1. Introduction

The importance of assessment in teaching and learning processes has long been acknowledged, although opinion is still sharply divided about whether (particularly in its high-stakes manifestations) it exerts a benign or malign influence. This brief paper, written as an introduction to a group of papers which will address the topic in more depth, touches on some of the ways that assessment can be used to promote and support good teaching and learning practices and approaches, as well as on some of the ways in which assessment can be considered to exert a baleful influence.

The paper contains the following sections, none of which are comprehensive:

- Some contextual factors are currently bringing assessment into greater prominence;
- Harnessing assessment practices to promote good learning habits and outcomes;
- Some ways in which assessment can negatively impact on teaching and learning

2. Some contextual factors that are currently bringing assessment into greater prominence.

2.1 Diminished contact time, greater class size and financial constraints

Gibbs (1998) emphasises the increasingly important strategic role of assessment in these times of larger class sizes and larger ratios of out-of-class to in-class time. He suggests that when students spend approximately the same amount of time in as out of class, what they do in class influences to a very great extent what they do out of class. When they spend a smaller amount of time in- as opposed to out-of-class, the influence is smaller. To illustrate this, consider the different study lives of an Arts and a Science student. The latter has (on average) greater class time and – partly because of smaller

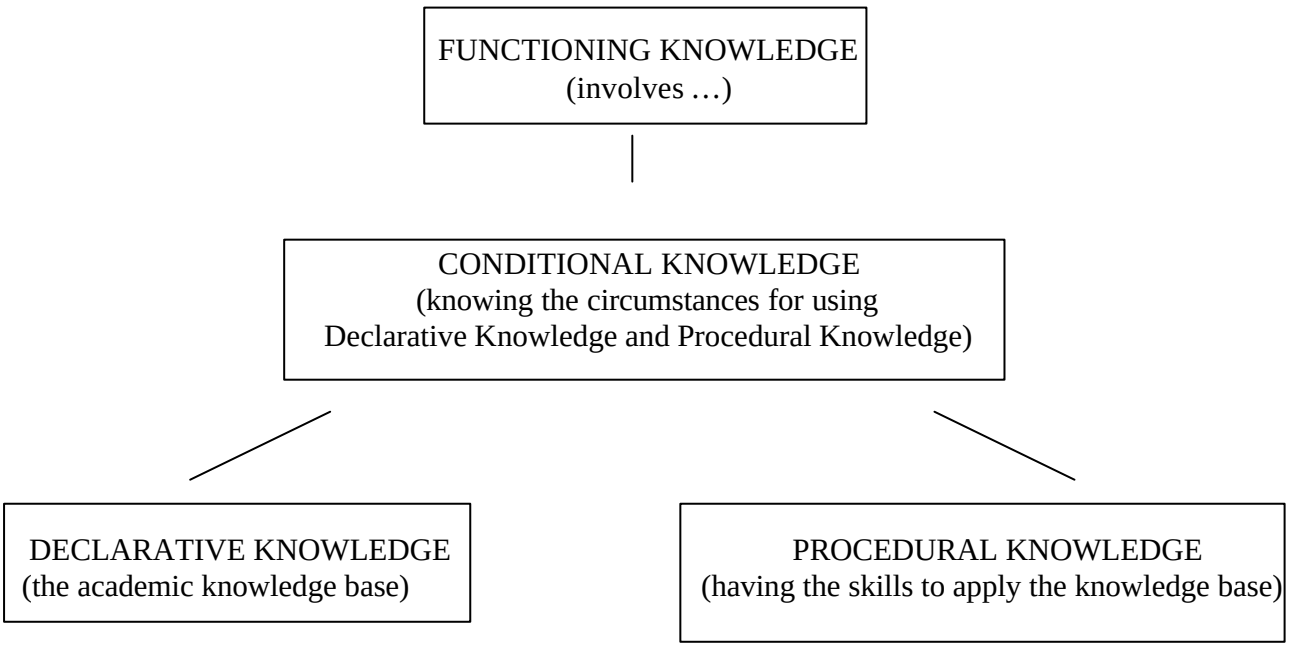
classes – greater contact with teaching staff. She spends a large proportion of her time out of class preparing for or reacting to in-class activities, such as writing up laboratory reports, preparing for field trips, and collaborating with class-mates on poster presentations. Arts students, on the other hand, do not find the allocation of their out-of-class time so clearly defined by class activities, which generally take the form of large, formal lectures aimed to prepare and guide students' reading in the discipline. They need, however, to make quite fundamental decisions about what to read/study, and how much of this is appropriate. In this situation, students "… become more strategic in allocating their out-of-class time to what counts: to what is assessed." (Gibbs 1998: 42). One only has to investigate the use of study areas on any HE library to see the truth of this: when a large class has an essay/test/project due, there is not a spare seat to be found: when it does not, there is space aplenty.

