

‘You can see the quality in front of your eyes’: grounding academic standards between rationality and interpretation.

Prof. Sue Bloxham, University of Cumbria.

Forthcoming in *Quality in higher Education*

Abstract

This article considers the failure of theory to provide a workable model for academic standards in use. Examining the contrast between theoretical perspectives, it argues that there are four dimensions for which the academy has failed to provide an adequate theoretical account of standards: documented or tacit knowledge of standards, norm or criterion-referenced grading, analytical or holistic judgement processes and broad or local consensus on standards. It concludes that whilst a techno-rational perspective poorly represents the actual practice of standards in use, alternative, interpretivist accounts do not satisfy demands for reliability, transparency and fairness. It concludes by outlining an alternative framework for safeguarding standards: systematising a range of processes for learning about and safeguarding standards particularly for new staff, reviewing the role and potential of documented standards, building staff awareness and assessment literacy and establishing trust in standards by students and other stakeholders.

Introduction

Academic standards are at the cornerstone of university education, a fundamental basis for universities’ credibility in the world. However, concerns about academic standards and grade inflation are widespread. Yet academic standards are poorly researched and understood, particularly in their everyday use by academics, managers and those involved in quality assurance. Sadler (2011) noted that the question of what is meant by ‘academic standards’ is rarely asked, never mind answered and Yorke (2008, p. 83) described it as an ‘elusive concept’. There are different theoretical perspectives for examining standards in use: on the one hand, the university sector has created quality assurance frameworks for standards that are underpinned by a techno-rational approach whilst, on the other hand, research and theory from a range of social science disciplines draws on alternative theoretical perspectives which emphasise the co-constructed, interpretive and local nature of standards. This paper argues that this bifurcation between a techno-rational and alternative, broadly interpretivist, theoretical perspectives for standards in use, which are elaborated below, is not helpful to academics in the day-to-day application of academic standards. Neither approach represents standards in everyday use nor provides a credible model that will reassure students and others that standards are fair and safeguarded. Following discussion of the weaknesses in existing theory, the article advocates a future research agenda to investigate how the higher education sector might create a framework of processes to build confidence and consistency into its academic standards in a way that coincides with knowledge of professional judgement as complex and tacit.

Defining the term: academic standards

It is necessary to commence with a brief discussion of the definition of academic standards. Authors have noted that academic standards are conceptually complex and difficult to define (Coates, 2010; Woolf and Cooper, 1999). Middlehurst (1996) has described the term as composite, including input, process and output elements. Harvey (2002) distinguished *academic* standards from standards of competence, service standards and organisational standards by centring his definition on academic attainment. This definition is similar to that of Alderman (2009, p. 12) who referred to academic standards as 'discrete levels of intellectual performance, the attainment of which results in the award of academic credit' and Sadler (1989, p. 129) who described them as 'a designated degree or level of performance or excellence'. This output approach is also evident in the definition used by higher education quality agencies, for example in Australia (Coates, 2010) and the United Kingdom (UK) (QAA, 2010). Based on this consensus, this article will use the term 'academic standards' as related to prescribed and recognised levels of student attainment.

The techno-rational perspective on assessment

Customary descriptions of academic standards hold conceptions of standards which lie in a techno-rationalist perspective on assessment; a standpoint that promotes the use quantitative data and measurement to ensure accountability (Kappler, 2004). It is based on an epistemology which holds that 'knowledge is monolithic, static and universal' (Delandshere, 2001, p. 127) with staff acculturated into 'guild knowledge' (Sadler, 1989). Shay (2008), writing in the South African context, argues that primacy is given to assessment 'technology', such as criteria, in current assessment (Shay, 2005, p. 676). On the basis of this theoretical perspective, although statements of standards were rarely articulated in the past, there was an assumption pre-massification (Silver and Williams, 1996) that a 'gold standard' existed; fixed benchmarks that enshrined the standards of the ancient universities. Such a perspective sees standards as independent of the individuals who created or are custodians of them. As Leach *et al.* (2001, p. 295) argue 'The prevailing hegemony [in assessment] holds that there are bodies of knowledge that are universally true, invests power in the teacher and has faith in scientific notions of objectivity and reliability'. Yorke (2011, p. 252) linked this perspective to a 'measurement' model and a 'realist' approach to assessment.

