

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

Purpose – Using the UK's recent disability legislation as a trigger, the paper explores issues in evidencing competence in the current context of using assessment as a tool for learning.

Design/methodology/approach - A modified version of the pragmatic method is used in which assessment theory is used to explore the tension in assessment which is both legislatively compliant and theoretically coherent.

Findings – In the current context of requiring assessment to be edumetrically sound, the legislation of competence standards is problematic in four respects. Task Validity is now a much more diffuse concept. Scoring Validity has to contend with many possible accommodations. Assessment Generalisability has to consider both the relevance and representativeness of the assessment task. Consequential Validity is essentially concerned with formative assessment; which is considered pedagogically important but, politically, is of less significance than summative assessment.

Research limitations – The analysis here is in terms of legislation within the UK and may not, therefore, refer to nuances of difference in the anti-discrimination legislation in other jurisdictions.

Practical implications – This study offers academics and administrators a framework within which to review the edumetric soundness of their assessment practices and policies. In so doing possible difficulties in equitable assessment can be made explicit. This, in turn, has implications for staff development.

Originality/value – The paper uses significant pedagogical theory to illuminate a legislative requirement and thereby contextualise difficulties in implementing equitable assessment.

Paper type - Conceptual Analysis

Keywords – edumetric assessment; authentic assessment, consequential validity

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

The Issue

The (2006) Further and Higher Education Amendment to the Disability Discrimination Act (1995) in the United Kingdom draws attention to the possibility of higher education institutions breaking the law when they (albeit unwittingly) exclude disabled students, either from participation or from success, by endorsing course requirements and/or academic standards which are spurious. Prior to the Amendment, while higher education establishments were legally obliged to make reasonable adjustments in the case of disabled students, institutions could insist that, following the reasonable adjustments, students should be able to demonstrate the achievement of certain academic standards. The Amendment has reduced the latitude that prevailed in describing and defending academic standards by introducing what it calls “competence standards”. According to the legislation (Disability Discrimination Act 1995 [Amendment] [Further and Higher Education] Regulations 2006) a competence standard must be for the purpose of determining whether a person has a particular level of competence or ability. While the Amendment pointedly draws attention to the undesirable consequence of treating disabled persons unjustly, its central tenet of justness for all is used to explore unresolved issues that attend the evidencing of competence in the current context of using assessment as a tool for learning.

Competence

Notions of competence have changed in recent years. Early notions lauded competence as outcomes-based achievement which was seen as democratic, transparent, supportive of learner autonomy and centred on necessary vocational skills. Unqualified support for this view crumbled in the face of evidence for its behaviourist overtones, its instrumental propensities and the unpredictability of student-centred learning (Barnett, 1994; Ecclestone, 2002; Jones, 1999; Tennant, 1997; Yorke, 1998). More recently competence has been described as a multifaceted concept which can refer to the:

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

- skills and abilities a person has developed;
- degree to which the person is effective in his/her interactions with the environment
- level of success of a person's performance (Schultheiss & Brunstein, 2005).

Either singly or in combination, each of these notions of competence – as ability, as effectiveness or as success – can be used to make judgements about self or other. The precise use of competence in any one situation or context will influence the psychological meaning that is exuded. The idea that competence is some profile of skills or overarching ability suggests some standard that is inherent in the particular task. The idea that competence is an interpersonal skill which may change over time suggests that it might be a general goal which energises other behaviour. The idea that competence is a measure of success suggests some outcome or grade and may invoke normative comparisons. Quite which, if any, of these distinctions was in the minds of those who developed the referent, competence standard, for the amendment to the Disability Discrimination Act is not clear. Sternberg (2005), however, argues that these different facets of competence are not mutually exclusive since all are brought into the service of "developing competence". For Sternberg, competence is the interface between intelligence or ability and expertise. The acquisition and consolidation of skills for performance at journeyman-level, to use Sternberg's nomenclature, is competence while the on going process of the acquisition and consolidation of skills for performance at high-level mastery is expertise. Ability, competence and expertise are on a continuum, along which one moves when developing broader and deeper levels of skill together with increased effectiveness in utilising the skills. Competence is therefore a malleable ability, goal or achievement to be developed under conditions of practice.

