

## ASSESSMENT

A paper by Professor Jim Tognolini  
 Director, Educational Testing Centre  
 University of New South Wales

My view of assessment is that it is the process of collecting information for a specific purpose. It occurs in education whenever one person (generally the teacher) collects information on the skills, understanding, knowledge, attitudes, etc of the persons (generally the students).

If one is prepared to accept such a definition, then it is possible to imagine that a teacher can and does collect information about a student in numerous ways. I find it useful to classify these various assessment procedures according to their degree of structure, and to represent them along a continuum.

*Figure 1: A range of Assessment procedures*

<i>Assessment is the collection of information for a purpose</i>			
Less formal			More formal
Unstructured assessment	Slightly structured assessment	More structured assessment	Most structured assessment
<ul style="list-style-type: none"> <li>• Chance meetings</li> <li>• Teacher initiated conversations</li> </ul>	<ul style="list-style-type: none"> <li>• Questionnaires</li> <li>• Observation schedules</li> <li>• Student self-assessment</li> </ul>	<ul style="list-style-type: none"> <li>• Classroom tests</li> <li>• Checklists</li> <li>• Practical work</li> <li>• Project assessment</li> <li>• Case studies</li> </ul>	<ul style="list-style-type: none"> <li>• Examinations</li> <li>• Standardised tests</li> <li>• Published aptitude tests</li> </ul>

Teachers use most of these procedures to build up a profile of what each individual student knows and can do. They generally maximize the validity of the results by matching what is being assessed, to the particular assessment procedure. They take each piece of information as an indicator and weave it into the profile of the student. When this information is at variance with what they expect, they ask why.

It would seem apparent that the best profile of a student would emerge from using multiple sources of assessment. If we rely on only one or two measurements from only

one end or the other of the continuum, then it is likely that we will get an incomplete or biased profile.

Some students perform unexpectedly on such assessments. The question for the discerning teacher is why? Grappling with such questions helps us become better teachers.

## **Norm-referenced assessment**

Measures, or test scores, by themselves do not have inherent meaning. For example, consider a score of 88. The first question to be asked is, 88 out of what? Most people assume a percentage. However, this is not necessarily the case. If the score is 88 out of 90, it conveys a different meaning than 88 out of 200.

The next question that would be asked is, What did everybody else get? This gives us an idea of the relative difficulty of the test. We have traditionally used class averages to convey this meaning to parents. For example a score of 88 when the class average is 60 implies that the student has achieved a good score on the test. A score of 88 when the class average is 98 would convey a different meaning again.

Other measures that have been used to represent norm referenced assessment include percentiles, deciles, reading ages.

The process of giving meaning to scores by referencing them to the cohort or normative group is called norm or cohort referencing .

**The process of giving meaning to scores by referencing them to the cohort or normative group is called norm or cohort referencing.**

Traditionally, test results have gone hand-in-hand with norm referenced reporting of results.

The weaknesses associated with using such information have been articulated on many occasions and I don't wish to repeat all of them here. However, let me illustrate the issue by reference to a couple of problems.

The first problem is that the results are not referenced to standards. They give no indication of what it is that students know and can do. Students scores are determined by the performance of the cohort group. If they happen to be a member of a very able cohort, then they will generally be given a lower score, irrespective of how hard they work or how much they improve (if the other students also improve as well).

There is no doubt that the idea of assessment evokes emotive reactions. Jerold Zacharias once wrote:

“I feel emotionally toward the testing industry as I would toward any other merchant of death. I feel that way because of what they do to the kids. I 'm not saying they murder every child - only 20 percent of them. Testing has distorted their ambitions, distorted their careers.”

Of course Zacharias was not upset with tests themselves. Rather he is disturbed by the way the scores are referenced to norms. Some students will always fail and as a consequence they will have their ambitions and aspirations distorted.