The techno-rational perspective can be seen most visibly in the processes that have been put in place to support the development and maintenance of academic standards through a range of quality assurance frameworks and measures. Alderman (2009) argued that historically academic standards were sustained over time by an oral tradition through contact between universities and subject communities, what Brennan (1996) referred to as a private and implicit approach to communicating standards. However, such 'shared understandings of an academic elite are insufficient as a basis for standards in a mass system of higher education' (Brennan, 1996, p. 16), a point reinforced more recently by Salmi (2009) and the last twenty years have witnessed considerable work attempting to make academic standards more transparent and consistent. In the UK this has included development of an 'Academic Infrastructure' including efforts to codify standards through a Framework for Higher Education (QAA, 2010).

These developments have had a focus on both the scope and level of higher education programmes. They have brought into the public domain greater information about the stated course content, aims, assessment methods and criteria of programmes of study. They have attempted to make previously elusive standards more explicit through

techniques such as subject benchmarking (Jackson, 1999). Threshold standards are now common in the professions (Gaunt, 1999) where statements of knowledge, skills and values that students must demonstrate in order to gain their professional recognition are normal practice in many national systems and often predate the growth in higher education quality assurance over the last twenty years. Statements of standards are also increasingly available to assessors and students in the form of rubrics (marking schemes) and assessment criteria and it is in relation to the latter where most research on academic standards in use (for example, grading, consistency of judgement and communicating standards) has taken place.

Alternative perspectives on academic standards

There are alternative theoretical perspectives that argue that we cannot establish standards independently from the individuals who are using them. This tradition includes a socio-cultural perspective that 'emphasises the situatedness of practice' and a social constructivist approach that 'emphasises the constructedness of knowledge' (Shay, 2008, p. 596). These philosophical perspectives are interpretive and conceive of staff as active agents in co-creating standards through local assessment practices and communities. As Shay (2005, p. 669) eloquently explained, the interpretive framework that each examiner holds combines objective elements learnt from participation in the relevant field and subjective elements arising from the local context of assessment and dependent on the assessor. 'From this perspective assessment is situationally contingent, rooted in local cultures and reliable and robust only in terms of sets of assumptions, attitudes and values which are, in part at least, localized'. Yorke (2008) linked this theoretical perspective to a 'judgement' model and a 'relativist' approach to assessment, which challenges the potential for shared and stable standards.

In relation to marking, writers have argued that lecturers learn their standards through an informal and differentiated process involving active participation in relevant communities and practices (Shay, 2005, Orr 2010). The socially situated nature of this learning creates the potential for individual differences in marking judgements (Crisp, 2008; Broad, 2003; Hunter & Docherty, 2011). Studies have suggested that lecturers' standards are influenced by their values (Leach *et al.*, 2001; Ashworth *et al.*, 2010; Morgan & Wyatt-Smith, 2000), access to specialist knowledge (Shay, 2005), socialisation processes, relationships with students (Shay, 2004; Orr, 2008), the other social worlds they inhabit (den Outer & Handley, 2010), their history and previous experience (Shay, 2005; Dobson 2008a; Milanovic *et al.*, 1996). Shay (2005, p. 664) suggested 'differences between markers are not 'error', but rather the inescapable outcome of the multiplicity of perspectives that assessors bring with them'.

Researchers working within these alternative theoretical perspectives have argued that individuals construct their own standards' frameworks (Ashworth *et al.*, 2010). Such highly complex frameworks represent how various influences combine to create a unique lens through which each tutor reads and judges student performance. Standards' frameworks are dynamic; constructed and reconstructed through involvement in communities and practices including engagement with student work, moderation and examiners' feedback (Sadler, 1989; Crisp, 2008; Bloxham *et al.*, 2011). Thus an assessors' standards' frameworks leads them to focus on different aspects of student work (Elander & Hardman, 2002), for example their first impressions (Mullins & Kiley, 2002) or presentational features (Hartley *et al.*, 2006; Kangis, 2001). This conceptualisation is

supported by work in the broader fields of research on university assessment and marking where similar notions have been theorised such as 'teachers conceptions of quality' (Sadler, 1989, p. 127), assessors' 'evaluative frameworks' (Broad, 2003), 'assessors' interpretive frameworks' (Shay, 2005, p. 665) and 'pre-formed knowledge structures' (Crisp, 2008, p. 250). Such perspectives emerge from different academic disciplines but they all emphasise the tacit, interpretive nature of standards, learnt informally through active participation in relevant communities and practices.