Determining Competence

Within formal education, student competence is determined through assessment.

Competence, as outlined above, is not directly observable but is an underlying abstract system which may or may not match observable performance. The determination of competence is

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

therefore by inference. In making the inference, we are touching on a central concern of assessment: how to ensure that the relationship between the observed behaviour and the internal, unobservable process (or the construct) is as good as it possibly can be. To the extent that we can claim a relationship, we can justify the validity of our assessment. Within the current assessment culture of integrating instruction and assessment (Birenbaum & Dochy, 1996) the formats for gathering information about students' achievements, and the processes through which such information is synthesised (in order to determine/diagnose competence) are much more extensive (both in number and range). There has been a consequent demand (Linn, Baker & Dunbar, 1991; Messick, 1989) that the traditional psychometric qualities (fairness and fitness-for-purpose) of assessment be extended (for example, Pellegrino et al, 2001), so that the edumetric quality (Dierick & Dochy, 2001) of providing the best possible opportunity for students to demonstrate their skills or achievements is also a criterion of competence. The purpose of this article is to tease out some potential tensions in evidencing competence in the current context of using assessment as a tool for learning.

METHOD

The argument posited here is constructed on the maxim of pragmatism, a method of reflection to render greater clarity, (Peirce, 1905) in which the regulative principles underlying particular practices are explored. This article calls on a modified version of the pragmatic method arising from Dewey (1910) in which he lists the steps in the approach as:

- A felt difficulty;
- Its location and definition;
- Suggestion of a possible solution;
- Development by reasoning of the bearings of the suggestion;
- Further observation and experiment leading to its acceptance or rejection (Dewey 1910, p.72).

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

The pragmatic method confronts issues, dilemmas or problems by tracing their respective practical consequences. What difference would it practically make to anyone if this notion rather than that notion were true? If no practical difference whatever can be traced, then alternative practices mean (practically) the same thing, and all dispute is idle. Whenever a dispute is serious, we ought to be able to show some practical difference that must follow from one side or the other's being right. The power of pragmatism is in the potential it offers to contemplate new ways of thinking about educational problems through reflecting on false dichotomies, unchallenged assumptions and traditional practices (Biesta & Burbules, 2003). In this instance the felt difficulty was the definition of competence, which is explored through an examination of pertinent literature on assessment theory. This points to tensions in determining competence standards that are compliant with the legislation but also offers a framework within which to review the appropriateness and possible impact of competence standards in different institutions.

EDUMETRIC QUALITIES OF COMPETENCE STANDARDS

Gielen, Dochy & Dierick (2003) outline four elements to be considered when judging the quality of new modes of assessment:

- Task Validity
- Scoring Validity
- Assessment Generalisability
- Consequential Validity

1 Task Validity

While all claims to validity rest on the notion that the task is a good measure of the characteristics being assessed, the new modes of assessment requires that tasks be authentic. In other words tasks need to represent accurately the real life problems that occur within the content domain being measured and they need to reflect the cognitive complexity required by

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

an expert working in the domain. According to Wiggins (1989, 1993), authentic assessment tests our intellectual ability, or our ability to understand. It is only when we can use our knowledge "wisely, fluently, flexibly and aptly in particular and diverse contexts" (Wiggins, 1993, p.200) that we can be said to understand. The reproduction of cued knowledge, the application of algorithms and the performance of drills and exercises do not of themselves reflect understanding: the quality that Wiggins argues is at the core of authentic assessment. Any number of assessment tasks could be authentic but the central criterion that any task must meet for it to be classed as authentic is its fidelity to the real world ways in which knowledge is used in the discipline or field of study into which the student is being inducted. But while authenticity can be evidenced in many ways, care needs to be taken in specifying the domain. If tasks are misspecified they may overestimate (if the task was insufficiently complex) or underestimate (if the task was overly complex) the student's competence. This issue can be particularly acute for a person who has limited competence in, say, the English language which is the language through which subject matter in say, mathematics is assessed. While traditional psychometric practices would require a tightly proscribed domain of competence in which assessment would be through standardised objective test-items, the edumetric practice is to provide open-ended tasks to encourage students to demonstrate the extent of their learning. However, by simplifying the task (say to display mathematical competence without unnecessary language interference) we are simultaneously removing the opportunity for the student to demonstrate understanding (because the limited demands of the language-free task preclude the evidencing of content-specific knowledge, which may be central to the domain). Assessing intellectual achievement can therefore be logistically problematic: on the one hand there is the legislative press that standards be only as complicated as strictly necessary and on the other that assessment tasks reflect real world tasks in all their complexity.

