In teachers' hands: Investigating the practices of classroom assessment,* Albany: State University of New York (SUNY) Press.

Stodolsky, S.S. (1988) *The subject matters: Classroom activity in math and social studies*, Chicago: University of Chicago Press.

Sweller, J., van Merrienboer, J., & Paas, F. G. (1998) Cognitive architecture and instructional design: Educational Psychology Architecture and instructional design. *Educational Psychology Review* , 10, 251-296.

Teaching Units in the Sciences: A set of 16 units in chemistry was developed in the 1970s, several units exist in physics while there is a range of units for biology: details can be obtained from one of the authors: dr_n@btinternet.com There is also a set of short units in a textbook: Johnstone, A.H., Morrison, T.I. and Reid, N (1980) *Chemistry about us*, London, Heinemann. Old copies of this out-of-print book are readily available.

West, l.H.T. & Fensham, P.J. (1974) Prior Knowledge and the learning of science: a review of Ausubel's theory of this process, *Studies in Science Education*, 1, 61-81.

Wills, D. (1993) Learning and assessment: exposing the inconsistencies of theory and practice, *Oxford Review of Education*, 19(3), 383-401.

Wood C.A. (2006) The development of Creative problem solving in chemistry, *Chemistry Education Research and Practice*, 7(2), 96-113.

Yang, M-J. (2000) Problem solving in chemistry at secondary school, PhD Thesis, University of Glasgow, Glasgow. [http://theses.gla.ac.uk/2161/]

Yang, M-Y. (2000) Problems Solving in Chemistry at Secondary School Science, PhD Thesis, Glasgow, University of Glasgow. [http:// theses.gla.ac.uk/ 2161/]

Zaitoon, H. H. (2003) *Teaching thinking: Vision applied to the development of the thinking mind*. Amman, Jordan: Dar Al Shorok