The subjectivity implied by these personal frameworks is considered to be mediated by membership of an academic community with its mutual engagement and joint enterprise creating 'a shared repertoire of terms, knowledge, understanding, tacit conventions and practices' (Crisp, 2008, p. 249); Sadler's (1989) 'guild knowledge'. Therefore, standards' frameworks, whilst unique to individuals, are grounded in the broader academic standards of the relevant academic community. Nevertheless, their individualised nature has an important impact on how lecturers perceive student work and the feedback they offer.

Thus, from this perspective, shared and consistent notions of standards that can withstand tests of reliability and objectivity, as portrayed by the techno-rational perspective, deny the necessarily elusive nature of academic standards in use that are continuously co-constructed by academic communities and extremely difficult to explain to a lay audience. Such standards are not 'immutable' (Sadler, 1987, p. 196) or independent of individuals (Shay, 2005, p. 677) and their tacit nature means that it is difficult to detect drifts in standards (Sadler, 1987). This perspective significantly challenges the assumption that there is a community consensus regarding national disciplinary standards that has enduring stability over time and across contexts (different staff, programmes and universities).

Standards in use

Overall, these two broad theoretical approaches to academic standards sit in clear contrast to each other. One sees standards as fixed, capable of being reliably interpreted, analysed and communicated. The other conceptualises them as shifting, tacit and holistic. In exploring these alternative approaches a series of dimensions of standards in use can be identified. These dimensions are: knowledge of standards; the referencing of standards; judgement processes and the nature of consensus (Figure 1). This paper argues that in the contrast between the techno-rational and interpretive emphasis in these dimensions, there is a failure to provide an adequate theoretical account of standards in use; an account that provides a firm basis for practice and that is understandable and credible to students.

Taking the second dimension as an example, whilst programmes now generally claim to draw on explicit criteria or rubrics to grade students' work, research suggests that lecturers rely on a level of norm referencing in order to ground their interpretation of criteria and develop their sense of appropriate standards (Orr, 2008). The academy does not have a theory that tackles this disparity and provides lecturers and students with an alternative theory of standards' referents that can encompass the false division between criterion- and norm-referencing. Indeed, as Orr (2008) noted, text book guidance for lecturers can swing from exhorting objectivist approaches in relation to explicit learning outcomes to recognising and discussing tacit practice. They offer no middle ground.

Figure 1: Dimensions of standards in use

Dimension	Techno-rational perspective	Interpretive perspective
Knowledge of standards	Explicit documentation of standards	Internalised, tacit standards
Referencing of standards	Criterion-referenced grading	Norm referencing of judgement needed
Judgement processes	Analytical judgement processes	Holistic professional judgement
Nature of consensus on standards	Broad consensus on standards possible	Individualised standards or localised consensus

Advocating one perspective over another is not an option. The notion that universities might abandon the panoply of quality assurance concepts and artefacts that are derived from the techno-rational approach in favour of an explicit return to a more tacit and localised interpretation of standards is inconceivable in the current climate of accountability. Indeed, in the UK, transparency for students has been part of the general direction of policy for many years, including that of the 2010 Coalition government. It places increasing stress on students as well-informed consumers of higher education. Therefore, although research indicates that there are major weaknesses with the techno-rational approach, most of all its failure to represent how academics make their academic judgements (Orrell, 2003; Sadler, 2009a), replacing that approach with a simple reassurance that standards are secure because assessors are steeped in their disciplinary communities will not stand up to scrutiny either.

In a period of growing pressure and accountability surrounding academic standards, the need to defend academics' judgement processes will become increasingly important not least for students. The procedures for quality assuring standards may, on the face of it, imply a robustness to assessment practice but the evidence of individuals' varying standards' frameworks provides limited confidence in the consistency and comparability of academic standards more widely. This 'softness' (Sadler, 2005) of assessment data leaves the academy vulnerable to challenge; lying as it does barely hidden beneath the surface appearance of reliable and robust procedures. Therefore, the academy needs a scholarship of professional judgement in academic grading and its quality assurance that has credibility to both the assessors and the assessed; where the expressions of academic standards presented to students accurately reflect the formulations of standards used by lecturers themselves. The following sections explore that challenge by examining each of the dimensions of standards in use (Figure 1) and then conclude by proposing a research agenda for developing theory and practice of standards in use.