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

2 Scoring Validity

Underpinning scoring validity are three related matters:

- that the task is congruent with the instruction;
- that the task permits all students have equal opportunity to demonstrate competence;
- that the criteria are relevant, transparent and comprehensible.

Approaches that make assessment fair and accessible are a requirement of the United Kingdom's disability legislation. But tensions exist between this requirement and the requirement that tasks be authentic. While it would not be controversial to assert that an assessment task that required a student to dance, when the instruction had been on dance appreciation, was invalid (thereby avoiding inappropriate disadvantage to a wheelchair-user), the accommodations necessary for all students classed as disabled may not be so straightforward. For example, that a student is permitted to use a calculator rather than literally effecting computations either on paper or mentally may in some circumstances be reasonable. On the other hand, if it is necessary for students to apply their understanding of particular mathematical concepts, the facility to get the correct answers on a calculator will of itself be an incomplete performance. Authentic assessment requires longer tasks, more complex cognitive processing and deeper subject-matter knowledge. It also requires elaboration typically displayed through writing, according to standards (Newmann et al, 1996; Newmann, 1997) by which the assessment of intellectual achievement can be judged authentic.

Take in Table 1

There is considerable encouragement for higher education to increase diversity in its assessment methods (Yorke, 1998; Biggs, 1999), not least to avoid compromising the validity of the assessment. Given that writing tasks may be problematic for disabled students in higher education, it would not be unreasonable if tutors attenuated the writing demands they

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

make of students in favour of alternative media. However, whilst there are very good reasons for enabling students in the assessment process, it is important not to discount the value of writing in enabling learning. Bereiter and Scardamalia (1987) make a clear distinction between writing which records the thoughts that are already in one's mind (characterised as the knowledge-telling model) and writing which is shaped to achieve specific purposes and which reorganises one's knowledge in the process (characterised as the knowledge-transforming model). The knowledge-transforming model subsumes (but is not merely an elaboration of) the knowledge-telling model, and notwithstanding the everyday situations for which the knowledge-telling model is perfectly adequate, it is the knowledge-transforming model which is the more powerful because it enables us to rework our thoughts. Thus while writing is demanding and may make even more demands of a disabled student, there are very good pedagogical reasons why writing should be an integral part of coursework for many students. Because writing tasks may be problematic for some disabled students, the issue is not on *replacing* writing with an alternative medium, but on finding ways in which students' needs can be *accommodated* to the writing demand. Interpretation of the student's performance thus becomes very important.

In authentic assessment students can have considerable latitude in interpreting the stimulus task and constructing their responses, making for difficulty in the reliable interpretation of performance. The historically dominant method of interpreting performance in educational assessment has been by comparing the results of one individual with those of a well-defined reference group. While such norm-referencing usefully gives meaning to measures such as blood pressure or cholesterol level, it is arguably less useful in educational assessment because it does not describe students' actual achievements (Glaser, 1963; 1990). To redress this perceived deficiency, predetermined levels or standards of performance have become the basis for comparison in order to be able to provide explicit information as to what students can do. However, referencing interpretations of performance in terms of criteria is not unproblematic. The specification of what constitutes competence can become so precise and

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

elaborate that assessment task is reduced to a set of routine, algorithmic subtasks making no authentic demands of the student, and thereby negating the pedagogical and philosophical underpinnings of edumetric notions of validity (Wiliam, 1998). Evidencing competence in overly detailed and mechanistic 'standards' may encourage a hunter-gatherer approach to information (Ecclestone, 2002) at the expense of deep engagement in either process or content. In addition to issues in determining the criteria, there can be difficulty in their use. Although fairness is often seen as being resolved in the specification of clear criteria, consistency does not reside in external, pre-specified criteria (Wiliam, 1996) and so to believe that reliable marking is a function of specifying clear criteria is naïve. It is perhaps a commonly held view that the more concrete the criteria are, the more reliable the assessment. But if criteria are too concrete, they may merely stimulate students to focus on superficial, behavioural actions (Gulikers et al, 2006), which is contrary to the intentions of authentic assessment.