This point may seem obvious. Its significance should not be underestimated, however, in our current context. National policies emphasise 'mixed mode' delivery of instruction: that is, a mix of distance and contact. Asynchronous instruction (that is, instructional interactions that are not face-to-face and take place over time, for example by email or correspondence) is often focused on some kind of assessment activity (a project, an assignment etc.). Much of this mode of instruction depends on interactions leading up to the completion of an assignment – whether it is an electronic conversation between the professor and her student/s, or a chat room set up for a group to share their ideas or produce a joint project. Increasing class sizes, even in the Sciences (and, with budget constraints, a cut in the number and duration of tutorials and laboratory sessions i.e. a cut in 'in-class' time) are forcing students to become more strategic about how to spend their time, and thus assessment plays a greater role, as it sets the deadlines to which students work, in the main.

2.2 Outcomes-based education

The increased emphasis currently given to how students can apply their learning and understanding has important consequences for assessment. Figure 1 below illustrates how Biggs (1999) understands 'functioning knowledge' – surely the aim of outcomes based education. The figure shows the important place accorded to the knowledge base, something that in the original Curriculum 2005 in South Africa was fatally downplayed – and which has been somewhat reinstated in the National Curriculum Statements.

Figure 1: Relationships between different kinds of knowledge (Biggs 1999 41)



A wonderful example of how assessment can promote functioning knowledge is given by Gibbs (1998:50-51), in his "case of the philosopher of education". In a philosophy of education course, students were asked in their final exam each year to compare and contrast the philosophies of two philosophers and apply these to classroom practice. Predictable rote-learned responses resulted – these were not the planned outcomes, however. When a new lecturer took over the course, she kept the same content and

teaching styles, but dramatically altered the examination questions. Students were told there would be two compulsory questions, following the viewing of a 10-minute video of a teacher in a classroom. The two questions were: Comment on what is going on in this class from a philosophical point of view:" and "Advise this teacher on her future practice, from a philosophical point of view". The same questions were used every year – the difference was in the video, which changed. The upshot was that students, in preparing for the exam, were forced to examine what went on in classrooms (through videos in the resource library, and through actual classroom observation). They also had to relate what they were reading about and studying in philosophy of education, and apply this to the observations. In other words, as Gibbs observes, through the assessment the learning activity had become appropriate to the desired outcomes.

2.3 Greater range in preparedness

A growing body of evidence demonstrates that different assessment approaches elicit different performances (e.g. Johnson 1989, Gipps & Murphy 1994, Yeld 2001, etc.). While this is true for educationally homogeneous student cohorts (homogeneous in terms of educational background, or gender, for example), it is even more striking in heterogeneous groups. The work done by colleagues and me in the AARP project has demonstrated the dramatically different results that different kinds of assessments produce (between different groups of students, and for the same students).

Meeting the assessment challenges posed by heterogeneous groups requires a considerable degree of assessment expertise and experience: something long neglected by teacher-training programmes.

3. Harnessing assessment practices to promote good learning habits and outcomes

According to Chickering and Gamson 1987, good practice in undergraduate education has the following characteristics. Good practice, they argue: encourages student-staff contact; encourages cooperation between students; encourages active learning; ensures that prompt feedback is provided; ensures that learners spend appropriate time on learning tasks; insists on high but attainable standards; and respects 'diverse talents and ways of learning'.