Dimension 1: Explicit to tacit knowledge of standards

The alternative perspectives discussed above argue firmly for the tacit nature of academic standards and that explicit statements of standards promise more than they can deliver (Hawe, 2002). Explicit documentation falls into two broad groups; first, that produced at the system, discipline or profession level and, second, those that are usually determined at the institution, programme or course level. They will be considered separately here.

In relation to the first group, explicit documentation produced at the system, discipline or profession level, Moss and Schutz (2001) have provided a useful exploration of the creation and use of explicit standards for judgement, drawing on standards for

qualifying teachers in the USA as their exemplar. Their philosophical argument is based on Habermas' view that the moral justification for laws (or in this case standards) comes from the fact that they are created through rational consensus with relevant representatives of the community involved in drawing them up. This has been the rationale underpinning many groups tasked with writing or reviewing standards such as Subject Benchmark Statements and statements of threshold standards for professions. However, they argue that once standards are written, the system operates outside the control of the standard setting group and this presumes that the interpretations of the words will not fall outside that which was determined by the community consensus that drew them up. In addition, Moss and Schutz argued that consensus seeking leads groups to produce standards that are too general or abstract, hides complexity and can mask diversity as those involved strive to complete the task in a limited time frame. What is produced may be a compromise rather than a consensus and even with elaboration, provides plenty of room for different interpretations. Thus, they argue that judgement becomes an 'interpretation' of standards not a 'reflection' of them. Hawe (2002) in New Zealand and Martin and Cloke (2000) in the UK found similar problems with standards for qualifying teachers and, as Yorke (2008, p. 204) has stated: 'prescriptions on their own did not carry sufficient meaning for consistency in their application across a range of different contexts'.

A further difficulty with explicit professional standards, as discussed by Yorke (2008) and Tang (2008) is that, using a techno-rational approach, they tend to negate the impact of context. The importance of professional judgement becomes obvious where assessors decisions need to take account of the particular elements and difficulties of the context in which the assessment takes place. Standards that can operate without reference to context are probably not actually useful in judging future professional competence. As Moss and Schutz (2001, p. 61) point out, context and standards together 'co-determine' the interpretation.

In relation to the second group, explicit documentation produced at the institution, programme or course level, similar problems occur when attempting to make standards explicit through assessment information such as rubrics, assessment criteria and grade descriptors. Although common parlance gives credence to these 'validated evaluative tools' (Ross, 2009, p. 475) as something that will lead to comparability and reliability of standards, the evidence is less reassuring. Broad (2003, 2000), in the context of writing assessment in the USA, and others have thoroughly explored the purposes and difficulties of making standards explicit in this way (Sadler, 2009a; O'Donovan *et al.*, 2008). In particular, the interpretive nature of judgement in using rubrics at this level renders them weak as a means of assuring consensus.

Furthermore, research on grading found that a significant proportion of academics did not physically refer to criteria or learning outcomes whilst marking and used them, if at all, at the end of the marking process to check or confirm their holistic judgement (Bloxham *et al.*, 2011). Orrell (2008) found little espoused concern by lecturers about using assessment criteria or other ways of achieving accuracy and consistency in their grading and did not tend to refer to written criteria. Research has also shown that lecturers interpret rubrics differently (Delaney, 2005). Indeed Shay (2004) found that staff said that they could not use a rubric and found it hard to articulate what they are doing.

On the other hand, academic standards in use may be changing as many quality assurance processes for assessment have become normative. More recent research (Orr 2010) suggests that lecturers' personal experience is anchored in or merged with written

learning outcomes and assessment criteria to make effective judgements (Orr 2010). Indeed Boyd and Bloxham (2011), using relatively recent data, found that a proportion of assessors drew on explicit information in some way to give confidence to their judgements. In these latter cases, it could be argued that explicit standards are now contributing in some way to determining individuals' tacit standards frameworks in a way that was less evident before.