A related problem to this is that most people do not understand the nature of norm-referenced scores and interpret them in an absolute manner. This has led to nobody having any faith in the measures that are assigned in the social sciences. Consider a situation whereby a student attends a school and is given a test in mathematics. The student gets 25 out of 40 on the test. The cohort in the school is a very able mathematical group and the average score is 35 out of 40. The student, parents or teachers reading the report at another school are given the impression that the student is not very mathematically able. If the student attended a different school did the same test and got exactly the same items correct: in this school, however, the students are not as mathematically able (average 20 out of 40). The student is given an entirely different message. Which is correct?

## **Criterion-referenced Assessment**

In the 1960 s, for the types of reasons outlined above, there was a move away from referencing the scores of individuals to the norm, towards referencing the performance to some specified behavioural criterion or proficiency, this is referred to as criterion referencing. The focus for a criterion referenced measure is on what students can, and cannot do.

**The process of giving meaning to scores by referencing them to some specified behavioural criteria or proficiency is called criterion referencing.**

The support for criterion-referenced measurement has originated in large part from

- the emphasis on behavioural objectives;
- the sequence and individualization of instruction;
- the development of programmed materials;
- a learning theory that suggests anybody can learn anything if given enough time; and,
- a belief that norm-referencing promotes unhealthy competition and destroys the self esteem of low-scoring students.

The major problem with criterion referencing is that it became closely linked with the atomization of the curriculum and the minimum competency movement. Large corporations (like the International Objectives Exchange) were set up and sold sets of objectives and accompanying sets of items. This link to competency and mastery introduced a range of issues about setting cut-off scores, about defining when a student had demonstrated competence enough times and about answering the question, Once competent, always competent?

However, the basic principles underpinning criterion referenced assessment are sound. That is, a student's image should be referenced to what it is that students know and can do.

## **Standards-referenced Assessment**

Traditionally, experienced teachers have internalised a **standard based** upon years of teaching, assessing and recording. They then compare the image to this **standard** and make comments such as, this student is an A student or a B student; or this piece of work is worthy of a high distinction or a distinction.

**This process of giving meaning to marks by referencing to standards is referred to as Standards Referenced Assessment.**

The problem is that while such teachers are often consistent in their judgements, there is no articulation of these internalised standards. Hence there is no way that people can be confident that the teachers are using the same standards. There is also no way that reporting of performance to students and parents can be standardized. In addition, some inexperienced experienced teachers may have difficulty internalising the standards unless they were written down. The process of articulating these standards is quite difficult, but once they have been expressed, the process of referencing the image to the standard and making an on-balanced judgement is straightforward and is consistent with the way that teachers have operated over the years.

The NSW Board of Studies has gone through a rigorous process using the latest advances in psychometric theory, in conjunction with the professional knowledge and judgement of experienced teachers and examiners, to establish performance standards. The performance standards are the levels of student achievement linked to the syllabus outcomes. The task of test constructors and teachers is to construct an image of each student's achievements based upon the outcomes statements that are contained in the syllabus. Student performance is then referenced to the performance standards and then reported against these standards.

A Standards-referenced System requires examiners to develop examinations that faithfully reflect the syllabuses and, more importantly, to produce and use marking keys that are explicitly identified with the outcomes being measured.

The advantages of a standards referenced system are that:

- student performance is assessed against pre-determined standards of achievement
- once established standards remain constant
- marks reflect the standards achieved rather than a pre-determined distribution
- assessment tasks are linked to outcomes
- students know what is expected of them.

## **Summary**

In summary, therefore, a standards referenced assessment model is characterised by

- 1 **A Syllabus** which contains explicit statements in the form of syllabus aims, objectives, outcomes and content.
2. **Teachers implementing the syllabus statements** using a variety of Teaching/Learning Programs.
3. **Teachers using a range of assessment devices** to produce a faithful image of what it is students have learnt.
4. **Students performance being referenced** to the pre-determined performance standards.