These principles of good practice provide a useful framework for a discussion of how assessment can impact positively on teaching and learning. In the interest of space, however, not all are discussed here at the same level of detail.

3.1 Assessment as a tool to encourage active learning

According to Biggs (1999:35) "real understanding is *performative* …". By this he means that learning changes the ways in which students perceive and act on the world, and that this is best developed and demonstrated through actions on the world (i.e. the curriculum). So asking students to reproduce, to list, to describe (for example) is not getting them to 'perform' their understanding. Biggs's well-known SOLO taxonomy attempts to capture how assessment can be used to encourage students to demonstrate their understanding in progressively more cognitively complex and performative ways. Briefly, he attaches verbs to denote levels of understanding, as follows:

Level 1: identify, do simple procedure (unistructural)
Level 2: enumerate, describe, list, combine, do algorithms (multistructural)
Level 3: compare/contrast, explain causes, analyse, relate, apply (relational)
Level 4: theorise, generalise, hypothesise, reflect (extended abstract)

It should be noted that movement from one level to another comes about through the increasingly complex cognitive operations that are brought to bear on the content, and

not on the amount of content displayed. One of the techniques discussed by Biggs that I have found most meaningful in this regard is his notion of "ordered outcome items". In appearance, on ordered outcome item looks like a multiple-choice question, with a 'stem'; the difference is that the candidate has to answer all the items (not choose the correct one). The items are carefully ordered into a hierarchy of complexity: the students indicate their level of competence by correctly answering the items as far as they can up the sequence.

This requires, however, a very high degree of functioning knowledge (in terms of assessment and of the topic being assessed) on the part of the assessor – evaluating degrees of cognitive demand is notoriously difficult.

3.2 Assessment as a means of providing prompt and effective feedback

As a beginning teacher, many years ago, I found myself increasingly demoralised under the twin pressures of large class sizes (as a language teacher, I had six classes, with about 40 students in each class) and a student body highly alienated from and cynical about schooling. Homework assignments were, on the whole, sloppily completed, and only a minority of students handed in work with any regularity. Given the workload, I was unable to provide feedback of a kind that the students found meaningful (indeed, I suspected it was largely ignored, as it seemed to have no effect on their subsequent performance).

In an attempt to address the situation, I instituted a system in which everyone had to continue to hand in their work, for DP purposes. However, only one tenth of the class had their work assessed and commented on in some detail. This tenth was randomly selected: so the same student could have his/her work marked repeatedly – or not at all (although I did try to check that no-one was completely ignored over time). To my surprise, it was not long before there was a clamour from those whose work had not

been assessed – they wanted it marked! One of the most startling aspects was that this demand come not only from the more conscientious students: students who had scarcely bothered to hand in before – or at least who had seemed totally untouched by the whole process – sidled up to me and asked me please to choose them next time.

Sadly, the unanticipated success of the scheme proved its downfall: with numbers back to their former unmanageable level – but an increased demand for meaningful feedback – I was once more swamped and unable to deliver in the same way – however, we worked out a plan to address this (I would comment fully on a proportion of the scripts but all would be assessed).

4. Some ways in which assessment can negatively impact on teaching and learning

In strong contrast to the assessment-led instruction approach, many educationists and others argue that measurement-driven instruction has a pernicious effect on the curriculum. In an article written at a time in the United States when proposals to develop common examinations for each of the secondary school curricular areas were being formulated, Madaus (1988) derived seven principles, based on the findings of studies on the power of testing to influence the curriculum. Three of these, which are particularly relevant to this paper, are discussed below.

The first principle asserts that the power of tests derives largely from how important they are perceived to be by test users and takers – it is, he argues, a 'perceptual phenomenon'. This principle includes the power of tests to mean or stand for various phenomena. In this way, Madaus states, "[T]est results become a synecdoche for standards" (1988:89) – a sort of substitute, or shorthand. Thus, when test scores rise, it is believed that standards in that area have risen. However, as is illustrated below, this change could as easily mean that candidates have become more adept at taking the test,

or that the results have been manipulated in some way, as that they are more proficient in the area being assessed.