In conclusion of this discussion, it is instructive to note that 25 years ago, Sadler (Sadler, 1987) pointed to the 'fuzziness' of standards. He argued that 'fuzzy' standards cannot be transformed into sharp standards by simply using more detailed language. However, this leaves a dilemma for lecturers in using and representing standards to students. How can theory help bridge this divide between the explicit and the tacit in the everyday act of judging student performance?

Dimension 2: Criterion-referenced or norm-referenced standards

Policies and practices to secure academic standards continue to rest on a fairly 'objectivist' rationality (Shay, 2004). This frequently includes a presumption of criterion referencing against published standards. Most guidance to lecturers advocates that they should base their assessment on criterion rather than norm referencing so that a student is judged against a set of standards, not against his or her peers. This distinction has been criticised (Orr, 2010; Neil & Wadley, 1999; Yorke, 2009). In particular, Yorke has made the point that assessors' grading behaviour is tacitly influenced by norm referencing and Vaughan (1991), Delaney (2005), Orrell (2008), and Crisp (2008) also found that lecturers draw on their knowledge of different students' work in order to make their judgements.

One explanation of the continued influence of norm referencing is that staff are unable to interpret semantically 'loose' criteria without some kind of referent and students' work is crucial in that role. There are wider implications for this in relation to warranting standards across institutions. This has been noted in critiques of external examiner systems which are an explicit effort to test local standards against national norms. In the UK, examiners for any given university tend to be drawn from the same type of university and it is argued that the influence of norms in their own 'mission' group will limit their ability to calibrate standards more widely.

Norm referencing is also problematic as Orrell (2008, p. 259) points out because 'the qualities of other students' performances.... do not provide a stable basis for maintaining standards because as a basis for grade decisions they are unpredictable and highly variable'. Yet an explicitly criterion-referenced approach located in the techno-rational tradition appears too difficult for the complex judgement that characterises most higher education assessment (Sadler, 2009a). In its place, a theoretical perspective is lacking that tackles this disparity and provides lecturers and students with an alternative theory of standards' referents which can encompass the false division between criterion- and norm-referencing.

Dimension 3: Analytical or holistic judgement processes in marking

This dimension is intrinsically connected to the first dimension: knowledge of standards. The typical academic in many systems is now faced with (and may contribute to creating) various artefacts to guide their academic standards in use, for example assessment criteria, grade descriptors and rubrics. Such artefacts tend to represent standards in an analytical form as in a typical set of assessment criteria or learning outcomes. However, empirical and theoretical studies have emphasised the extent to which marking judgements are holistic (Sadler, 1989, Crisp, 2008; Sadler, 2009a, 2009b; Grainger *et al.*, 2008). Grainger, Purnell and

Zipf (2008) found that staff work backwards from an holistic judgement, awarding marks to individual criteria afterwards. Likewise, Bloxham *et al.*, (2011) observed that assessment criteria were largely used to check or confirm holistic judgement at the end of the marking sequence. Orr (2010) noticed that criteria were used more for adjudication in difficult cases rather than as guidance to the marker and Elander and Hardman (2002) found that marks given for individual assessment criteria are not statistically independent of each other. In relation to judgement of professional performance, Tang (2008) found that judgements about student teachers' performances are made holistically rather than by reference to the standards. Further discussion and evidence regarding the practical failure of the analytical approach to grading can be found in Yorke (2011).

Accordingly, this dimension describes a further gap between the approaches as advocated by the techno-rational and alternative perspectives. Whilst lecturers can, *post hoc*, rationalise their holistic judgement against analytical referents, it would be preferable to develop a theoretical approach that more honestly reflected actual practice and also provided confidence in that practice.

Dimension 4: Consensus in standards

Central to the professional discourse of standards is the idea that there can be a consensus in standards across staff, departments and universities, if not systems. Whilst the techno-rational perspective increasingly locates that consensus in 'transparent' statements of standards and accountability through quality assurance processes, alternative approaches have argued that tutor subjectivity is mediated through membership of an academic community. Whilst the latter approach does not see standards as 'fixed', it does offer a level of intersubjectivity and continuity as standards are shaped by the norms of that community.

