

# **Tensions between assessment for learning and assessment for qualifications**

**Terry Crooks  
Educational Assessment Research Unit  
University of Otago, New Zealand**

## **Abstract**

Assessment for learning and assessment of learning (or formative and summative assessment) are always somewhat in conflict with one another. For maximum learning benefit, students need to trust teachers and be willing to expose and discuss any difficulties they are having, and the assessment has to occur early enough to be of use for learning purposes. Teachers' feedback should have particular characteristics, including tailoring to the characteristics and needs of individual students. These requirements are not well matched with the requirements of high stakes assessment of learning, particularly where internal (teacher) assessment plays a major role in the award of qualifications.

**Paper presented at the Third Conference of the  
Association of Commonwealth Examinations and  
Accreditation Bodies (ACEAB)**

**Nadi, Fiji, 8-12 March 2004**

## **Tensions between Assessment for Learning and Assessment for Qualifications**

In the past 20 years, educational researchers have given substantial attention to investigating the impact of assessment practices on student learning and motivation (Crooks, 1988; Madaus, 1988; Black & Wiliam, 1998a,b; Assessment Reform Group, 1999, 2002). The substantial effects identified have suggested that it is important to think about how assessment can be used to support and improve students learning and motivation. This formative use has been described as “assessment for learning” (Assessment Reform Group, 1999), while the more familiar summative use is now commonly described as “assessment of learning”.

Over a similar period, there have been substantial advances in thinking about the validity of educational assessments, with greater emphasis on the consequences of these assessments as an important factor in validity (Cronbach, 1988; Messick, 1989; Shepard, 1993). This has been accompanied by greater emphasis on the primacy of validity in evaluating the effectiveness of assessments, and ways of tackling the difficult task of evaluating the validity of assessment processes (Kane, 1992; Shepard, 1993; Crooks, Kane and Cohen, 1996).

In most educational settings, both assessment for learning and assessment of learning are taking place, with varying degree of relative emphasis. Despite the prominence given to the recent research on assessment for learning, however, very little consideration has been given to how validity theory might be applied to assessment for learning. Indeed, some scholars have argued that while the consequences of assessments are clearly important, it would be wise not to consider them separately from validity (Popham, 1997; Mehrens, 1997). However, it seems to me that consequences are absolutely central considerations in the validity of assessment of learning (Crooks, 2001).

The following statement is made in my university’s policy on assessment of student performance (University of Otago, 2001):

Validity requirements are different for the formative and summative purposes of assessment. Validity is high for formative purposes when the assessment procedures help the student considerably towards achieving the real goals of the paper [course, learning]. Validity is high for summative purposes when the assessment gives an accurate account of the student’s capabilities at the time the final grade is awarded or the selection decision is made. (p.3)

This statement takes a whole course perspective: an individual assessment during the course could be described as valid for summative purposes at the time it occurred if it gave an accurate account of the student’s capabilities at that time.

As Crooks, Kane and Cohen (1996) have explained, the different requirements for validity for these two different purposes mean that different factors need to be considered, or different relative weights given to the same factors. This can lead to a tension between the two purposes, especially if the same assessments are to be used for both purposes. For instance, assessment relatively early in a learning experience may be particularly valuable for learning purposes, but be distinctly inappropriate later for end-of-course summative

purposes. Indeed, if it makes a real difference to the student's learning, perhaps as a result of the feedback resulting from it, that advance of learning means that the assessment is no longer accurate as an assessment of learning.

This paper explores this tension between the two purposes, and the consequences of various assessment practices for the resulting validity for the two purposes. While not restricting itself to high-stakes assessments leading to the awarding of qualifications, it gives particular attention to qualifications that are based on a mix of assessments during and at the end of the course. Assessment toward the final grade during the course is described here as "internal assessment" (which might elsewhere be described as teacher assessment, continuous assessment, coursework assessment, or school-based assessment). Assessment at the end of the course is described here as a "final examination", whether it is prepared by the teacher or by an external agency or group. It would usually take the form of a written test or examination.

Because the final examination occurs at the end of the course or learning experience, it might be regarded as solely assessment of learning, or summative assessment. In many cases the student will receive no feedback other than the resulting mark or grade, but this does not mean that the final examination has no formative effects. Indeed, many of the effects described by Crooks (1988) can be associated with final examinations. Students' motivation, study habits and specific learning activities are affected by the existence, expected nature and content of the impending examination, while the experience of sitting the examination and the result obtained can have a major impact on student perceptions of ability or self-efficacy, and on their future study options and choices. If additional feedback is available, it may also help students to handle future learning and assessment better.