In South Africa, this presumption can be seen very clearly in the publicity surrounding pass-rates in the Senior Certificate examination. If they rise, standards are believed to have risen, and vice versa. The controversy surrounding the 2000 SC results are a case in point, with government officials claiming improvements in schooling (Grey 2001:2), and others such as noted local educationist Jonathan Jansen stating that the nine-percentage point increase reported for the 2000 results was not credible (Monare 2001).

A further principle put forward by Madaus states that if a test has important perceived consequences, it will be taught to. In and of itself, of course, this need not be a bad thing. If the test is measuring skills which are agreed, through some legitimate process, to be important, then instruction aimed at teaching these skills ought to be beneficial. This argument is at the heart of the proponents of assessment-led reform. The challenge, however - which such proponents would acknowledge - is to develop and maintain methods of assessment that authentically represent the skills. The extent to which the skills and the measurements do not correspond determines the extent to which teaching to the test will be negative in its consequences.

An illustration of how scores can change without there being a corresponding, generalized change in knowledge is provided by Shepard (1997). As she states, students "… can appear to know when they don't really know" (op cit:8). The example is based on the performances of 9 year-olds in a high-stakes testing district, i.e. a testing district where accountability testing was exerting pressure on schools to raise scores, and is derived from a study by Koretz (1991).

STANDARDISED TEST ITEM

Instruction	Options
87 **-24**	A 63 B 53 C 64 D 62

ALTERNATIVE TEST ITEM

Instruction	Options
Subtract 24 from 87	A 63 B 53 C 64 D 62

As can be seen, the vertical subtraction problem of the standardised test item was changed to a horizontal problem in the alternative test version. In both cases, however, candidates were required to manipulate the same two numbers, 87 and 24, and in both cases the same multiple-choice options were supplied, and in the same sequence. However, the performance of the candidates changed depending on which format was presented. 83% of the candidates got the standardised test item right, whereas only 66% of the candidates got the alternative test version right.

This difference in performance raises serious questions about the 'robustness' of the mastery of the skill. As Shepard suggests, students can appear to understand, "… if we keep asking them to demonstrate skills in exactly the same format" (1997:20). As this example demonstrates, however, changing the format can change performance levels. It also raises an important issue related to the existence of 'generic' skills, as illustrated below.

The issues are complex. If a candidate can perform similar tasks in one context but not in another - or in one format but not another - it is not clear what this performance means. It could, for example, mean that the candidate's understanding or knowledge is incomplete - that s/he has learned the format strategy but does not have a sound understanding of the skill. On the other hand, it could mean that one of the contexts

contains inadequate or misleading items/item formats which prevent her/him from being able to demonstrate mastery.

The difficulty of explaining why there is a difference in scores when different formats are used, however, does not lessen the importance of the existence of this difference.

What this evidence points to is the need for the developers of assessment procedures and instruments to insist on the inclusion of as wide a variety of formats as possible. Equally important, however, is that the dependence of performance on format and context is not just a measurement problem. It is, rather, the responsibility of all involved in the educational process to ensure that every avenue possible is used to "… ensure transfer and generalized knowledge" (Shepard 1997:27), and this imperative includes the avenue of assessment. A candidate ought not to be assessed as proficient until s/he has demonstrated mastery in more than one context.

The third principle discussed here asserts that irrespective of context, wherever a high-stakes test (and end of year or course module examination, for example, is established, "… a tradition of past exams develops, which eventually de facto defines the curriculum." (Madaus 1988:93). The relationship of the syllabus to the examination is crucial in this regard. If there is conflict between the aims of the syllabus and the way the course is assessed, the examination's implicit curriculum will win.