3 Assessment Generalisability

Assessment generalisability is concerned with the extent to which competent task performance can be generalised to the domain overall, invoking considerations of construct-relevance and construct representation (Messick, 1989). Construct-irrelevance occurs when extraneous clues or assessment formats permit or inhibit students to respond correctly or incorrectly in ways irrelevant to the construct being assessed. For example, if multiple choice items are constructed such that the distracters are all obviously shorter than the target response, or target responses follow regular ordinal patterns, respondents may be responding correctly in ways that are irrelevant to the construct being assessed. The most common construct irrelevant variables are ancillary skills and knowledge which can contaminate inferences about assessment performance, and contribute to inequitable assessment practice (Baker & O'Neil, 1994; Sackett et al, 2001). So, for example, the requirement to submit an assignment on-line could be a reasonable competence in a module on ICT, but would

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

probably be incompetent to require an on-line submission of an essay in philosophy. What this highlights is that skills which may be ancillary for one interpretation of assessment performance may be relevant for another interpretation, making it pedagogically inappropriate to legislate that particular assessment formats do/not necessarily discriminate against all disabled students. While within up-to-date understandings of assessment there is no reason to expect performance on one task to be similar to that on another, or to expect that assessment rubrics be standardised (since this very standardisation could preclude the assessment of important skills such as conceptualising a problem), the possibility of construct irrelevance contamination complicates the extent to which performance on 'equivalent' tasks can be considered comparable. Although information from multiple tasks may improve the validity of the judgement made, and although competence may be evidenced through a range of heterogeneous devices, there has to be consistency among independent measures intended as interchangeable (Moss, 1994). This is, and continues to be, a pressing issue in the light of accommodations that might be considered appropriate for disabled persons.

The issue of construct-representation is invoked when determining the nature of the assessment task(s). The requirement to demonstrate competence by doing or producing something means that the constituent task-specific skills of some performance may well constitute the evaluation criteria (Motowidlo et al, 1990; Russell & Kuhnert, 1992). If all that counts is the artefact or performance, then task-specific assessment can be perfectly adequate. So long as the assessment task elicits the skills underlying the performance in the domain of interest (as in acting, dancing, painting, participative sport and so on) there can be little quibble about the validity of the task. That the performance per se and the target of assessment are the same is known as task-driven assessment (Messick, 1994). But this is a theoretically pure characterisation, reminding us of why initial accounts of competence were inadequate: no matter how task-driven the performance is, it will invoke knowledge even if the knowledge is so embedded as to be a way of 'thinking and practising' (Hounsell & Hounsell, 2007). However, if making explicit some tacit aspect of task performance is a

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

necessary part of assessment (as, say, in commenting on the consistency and ingredient of paint in relation to the porosity of the wall to be painted) and if disabled students have difficulty in articulating their knowledge, there are implications for meeting anti-discrimination legislation. But even so, task-driven performance may not always be appropriate. Higher education is rarely concerned with one particular performance. If people are to learn to think, reason, plan and make good decisions (which is a significant aim of higher education), they must be able to generalise what they have learned in the past to new learning (Haskell, 2001). Therefore students may be assessed on the application of new knowledge to a range of novel situations or on the extension of their existing understanding into new knowledge: both situations being potentially unknowable and unpredictable.

Because of this need to generalise abstract concepts (Bereiter, 2002) from one situation to another, task-driven performance should not be a significant part of educational assessment. Rather, the concern to assess whether or not a person understands the underlying attributes or variables which represent the crucial components of the skilled performance (and thus draw on them at will) means that the performance assessment should be what Messick (1994) terms construct-driven (in which the knowledge, skills or other attributes to be assessed guide the selection of the task as well as the development of the scoring procedures).