However, in examining this assertion of consensus, numerous studies over time have indicated the low agreement between higher education assessors (Leach *et al.*, 2001; Elander & Hardman, 2002; Sadler, 1987; Wolf, 1995). Empirical studies (Broad 2003; Dobson, 2008a; Greateorex, 2000) have found that assessors use personal criteria beyond or different to those stated. Wolf (1995) argued that this variation does not mean that assessors are marking at the incorrect standard but are using other evidence (for example, about the person, other work) because of the variability of assessed tasks. However, Dobson (2008b) suggested that this use of contextual information about candidates is a threat to standards and Milanovic *et al.* (1996) pointed to the marker as a recognised source of error. 'Standards refused to be as solid, stable and portable an entity as participants wished' (Broad, 2003, p. 74) and, as Moss and Shultz (2001) found, even where assessors agreed pass and fail, they did so for different reasons.

Indeed, Knight (2006) argued coherently that a consensus regarding standards can typically only be held at the local level and studies on how assessors learn to mark confirms the influence of the local academic community in helping novice assessors learn appropriate standards (Jawitz, 2009). So whilst it might be argued that individuals' personal standards frameworks influence and are influenced by local 'guild knowledge', this provides limited confidence in the consistency and comparability of academic standards more widely than the local team. Overall, the notion of broad comparability of standards within subject disciplines and across universities is yet to be empirically demonstrated.

A new research agenda for academic standards?

In summary then, the alternative perspectives, whilst recognising the socially constructed and interpretive nature of academic standards inherently accept the normalisation of bias, changing standards, norm referencing and other features of professional judgement that generate concern. On the other hand, the foregoing discussion suggests that the techno-rational approach, with its beliefs in reliability and transparent knowledge of standards, poorly represents the actual practice of grading in higher education. Neither theoretical approach serves as an effective basis for securing standards in universities, giving confidence to students or supporting university teachers in their exercise of professional judgement in assessment.

However, it could be argued that the quality assurance machinery that has emerged from a techno-rational conceptualisation does provide a basis for trust in standards amongst the wider community and certainly at the official level of institutional audit. In addition, research (Boyd and Bloxham, 2011) has shown that those quality assurance processes do influence and give confidence to some academics' judgement. The research involved a mixed-method study of university teachers from four disciplines and three universities in the UK. 'Think aloud' protocols and interviews were utilised to access both actual and espoused data regarding grading practices. The findings illuminate how lecturers are implicitly combining different epistemologies in negotiating work practices. Whilst recognising that much of what they do is internalised and not explicit, markers appear to gain a sense of security about their marking judgements from various quality assurance documents and procedures; statements of criteria and standards, moderation and cross-marking and interaction with external examiners. Indeed, the research found that an interpretivist viewpoint was only embraced to a certain extent with academics continuing to believe in fixed standards and 'right' marks and vesting expert examiners with privileged knowledge of those standards.

Nevertheless, the study suggests that the act of creating and discussing, for example, assessment criteria provides a clear opportunity for detailed dialogue about standards. Therefore, there appears to be a strong influence of 'techno-rational' artefacts on interpretive practice, as many quality assurance processes for assessment become normative. Indeed there appeared to be an easy interplay for these staff between the tacit, norm-referenced, holistic and local traditions of grading and the explicit, criterion-referenced, analytical and national imperatives of techno-rational quality assurance. These extracts from lecturer interviews reflect this interplay:

...essentially the descriptions which exist in written documents which you've probably seen about what a First Class grade means, what a Second Class grade means and so on, they are rooted in my mind and have become part of my sort of experience really and I feel I can judge, I mean I could sit here and list all the criteria but there's no point in that. I feel I can judge now myself without referring to any kind of written standards but we do operate in accordance with those standards. (T5 History)

...the detail, the fact that they've met the learning outcomes and they've met the learning criteria. You know they have to tick all the boxes in the tickable bit but fundamentally you can see quality in front of your eyes especially when we're in a

visual subject. (T13 Product Design) (Cited in Boyd and Bloxham, 2011, which also describes how the data was collected and analysed).