One of the difficulties is of course that tests have to *sample* what is in a curriculum, as complete coverage is not possible. However, because some parts of the syllabus are more 'testable' than are others, these parts will tend to appear more frequently, and thus will assume a more important place in the curriculum than was originally intended. Understandably, in a high-stakes situation, teachers try to maximise overlap between what they teach and what is examined - a strategy referred to by Resnick and Resnick

(1992:57) as 'curriculum alignment'. Generally, scores rise as the curriculum is aligned to the tests, i.e. where there is increasing overlap. In situations where syllabuses are not available, or where confusion exists about which version of a syllabus is in force[1], of course, it is easy to see how examinations begin to define the curriculum.

From the discussion above, it can be seen that using examinations to raise the level of knowledge and skills in an educational system is an extremely complex undertaking. In order to minimise negative consequences, great care must be taken, before such an undertaking is embarked upon, to ensure systematic and comprehensive training of teachers in respect of the design, content, and administration of the tests.

5. Conclusion

It has been repeatedly emphasized in this talk that if assessment is to promote learning and teaching (and result in meaningful certification), considerable expertise is required. It is a serious concern that so few teaching training programmes pay sufficient attention to assessment – and, too, that many assessor training programmes appear to focus more on the arrangements surrounding assessment than on the actual processes involved in designing the approaches and instruments. Until this changes, it is a bit hard to be optimistic … and assessment will continue to be viewed with suspicion.

[1] Such confusion was found in the investigation of the Ministerial Committee looking into the Senior Certificate examination in 1998, where teachers and regional coordinators were using different syllabuses for ESL-HG, and where not even the National Department could produce copies of the most recent form of the syllabus (DoE 1998).

REFERENCES

Biggs, J. (1999). *Teaching for Quality Learning at University*. Buckingham; Society for Research into Higher Education/Open University Press.

Brown, S. and Glasner, A. (Eds.)(1998). *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches*. Buckingham; Society for Research into Higher Education/Open University Press.

Chickering, A.W. and Gamson, Z.F. (1987). Seven Principles for Good Practice in Undergraduate Education. *Wingspread Journal* 9(2), special insert.

Ekstein, M.A. and Noah, J.N. (Eds.)(1992). *Examinations: Comparative and International Studies.* Exeter: Pergamon Press.

Gibbs, G. (1998). Using assessment strategically to change the way students learn. In Brown, S. and Glasner, A. (Eds.)(1998), 41-53.

Gipps, C. and Murphy, P. (1994). *A Fair Test? Assessment, Achievement and Equity*. Buckingham: Open University Press.

Grey, J. (2001). Exam Results Show a Long Road Ahead. *The Teacher* 5(1), 2.

Johnson, S. (1989). National Assessment: the APU Science Approach. London: HMSO.

Koretz, D., Linn, R.L., Dunbar, S.B., and Shepard, L.A. (1991). *The Effects of High-stakes Testing on Achievement: Preliminary Findings about Generalization across Tests.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Madaus,G.F. (1988). The Influence of Testing on the Curriculum. In Tanner L.N. (Ed.)(1988). *Critical Issues in Curriculum*. Chicago: University of Chicago Press, 83-121.

Monare, M. (2001). 'Impossible' Matric-Pass Rise to be Probed. *The Star* January 17, 2001.

Resnick, L.B. and Resnick, D.P. (1992). Assessing the Thinking Curriculum: New Tools for Educational Reform. In Gifford, B.R. and O'Connor, M.C. (Eds.)(1992).

Changing Assessments. Alternative Views of Aptitude, Achievement and Instruction. Boston: Kluwer Academic Publishers, 37-75.

Shepard, L.A. (1997). *Measuring Achievement: What Does It Mean To Test for Robust Understandings?* William H Angoff Memorial Lecture Series. Princeton NJ.: Educational Testing Service.

In terms of the format of your paper, please provide it in Microsoft Word 97, 98 or 2000 with your document settings on

Paper Size:	A4
Font:	Times New Roman
Font Size:	12 for text
Line Spacing:	1.5

You can courier it on 3.5” stiffy disk or on compact disc to:

The Chief Executive Officer
UMALUSI
Building A, The Woods
41 De Havilland Crescent
Persequor Technical Park
Lynnwood Manor, PRETORIA
South Africa