Arguments for construct-driven assessment being preferable, because generalisable learning can get lost in task-driven assessment (Schavelson et al, 1992), nevertheless raise questions as to the type of evidence deemed sufficient for validity (Messick, 1989). Given that performances vary in what underlying cognitive processing they demand (Hamilton et al, 1997) and reveal (Baxter & Glaser, 1998), clarity in legitimate evidence of competence is critically important, yet difficult to achieve. Debates about the behavioural manifestations of abstract, psychological constructs such as higher order skills, problem solving and critical thinking (Baker et al, 1993) give rise to perceptions of relative importance (Messick, 1989). In the spirit of the disability legislation it is pedagogically unhelpful if the performance is conceived of as instrumentally rather than intrinsically important.

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

4 Consequential Validity

Consequential validity refers to the extent to which the actual and expected consequences of assessment match: in other words in what way(s) the effects of assessment influence the higher education student, curriculum, and assessment programme. The effects of assessment can be understood in three ways (Gielen et al, 2003). First the cognitive complexity of tasks can have a positive effect. By being sensitive to the demands of assessment (Entwistle, 2000) and following through with task appropriate behaviour such as deploying deeper-learning strategies, questioning content more critically, looking-up additional information students can progress their own learning. Second the information received by students, with and after completing assessment tasks, that is explicitly aimed at supporting and monitoring learning can be formative (Black & Wiliam, 1998). Feedback is key in formative assessment. However since few skills can be acquired satisfactorily simply through being told about them (Sadler, 1989), feedback is better understood as information that is used to alter the gap between one's actual or current level of performance and the goal, or reference level being aimed for. Third the transparency of, and the student's involvement in, the assessment process enables students to strive for domain understanding rather than, as within traditional modes, experience assessment as something that is done to them (Struyven et al, 2003). When students have experience of formulating criteria and using them to make judgements, they are developing the self-assessment that is necessary for self-regulation (Birenbaum, 2003) and authentic learning (Berlak et al, 1992; Gulikers et al, 2006).

However, the strength of match between expected and actual consequences of assessment can be quite slight, as there are a number of threats to consequential validity. The currently dominant view of higher education as one means of improving the country's social and economic goals means that it responds, at least in part, to particular political agendas causing the use of assessment for administrative purposes to acquire a primary rather than secondary purpose. In other words, external pressures may cause assessment to assume a primarily summative function. In such a context, the reliability of assessment becomes a dominant concern in order that society

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

can have confidence in the mechanisms through which inferences are made. This may proscribe the authenticity of the assessment task. Indeed this concern with reliability is clear in the Amendment's assertion that assessment "must be rigorous regarding standards so that all students are genuinely tested against a benchmark" (Section 9.30). As has been implied throughout, traditional notions of reliability can mean reducing the parameters of the assessment task to the point of sampling only a small part of the construct so that confident generalisations can be made. The potentially high-stakes nature of summative assessment, with its focus on high reliability, is no small threat to the anti-discriminatory intentions of the Code of Practice in the UK's (2006) Further and Higher Education Amendment to the Disability Discrimination Act.

Another threat to consequential validity is the assumption that formative function of assessment is necessarily motivating. In contemporary social-cognitive models of motivation (Pintrich & Schunk, 2002), motivation is dynamic, context sensitive and multifaceted. Motivation is actively shaped by students' perceptions of the control they have over the learning environment, their metacognitive processes, their perceptions of ability and their beliefs about the utility of effort (Birenbaum & Dochy, 1996; Prosser & Trigwell, 1999; Struyven et al, 2003). Cultural differences can also affect motivation: in some cultures motivation to learn can be assumed (Watkins, 2000) while in others, such as the United Kingdom, it may have to be fostered (Hidi & Harackiewicz, 2000). A final, and related, threat is in the oversimplification that formative assessment always leads to deep learning (Askham, 1997). Although effective formative assessment needs a constructivist or powerful learning environment (Birenbaum & Dochy, 1996), a dominance of didactic teaching, an emphasis on high-stakes assessment and a reluctance by students to relinquish their dependence on tutor-regulation of learning (Entwistle, 2000; Hounsell & Hounsell, 2007; Vermunt, 2007) can promote the surface apathetic approaches (Entwistle, 2000) that are so antithetical to new modes of assessment. The sheer range of diversity amongst persons (economic status, ethnicity, political affiliation, sex, religious persuasion) means that