Perhaps this interplay of approaches in practice provides a clue to a future theoretical basis for academic standards in use and helps identify the focus for a new research agenda. Such an agenda would embrace the idea that assessment judgement informs and is informed by a range of interacting processes:

- involvement in the creation and use of explicit statements of standards (for example, rubrics, assessment criteria, learning outcomes, grade descriptors, professional standards);
- use of learned, internalised, tacit standards;
- engagement with students' work;
- discussion with, and feedback from, internal and external colleagues during second marking, moderation, assessment design;
- use of previous experience including relevant professional and industrial experience.

A recognition of these contributing influences could help identify how quality assurance processes can reflect the reality of assessment in action, making the processes (if not the judgement) explicit, particularly for students. Rather than reject the interpretivist approach, it can provide a framework for considering how consistency can be built into assessment judgements in a way that coincides with knowledge of professional judgement as complex and tacit. The following paragraphs identify some of the component parts of such a framework, which deserve further research.

Systematising processes for learning and safeguarding standards

A new theoretical approach would reject the idea of complete reliability of judgement. Instead it would recognise the power of 'collaborative professional judgement' (Moss and Schutz, 2001, p. 62), exploring how greater confidence can be established in that concept. Indeed, this is not new as many writers are now stressing social constructivist approaches to building shared standards through dialogic communities (Orr, 2010; Crisp, 2008).

What is possibly new is the notion of how this building of standards could be made more systematic and demonstrable. There is significant evidence that lecturers consider dialogue helpful to agreeing standards and Yorke (2008) draws on a range of studies and reports to emphasise the potential of promoting dialogue amongst academics regarding standards. Tang (2008) also identifies discussion about the interpretation of criteria as a way to tackle consistency in summative assessment. However, empirical evidence about the power of these processes in practice is thin and certainly needs further systematic evaluation. The types of process which might be examined more closely are moderation discussion (internal and cross institutional), Assessment Benchmarking Clubs (Woolf & Cooper, 1999), marking bees and the process of generating sets of assessment criteria or other guidance for students (Price *et al.*, 2001), as discussed below. Can evidence be built that these processes, managed appropriately, have greater potential to elucidate and safeguard standards?

Such a theoretical approach would also need to explore the impact of engagement with students' work in building confidence in standards. An interpretivist perspective recognises student work as an important referent in building lecturers' standards' frameworks but is it possible to regularise this norm referencing by greater explicitness and

discussion of those norms; for example by discussion of past papers and exemplars as part of a process of surfacing the approaches to judging work that lecturers actually adopt.

Quality assurance documentation

The role and potential of explicit statements of standards needs to be foregrounded in this new theoretical approach, if for no other reason than the huge investment in these artefacts by both lecturers and students. There is considerable evidence of the failure of such explicit information to make standards transparent but is there a need for a better understanding of the role they play as part of a social constructivist approach to building shared standards. Whilst evidence suggests that academic teachers should admit to students and others that written criteria alone are not able to function as a basis for consistent judgement, they can serve as a guide to staff by acting as a focus for their discussion; a concrete set of statements for them to test their holistic judgement against. To do this, it is necessary to investigate the potential for such statements to more closely reflect actual criteria for judgement. This is important because of the failure of rubrics to be grounded in empirical study of the standards that assessors actually use (Milanovic *et al.*, 1996) or students' actual performance (Greatorex, 2000) which is fundamentally important. As Sadler (1987) said, standards must be grounded in field experience not in theory. Likewise, Norcini and Shea (1997) argued that establishing effective standards needs them to be evaluated for their realism. Woolf (2004) found that criteria were drawn up by individuals with little input from colleagues. It is important to investigate whether it is possible to develop more inductive ways of building explicit statements of standards, for example by developing them from real exemplars of student work.

However, regardless of whether explicit articulation of standards is a futile endeavour, dialogical engagement in using or creating such quality assurance artefacts may have an important role to play. This accords with the views of O'Donovan *et al.* (2004), Rust *et al.* (2005) and Javitz (2009, p. 603) who stress the need for 'meaningful understanding and application of assessment criteria within an interpretive community or community of practice'. The emphasis is on participatory processes that enable the co-construction and application of meaningful criteria (Shay, 2008). Overall, research needs to consider how the academy can capitalise on the processes of determining explicit standards in order to meet the aspiration to build greater consensus in the tacit standards of academics.

