

**Some notes on assessment theory. L Higgins (2006). TLC Project  
Partner. Southampton Solent University.**

## **Introduction**

This paper is not meant as an all encompassing thesis on 'assessment theory'. It is more of a brief overview of the main elements and issues in assessing undergraduates at university.

Assessment has been defined as, '... to determine the importance, size or value of' (Longman,1984:86). Assessment is how we judge whether learning has taken place. Assessment is different from learning. It is not how people learn, but about how much and what have they learnt? Assessment answers the question, 'How do we know that a person has learnt something?' To better explain what assessment is in higher education, this paper will use Jarvis' (2005) headings to briefly note the key features of assessment literature.

## **Formative and summative assessment**

There are clear differences here in both their purpose and power base. The purpose of summative assessment is to get a secure estimate of a mark or grade. The purpose of formative assessment is to give feedback to help with learning. The power base is with the teacher in summative assessment, and is more shared with formative assessment. In short, summative assessments are high stakes, high costs, formative assessments are low stakes, low cost. (Knight and Yorke,(2003).

## **Assessment as measurement**

According to Biggs (2003) this approach is designed to measure an underlying trait or ability, and express that measurement along a graduated scale so that individuals can be compared with each other. The measurement model is based on five assumptions: it is based on a quantitative view of

knowledge; that percentages are a universal currency; that a test needs to spread students out clearly, sorting the high from the low performers; that quantitative approaches to assessment are scientific, precise and objective; and that university education is selective.

This model is prevalent in UK higher education and Biggs (2003) suggested the following reasons why: tradition and habit; bureaucratic convenience; teaching convenience; and a genuine belief in the measurement model ('my job is to sort the sheep from the goats').

### **Assessment as judgement**

This approach is designed for assessing learning in a teaching situation. This is based on a qualitative view of learning, and has three different implications for assessment. First, judgements are used to compare a learner with him or herself at another time. Second, judgment is used to make a comparison with other people ('she came second in class'). Third, a judgment can be used to compare a learner against some chosen criterion of performance. Moreover, the judgment approach to assessment is based on certain assumptions: we can set standard criteria as course or unit objectives; most students should be able to reach those standards at an acceptable level; different performances can reflect the same standards; and teachers can judge performances against the criteria (Jarvis, 2005; Biggs, 2003).

### **Validity and reliability of assessment**

The assessment literature insists that both measurement and judgment should be valid. There are different forms of 'validity'. For example, 'Face Validity' of assessment is defined as, "... the extent to which an assessment is seen to assess what it is supposed to assess" (Elton and Johnson, 2002:9). The main issues here are that in education it is seldom possible to measure exactly what learning has taken place. Assessors are forced to measure a 'surrogate' for what they are really interested in (Jarvis, et al. 2005). This led to lots of concerns about how valid assessments are. For example, according

to Knight and Yorke, (2003) assessments should engage with the programme or module's learning intentions and not with pallid, easy to handle simplifications.

In addition to being valid, the assessment literature stresses the need for assessment instruments to be reliable. The reason why is that generally, assessments are applied to large groups of people, not just an individual, and so needs to be fair. However, research has consistently suggested that marker reliability is low, except when pure knowledge recall is tested. This is probably because the markers lack training in assessment. Paper reliability research is rare because it is difficult to do. A student would have to sit two very similar papers (Elton and Johnson 2002). Reliability, then, really matters with summative assessment, and this is not always easy to do.

“... reliability is greatest when assessments are simplest. There is a problem, then, if we aim to engender complex learning, because assessment processes that tend to simplify will undermine it” (Knight and Yorke, 2003:18).

They go on to quote Gibbs and Simpson (2002:1):

“The most reliable and cheat-proofed assessment systems are often accompanied by dull and lifeless learning that has short-lasting outcomes...” (Knight and Yorke, 2003:20).

Due to the prevalence of a positivistic approach to assessment, for example a reliance on statistical models of distribution, dominance has been given to reliability over validity: the dominance given to “fairness” over educational relevance. A constructivist approach to learning and assessment challenges the assumptions made by a positivist approach. For example, what is “truth”, the right answer according to whom? It questions the arbitrariness of divisions in classes, where over 70% is a First, and 69% is a 2:1; and the assumption that marks should be normally distributed, since good teaching ought to legitimately skew the distribution (Elton and Johnson, 2002).

To add to the complexity of dealing with assessment, it should be noted that reliability and validity are not mutually independent from each other. In high

stakes assessment, as in degrees, there are stakeholders (the QAA, employers) who demand reliability yet:

“Unfortunately , these are the very objectives which are of comparatively little importance in degree programmes, where it is usual to expect higher abilities and skills to be developed, and where increasingly learning programmes differ for different students” (Elton and Johnson, 2002 28).

Thus there has to be a trade off between reliability and validity. For example, exams are often used to deter plagiarism in higher education, resulting in a testing of recall rather than higher order thinking: description rather than critical analysis.

### **The changing nature of assessment**

“Assessment is probably the most important thing we can do to help our students learn. We may not like it, but students can and do ignore our teaching...” (Brown, S., 2004:81).

Why is assessment changing? There are various factors involved, but perhaps the most influential is the increasing diversity of the student population: increasing numbers of part time students, more mature students, and more from non-traditional backgrounds. At the same time:

“Recent developments have concentrated on forms of assessment, which take account the movement to use assessment in order to encourage active and student centred learning, which makes it difficult to assess all students in the same manner, as well as on the necessity in these days to make assessment more cost effective” (Elton and Johnson, 2002: 32).

This change advocates less focus on tutor-led approaches and more concentration on the learning outcomes and how learners may achieve them (Rust, 2002; Brown, 2004).

### **Conclusion**

The context in which undergraduates are assessed has changed. On the one hand there is a drive for employment related educational qualifications by

employers, funders, and parents, with mass higher education and more diverse students. On the other hand, there has been the increasing recognition that assessment needs to be formative and developmental, and that assessment needs to be aligned with desired learning outcomes. Unfortunately these are competing demands. The underlying reasons are probably to do with the two fundamentally different approaches to assessment: positivist and constructivist, with different theoretical and philosophical bases.

This is central to the understanding of assessment theory, as it relates to reliability, and validity. Positivists assume that student ability is fixed, and consistent. Constructivists view reality, especially social reality, as mental constructions, rather than absolute, objective truths. Whatever the approach adopted, assessment should involve a rational debate among the assessors and an agreement on how to answer a particular problem

## References

Biggs, J., (2003). *Teaching for Quality Learning at University* 2<sup>nd</sup> ed. Buckingham: Open University Press.

Brown, S., (2004). *Assessment for Learning. Learning and Teaching in Higher Education*. Issue 1. Pp 81-89.

Elton, L. and B. Johnston, (2002). *Assessment in universities: a critical review of research*. York: LTSN Generic Centre.

Jarvis, P., et al., (2005). *The theory & practice of learning*. 2<sup>nd</sup> edition. Abingdon: RoutledgeFalmer.

Knight, P. and M. Yorke, (2003). *Assessment, Learning and Employability*. Maidenhead: Open University Press.