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

assessment phenomena cannot be understood independently of the people operating therein and individuals cannot be understood independently of the assessment contexts in which they are required to perform (Blaine, 2007). While this diversity would be consistent with the underlying value position of the Disability Discrimination Act, the Amendment's recommendation that assessment "must also be flexible regarding the mode of measurement so that each student has an equal opportunity to demonstrate their competence" is not without difficulty in its application.

Conclusion

In answering the question, "what is a competent competence standard?" a pragmatic approach has been taken to explore what might be problematic in the UK legislation's specification. The theoretical literature on assessment would suggest that a competent competence standard is one that is edumetrically sound. Edumetrically sound assessment is authentic, and cognitively complex. It can, theoretically, be evidenced in a variety of ways. Because of the political significance of assessment to be an agent of educational reform (Linn, 2000); the case for equity in assessment is incontrovertible. Not only is inequitable for there to be irrelevant barriers to accessing assessment, the Amendment, in its determination of 'reasonable adjustment', requires that steps be taken to compensate for the various ways in which disabled persons might be disadvantaged by traditional assessment practices. If educational assessment was an uncontested area of theory and practice, the legislation's introduction of competence standards might have much to recommend it. However, the considerable advances in understanding of how people think and learn, and the concomitant growth in the assessment culture mean that assessment, learning and teaching are now widely regarded as complex and contestable. This suggests that the fundamental educational aim of fostering the capacity of all to participate, democratically, requires disabled persons be party to deliberations on how competence is to be evidenced. The motivation for enabling the many alternatives possible in edumetrically sound assessment is clear but so too is the

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

tension: that of comparability of results from alternative formats. Efforts to make assessment more just for individual students through offering different assessment tasks and/or different performance conditions may reduce the justness of comparisons among students. For as long as we have to describe and differentiate between the achievements of students we are involved in a process of measurement. This is inherently a flawed process which can be attenuated but never fully ameliorated. Therefore, what ever promise the (2006) Further and Higher Education Amendment to the Disability Discrimination Act (1995) held out for equity in the assessment of disabled students, it places considerable demands on the system.

REFERENCES

- Askham, P. (1997), "An instrumental response to the instrumental learner: assessment for learning", *Studies in Educational Evaluation*, 23 (4) pp. 299-317
- Baker, E., O'Neil, H. & Linn, R. (1993), "Policy and validity prospects for performance-based assessment", *American Psychologist* Vol. 48, pp.1210-8.
- Baker, E., & O'Neil, H. (1994) "Performance assessment and equity", *Assessment in Education* Vol. 1 No. 1, pp.11-26.
- Barnett, R. (1994) *The Limits of Competence*, SRHE & The Open University Press, Buckingham
- Baxter, G. & Glaser, R. (1998), "Investigating the cognitive complexity of science assessment", *Educational Measurement: Research and Practice*, Vol. 17 No. 3, pp.37-45.
- Bereiter, C. & Scardamalia, M. (1987), *The Psychology of Written Composition*, Lawrence Erlbaum Associates, NJ.
- Bereiter, C. (2002), *Education and Mind in the Knowledge Age*, Lawrence Erlbaum Associates, NJ.

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

Berlak, H., Newmann, F, Adams, E., Archbald, D., Burgess, T., Raven, J. & Romberg, T.

(1992), *Towards a New Science of Educational Testing and Assessment*, University of New York Press, New York State.

Biesta, G. & Burbules, N. (2003), *Pragmatism and Educational Research*, Rowman & Littlefield Publishers, Lanham MD.

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

Birenbaum, M. & Dochy, F. (Eds.) (1996), *Alternatives in Assessment of Achievement, Learning Processes and Prior Knowledge*, Kluwer Academic Press, London.