Internal assessments have all of the effects associated with final examinations, but the work is returned to the students, usually accompanied by comments designed to indicate strengths and weaknesses. In addition to this specific feedback to individual students, there may be feedback to the whole class based on issues common to several or many students. The feedback may focus on justifying marks or grades awarded, helping the students do better in the future, or both.

A third form of assessment, present in almost all teaching contexts, I describe as "feedback assessment". These are not intended to have a summative role in the course mark or grade. They may be planned assessments, involving work prepared by the students and submitted to the teacher, or they may be incidental assessments, where the teacher provides feedback during learning activities. Some people might not regard the latter as assessment at all, but rather a part of teaching, but I prefer to view any judgement of student work and feedback on it as assessment.

### **Validity for Summative Purposes (Assessment of Learning)**

In this section, I briefly discuss high stakes assessment of learning, such as for qualifications, looking first at the advantages and disadvantages of using only a final examination, then at the advantages and disadvantages of internal assessment, and finally at determining an optimal combination of both. It is helpful to recall the earlier statement, from the University of Otago assessment policy, that validity is high for summative

purposes when the assessment gives an accurate account of the student's capabilities at the time the final grade is awarded or the selection decision is made.

### *Final Examination*

The advantages of assessment based on a final examination are that the assessment provides the most up-to-date information possible (immediately before the grade is awarded), the work usually is done under conditions that give high confidence that it is the student's own work, marking may be able to be standardised across several or many teachers, and if this is the case prior knowledge and expectations of particular students is unlikely to influence the marking.

Corresponding disadvantages are that the performance is on one occasion only (limited sampling of student behaviour, given known day to day variations), the intensity of the final examination causes severe examination anxiety for some students resulting in their capabilities being under-represented on the day, and there are major limitations in the range of types of outcomes that can be assessed in a final examination format. A less obvious disadvantage is that because the examination often has been set by someone other than the students' usual teachers, the curriculum choices and guidance given by the teachers may be poorly matched with the actual requirements of the examination. While the resulting student performances may accurately reflect what the students knew and could do on that day, they may not be very valid for the purposes for which the result is to be used, such as admitting students to the next stage of education. Students who have had ideal guidance may be unfairly advantaged over those who have not.

### *Internal Assessment*

Including internal assessment in the overall package of assessment of learning has four main potential benefits. First, it often allows important outcomes to be assessed that could not be assessed in a final examination, such as practical work, extended projects or assignments, and oral or aural work. Second, the assessments occur on multiple occasions, reducing the sampling risks associated with single occasion assessment. Third, it spreads the assessment load through the course, encouraging more consistent work and reducing the risk of extreme assessment anxiety. Finally, if particular outcomes are assessed more than once, it allows the possibility of identifying atypical performances or trends in performance, which can be allowed for in the final grading decision.

Against these advantages, there are several risks or disadvantages. First, internal assessment often provides less control than final examinations over who does the work and under what circumstances. Some or all of the work may be done outside of class, with little or no supervision. Second, teachers may give more help to some students or give inappropriate weight to what they already know or believe about particular students. Third, work done part way through the course can give a misleading picture of the final achievement of students by the end of the course, and this can be particularly unfair to students who learn from their mistakes and the feedback they have received. Finally, there is a danger of multiple assessment occasions creating a treadmill effect of work, so that the course becomes a succession of graded assignments rather than a programme that offers space for the student to think and explore.

### *Optimising Assessment of Learning*

In light of the advantages and disadvantages discussed above, I would like to suggest four principles for maximising the validity of assessment of learning:

- 1 The quality and appropriateness of the individual assessments should take precedence over the total quantity of assessment. More can be less!
- 2 Things that can readily be assessed in a final examination probably are best assessed in one, unless the material has been taught much earlier and is unlikely to have been reinforced, consolidated or enriched by later material (in which case it might better be assessed at the point of completion of that material).
- 3 Things that are important but cannot be assessed in a final examination should be assessed through internal assessment, at a time when students have been given ample opportunity to develop the knowledge and skills required and will have little further opportunity to progress in these areas during the course.
- 4 If the same knowledge and skills are assessed more than once, in multiple internal assessments or both internal assessment and the final examination, students' final grades should reflect their final levels of performance rather than earlier, out of date levels. Some compromise may be necessary here: arguments in favour of greater sampling for reliability and not placing too much emphasis on a single fallible assessment suggest that students should be given the benefit of the doubt if their results improve over time, but given some credit for their higher earlier marks if their results decline.

#### **Validity for Formative Purposes (Assessment for Learning)**

It is appropriate to commence this section by reminding readers of part of the earlier quotation from the University of Otago assessment policy:

Validity is high for formative purposes when the assessment procedures help the student considerably towards achieving the real goals of the paper [course, learning].