Birenbaum, M. (2003), "New insights into learning and teaching", in Segers, M., Dochy, F. & Cascallar, E. (Eds.), *Optimising New Modes of Assessment: in search of qualities and standards*, Kluwer Academic Press, Dordrecht, pp.13-36.

Black, P. & Wiliam, D. (1998), "Assessment and classroom learning", *Assessment in Education*, Vol. 5 No. 1, 7-74.

Blaine, B. (2007), *Understanding the Psychology of Diversity*, Sage Publications, London.

Dewey, J. (1910), *How We Think*, D.C. Heath, Lexington, Mass.

Dierick, S & Dochy, F. (2001), "New lines in edumetrics: new forms of assessment lead to new assessment criteria", *Studies in Educational Evaluation* 27, pp. 307-329.

Ecclestone, K. (2002), *Learning Autonomy in Post-16 Education*, Routledge-Falmer, London.

Entwistle, N. (2000) "Approaches to studying and levels of understanding: the influences of teaching and assessment", in Smart, J. & Tierney, W. (Eds.) *Higher Education: Handbook of Theory and Research*, Volume XV, Agathon Press, New York, pp. 156-218.

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

Gielen, S., Dochy, F. & Dierick, S. (2003), "The influence of assessment on learning", in

Segers, M., Dochy, F. & Cascallar, E. (Eds.), *Optimising New Modes of Assessment: in search of qualities and standards*, Kluwer Academic Press, Dordrecht, pp.37-54.

Glaser, R. (1963), "Instructional technology and the measurement of learning outcomes: some questions", *American Psychologist*, Vol. 18, pp.519-21.

Glaser, R. (1990), "Toward new models for assessment", *International Journal of Educational Research*, Vol. 14 No. 5, pp.475-83.

Gulikers, J., Bastiaens, T., Kirschner, P. & Kester, L. (2006), "Relations between student perceptions of assessment authenticity, study approaches and learning outcome", *Studies in Educational Evaluation*, Vol. 32, pp. 381-400.

Hamilton, L., Nussbaum, E., & Snow, R. (1997), "Interview procedures for validating science assessments", *Applied Measurement in Education*, Vol. 10 No 2, pp.181-200.

Haskell, R. (2001), *Transfer of Learning*, Academic Press, London.

Hidi, S. & Harackiewicz, J. (2000), Motivating the academically unmotivated: a critical issue for the 21st century, *Review of Educational Research*, 70 (2), pp. 151-179

Hounsell, D. & Hounsell, J. (2007), "Teaching-learning environments in contemporary mass higher education", *British Journal of Educational Psychology*, Monograph Series II, 4, pp 91-111.

Jones, A. (1999), "The place of judgement in competency-based assessment", *Journal of Vocational Education and Training*, 51 (1) pp. 145 160.

Linn, R., Baker, E. & Dunbar, S. (1991), "Complex performance-based assessment: expectations and validation criteria", *Educational Researcher*, Vol. 20 No. 8, pp.15-21.

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

Linn, R. (1999), Validity standards and principles on equity in educational testing and assessment, in Nettles, A. & Nettles, M. (Eds.) *Measuring Up: Challenges Minorities Face in Educational Assessment*, Kluwer Academic Publishers, The Netherlands pp. 13-33.

Linn, R. (2000), "Assessments and accountability", *Educational Researcher*, Vol. 29 No 2, pp.4-16.

Messick, S. (1989), "Validity", in Linn, R. (Ed), *Educational Measurement*, Macmillan, New York, pp. 13-103.

Messick, S. (1994), "The interplay of evidence and consequences in the validation of performance assessments", *Educational Researcher* 23(2), pp.13-23.

Moss, P. (1994), "Can there be validity without reliability?" *Educational Researcher*, Vol. 23, pp.5-12.

Motowidlo, S., Dunnette, M. & Carter, G. (1990), "An alternative selection procedure: the low-fidelity simulation", *Journal of Applied Psychology* Vol. 75, pp.640-7.

Newmann, F., Marks, H. and Gamoran, A., (1996) "Authentic pedagogy and student performance", *American Journal of Education*, 104, pp.280-312.

Newmann, F. (1997), "Authentic assessment in social studies", in G. Phye (Ed), *Handbook of Classroom Assessment*, Academic Press. London. pp.359-380.

Peirce, C. (1905), *Collected Papers of Charles Sanders Peirce*, Harvard University Press. Cambridge.

Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.) (2001). *Knowing what Students Know: the science and design of educational assessment.*, DC: National Academy Press, Washington

Pintrich, P. & Schunk, D. (2002), *Motivation in Education*, Prentice-Hall Merrill, New Jersey.

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

Prosser, M. & Trigwell, K. (1999), *Understanding Teaching and Learning*, The Society for Research into Higher Education & Open University Press, Buckingham.

Russell, C. & Kuhnert, K. (1992), "New frontiers in management selection systems: where measurement technologies and theories collide", *Leadership Quarterly* Vol 3, pp.109-36.

Sackett, P., Schmitt, N., Ellingso, J. & Kabin, M. (2001), "High stakes testing in employment, credentialing and higher education", *American Psychologist* 56(4), pp.302-18.

Sadler, R. (1989), Formative assessment and the design of instructional systems, *Instructional Science*, 18, pp. 19-144

Schavelson, R., Baxter, G. & Pine, J. (1992), "Performance assessments: political rhetoric and measurement reality", *Educational Researcher* 21(4), pp.22-7.

Schultheiss, O. & Brunstein, J. (2005), "An implicit motive perspective on competence", in Elliot, A. & Dweck, C. (Eds.), *Handbook of Competence and Motivation*, The Guilford Press, New York, pp.31-51.

Sternberg, R. (2005), Intelligence, competence and expertise, in Elliot, A. & Dweck, C. (Eds.), *Handbook of Competence and Motivation*, The Guilford Press, New York, pp.15-30.

Struyven, K., Dochy, P. & Janssens, S. (2003), Learners' perceptions about new modes of assessment, in Segers, M., Dochy, F. & Cascallar, E. (Eds.) *Optimising New Modes of Assessment: in search of qualities and standards*, Kluwer Academic Publishers, London, pp.171-223.

Tennant, M. (1997) *Psychology and Adult Learning*. Routledge, London

Vermunt, J. (2007), The power of teaching-learning environments to influence student learning, *British Journal of Educational Psychology*, Monograph Series II, 4, pp 73-90.

What is a competent 'competence standard'? Tensions between the construct and assessment as a tool for learning

Watkins, D. (2000), "Learning and teaching: a cross-cultural perspective", *School Leadership and management*, 20 (2) 161-73.

Wiggins, G. (1989), "Teaching to the (authentic) test", *Educational Leadership*, 46(7), pp.41-7.

Wiggins, G. (1993), "Assessment: authenticity, context, and validity", *Phi Delta Kappan*, 75, pp.200-14.

Wiliam, D. (1996), "Meanings and consequences in standard setting", *Assessment in Education* 3(3), pp.287-307.

Wiliam, D. (1998), Construct-referenced assessment of authentic tasks: alternatives to norms and criteria, Paper presented at the 24th Annual conference of the International Association for Educational Assessment – Testing and Evaluation: Confronting the Challenges of Rapid Social Change, Barbados, May 1998.

Yorke, M. (1998), "Assessing capability", in Stephenson, J.& Yorke, M. (Eds.), *Capability and Quality in Higher Education*, Kogan Page, London, pp.174-191.

Table 1 Standards of Authenticity in Intellectual Achievement

1. Analysis - The response reflects higher order thinking with content by organising, synthesising, interpreting, evaluating and hypothesising to produce comparisons, contrasts, arguments, application of information to new contexts, and consideration of different ideas or points of view.
2. Disciplinary Concepts - The response reflects an understanding of ideas, concepts, theories and principles that are central to the academic or professional disciplines into which the students is being inducted.
3. Elaborated written communication – The response explains understandings and conclusions. It is clear, coherent and provides richness in details, qualifications and argument.

Source: Adapted from Newmann, Marks & Gamoran (1996) and Newmann (1997)