The main issues in the validity of assessment for learning can also be seen as issues about good teaching. In our paper about threats to validity (Crooks, Kane & Cohen, 1996), my colleagues and I listed one of these threats as "Poor Pedagogical Decisions". What we were indicating was that in the decisions that teacher take about assessments, which include feedback to the students and perhaps their parents, teachers need to tailor their feedback and decisions to the individual personalities and histories of individual students. Very similar performances from two students may need totally different responses if optimal benefits for learning and motivation are to be attained. For instance, an overconfident student who has done unusually poorly may benefit from an expression of concern or disappointment and quite pointed references to weaknesses in the work, while a similar performance from a student for whom this was unusually good work may justify praise and a clear focus on the strengths of the work.

Two years ago, in a conference paper that still needs a lot more refinement and explanation before it could be called complete (Crooks, 2001), I identified twenty-five factors influencing the validity of assessment for learning (see Table 1). Rather than trying to explain these factors fully here, I wish to highly just a few of them. First, assessment for learning requires committed participants – teachers and especially students. If this commitment is missing, little is likely to be achieved. Second, trust is vital is students are to feel safe admitting the difficulties and uncertainties they are having, which maximises the chance that others can help them overcome those difficulties and uncertainties. Third,

students have to understand what is being asked of them and accept those goals and the associated criteria as appropriate. This requires that teachers have good subject knowledge and considerable expertise in explanation and negotiation. Fourth, strong emphasis should be placed on promoting students' self-assessment skills, for two reasons: self-assessment is a vital link between teacher assessments and students' actions to improve their work (Sadler, 1989) and teachers generally have quite limited time per student for individual interaction and feedback. Fifth, effective feedback from teachers to students requires insight into the difficulties students are having, good timing (feedback when students are most receptive to it), and excellent judgement about the choice of comments (balance between comment on strengths and comment on weaknesses, choosing a few important points to focus on, and tailoring the feedback to each student's personality and thinking).

This limited discussion indicates that effective assessment for learning cannot be guaranteed simply by having some assessments during a course and providing marks and some brief verbal feedback on each students' work. Indeed, there is ample evidence that much of the effort than many teachers put into such work is poorly directed and ineffective. A particular area of concern is the use of the same assessments for both formative and summative purposes.

### **Tensions Between Assessment for Learning and Assessment of Learning**

There is insufficient research on the dual use of assessments to allow us to understand the extent to which this strategy can produce high validity for both purposes, or the circumstance that determine its success or failure. There are, however, numerous reports, from teachers and students, that students often pay little attention to specific feedback if a mark or grade is also provided. These reports are accompanied by a few tantalising pieces of research. Black and Wiliam (1998) cite the research of Butler (1988), who found little learning benefit from feedback that consisted of marks alone or marks together with written comments, but substantial learning benefit where the feedback consisted solely of written comments

I think it would be a mistake to think that further research would reveal a simple conclusion about the dual use of the same assessment events to promote learning and to give credit for achievement. Indeed, the complex human interactions involved in education rarely seem to allow such simple conclusions!

Certain approaches to combining assessment of learning and assessment for learning in the same tasks seem promising. If students are allowed multiple chances to perform well in a particular area, for instance by being reassessed if they do not meet the desired standard initially, they are very likely to pay careful attention to specific feedback resulting from their initial unsuccessful attempts. Perhaps less strongly, if they are expected to perform several tasks that are similar in nature and demands, the feedback on the earlier tasks is more likely to be attended to than similar feedback on the first of quite unrelated tasks. Another option promoting attention to feedback is subdivision of a relatively large task into a series of smaller stages, with feedback available at each stage but the grade based principally or entirely on the final product. This often involves a tension for the teacher in deciding how detailed and extensive feedback can be before it raises doubt about whether the final product is more the work of the teacher or the

student – rather akin to a tightrope walking act, with dangers in both directions. In these examples of approaches that involve both feedback and assessment for credit, I think students can see a close association between attending to and trying to respond to the early feedback and the possibility of greater success on later or more important tasks.

This tension between formative and summative purposes for the same assessments does not only affect students – teachers can find the tension a challenge, too. I often find myself fighting an internal battle between providing feedback that aims mainly to justify my marks or feedback that is designed mainly to help the students learn how to perform better on future occasions.

Another important aspect of the tension between assessment for learning and assessment of learning is that assessment for learning needs to occur early enough for students to learn from it and take remedial actions, while assessment of learning should occur when those actions have taken full effect. No one assessment can serve both functions well, so a carefully crafted package of assessments is needed to achieve validity for both purposes.

While the summative uses of internal assessments may limit their value for formative purposes, this does not mean that assessment for learning based solely on a final examination guarantees good formative assessment during the course. Even if the final examination is external to the learning institution, as is the case with national examinations or examinations administered by a regional examinations board, the students' own institutions often have their own internal system of assessment of learning which can be of considerable summative importance. This was the case in the secondary school I attended 40 years ago, where the internal marks awarded were aggregated to allow reporting to parents and prizes for place in class. To me, these assessments seemed of similar significance to the external examinations.

In earlier sections, I have suggested guidelines for valid assessment both of learning and for learning. Those guidelines apply to programmes in which the teacher is involved in both processes, often using the same assessments. One further suggestion is to include some assessment activities solely for formative purposes – what I described earlier as “feedback assessments”. These might be practice or preparatory assessments leading up to important assignments or the final examination. Some students may decide not to bother with these because they “do not count”: that is their prerogative and deprives them of potentially helpful feedback, but I see this as part of students increasingly accepting responsibility for their own learning.

## **Conclusion**

There is no magic wand that can remove the tension between assessment of learning and assessment for learning in most courses leading to a qualification or other fairly high stakes outcome. However, this paper has suggested some strategies that can help to achieve validity for both purposes, if intelligently applied.

## References

- Assessment Reform Group, U.K. (1999). *Assessment for learning: beyond the black box*. Cambridge, U.K: University of Cambridge School of Education.
- Assessment Reform Group, U.K. (2002). *Testing, motivation and learning*. Cambridge, U.K: University of Cambridge School of Education.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, 5, 7-74.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80, 139-149.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Erlbaum.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438-481.
- Crooks, T. J. (2001). The validity of formative assessments. Paper for the annual conference of the British Educational Research Association, Leeds, UK, 2001 [Included in the conference proceedings, but not presented due to travel delays cause by the September 11 terrorist attacks in the USA].
- Crooks, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in Education: Principles, Policy and Practice* 3, 265-285 (1996)
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527-535.
- Madaus, G. F. (1988). The influence of testing on the curriculum. In L. F. Tanner (Ed.), *Critical issues in curriculum, Eighty-seventh yearbook of the National Society for the Study of Education, Part 1* (pp. 83-121). Chicago, IL: University of Chicago Press.
- Mehrens, W. A. (1997). The consequences of consequential validity. *Educational Measurement: Issues and Practice*, 16 (2), 16-18.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.
- Popham, J. (1997). *Educational Measurement: Issues and Practice*, 16 (2), 9-13.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Shepard, L. A. (1993). Evaluating test validity. *Review of Research in Education*, 19, 405-450.

University of Otago (2001). *Senate policy on evaluation of student performance*.  
Dunedin, New Zealand: University of Otago.

**TABLE 1**  
**(from Crooks, 2001)**

## **Key Factors Influencing the Validity of Formative Assessment**

### ***Affective Factors***

|                   |  |
|-------------------|--|
| <b>Motivation</b> | Teacher is devoted to helping student learn<br>Student cares about learning and wants to improve   |
| <b>Trust</b>      | Teacher is encouraging, constructive, sensitive to student's feelings<br>Class/peer relationships and attitudes support student's learning<br>Student feels safe to admit difficulties and uncertainties |

### ***Task Factors***

|                  |  |
|------------------|--|
| <b>Knowledge</b> | Teacher understands the key aspects and difficulties of the task   |
| <b>Criteria</b>  | Teacher identifies and explains well the qualities sought<br>Student understands clearly what is needed  |
| <b>Standards</b> | Teacher sets standards appropriate to student<br>Through descriptions and examples, the standards are explained<br>Student understands the standards and accepts them as appropriate |

### ***Structural Factors***

|                    |   |
|--------------------|---|
| <b>Connections</b> | Final version of task can benefit from the formative assessment<br>Work on subsequent tasks can benefit from the formative assessment |
| <b>Purposes</b>    | Formative use of task is not undermined by parallel summative use   |

### ***Process Factors***

|                         |   |
|-------------------------|---|
| <b>Self-assessment</b>  | Teacher helps student to develop self-assessment skills<br>Student takes increasing responsibility for his/her own learning         |
| <b>Peer involvement</b> | Teacher encourages collaboration among students to improve work<br>Peers learn to be constructive and generous in offering feedback |
| <b>Monitoring</b>       | Teacher monitors student's work to track both process and progress  |
| <b>Insight</b>          | Teacher detects misunderstandings or other obstacles to success<br>Teacher detects exciting possibilities in student's work         |
| <b>Timing</b>           | Feedback is given at times when student is most receptive to it   |

**Balance**

Feedback gives attention to strengths as well as weaknesses

**Selectivity**

Feedback addresses mainly the aspects likely to have biggest benefit

**Wisdom**

Feedback is convincing, appreciated, and useful to student