New academic staff

Warren Piper (1994) discussed how learning standards and gaining agreement on them takes place informally through the experience of sharing the judgement of actual student work. More recently, Javitz's (2009) work on how academics learn to assess reinforces this idea that relevant knowledge can only be learnt participatively, through the 'interpretive support' of the relevant academic community. A new theoretical approach to standards in use would need to investigate how the 'community' can most effectively support the learning of standards amongst new academic teachers. The combined influences of public knowledge and practical wisdom should be central to this investigation (Boyd & Bloxham, 2011).

Staff awareness and assessment literacy

Academics' relaxed view of standards in use may partially result from the idea that they do not have conscious knowledge of all that they know in relation to academic standards and

grading judgements (Orr, 2010; Broad, 2003). Indeed Broad claims that lecturers do not have satisfactory access to their own standards by merely reflecting on them. In his view, this access is only gained through discussion of actual student work. An academic community that is more knowledgeable about assessment and professional judgement might be in a better position to defend their practices to students who are likely to become ever more concerned about fair assessment as their investment in higher education rises.

Thus, a further field of investigation is the development of 'assessment literacy'; a term used in school-based education (Stiggins, 1991; Popham, 2009) to refer to an understanding of the proper design and use of reliable and valid assessment. In relation to assessment standards in higher education, assessment literate teachers would be sufficiently experienced, alert to the vagaries of professional judgement and conscious of developments in good assessment practice. It would explore how assessors might be made more conscious of their tacit knowledge and aware of how that might influence and be influenced. Moss and Schutz (2001) argued that examiners should be helped to see the potential one-sided nature of judgement or, at least, to understand their own perspectives better. Orr (2008, p. 141) suggested that academics might be encouraged to reflect on and evaluate their use of criterion and norm-referencing in their grading so they develop more understanding of 'intuitive practices and unwritten rules'. Orr (2008) also recommended that someone might observe moderation meetings and feed back to staff: 'In this way, lecturers can explore the implicit and explicit rules and practices relating to the allocation and spread of marks from year to year' (Orr, 2008, p. 141).

Trust in standards by students and other stakeholders

An important aspect of any changes in assessment is to maintain and improve the trust in standards held by students and other stakeholders such as employers and professional, statutory and regulatory bodies and this would form a final, crucial component of a research agenda. As Carless (2009, p.78) argued, where trust is replaced by high levels of accountability and surveillance through quality mechanisms, systems are inefficient and staff are distracted from their 'core roles'. None the less, it is important to demonstrate that alternative approaches justify a trust in the consistency and reliability of standards both within and external to the academy.

Perhaps most importantly, investigation is required into how students' knowledge of and trust in academic standards can be established and maintained. Studies have been conducted to better understand how students can be helped to grasp tacit standards as they work through their studies (Rust *et al.*, 2003). However, this is a separate point. A new approach would need to explore what should be communicated to students to build their confidence in fairness and consistency, particularly given their lack of satisfaction with assessment. Key elements might involve explaining:

- that explicit standards are only guidelines, not measurement scales. They are a genuine effort to express standards but they cannot reflect the full complexity of professional judgement;
- the nature of professional judgement (holistic, intuitive, tacit);
- the training/ dialogue that takes place to build consensus internally and externally and to develop the assessment literacy in academic staff;
- the range of checks and balances put in place to provide for consistency and fairness.

Yorke (2011) stressed the importance of tackling the issue of trust. As he pointed out, once a 'judgement' rather than a 'measurement' approach to assessment is accepted, the issue of trust needs to be faced openly. Can students' trust be built if there is greater honesty about professional judgement? Some writers (for example, Carless, 2009) have advocated greater transparency in this matter coupled with involving students in assessment as an important skill for lifelong learning and employability (practising professional judgement itself). However, it has yet to be explored whether this is achievable in the current climate of accountability. Building students' and others' assessment literacy would seem an important pre-requisite for creating this trust as are demonstrable systems for quality assurance, albeit based on social constructivist rather than techno-rational principles.

Conclusion

The academy needs a theory of standards that can be explained to students, novice academics, parents, politicians and those in higher education quality assurance. Above all, it needs to have credence within the academy and reflect or improve actual practice in a pragmatic way. This article has attempted to elucidate the failure of current theory to underpin standards in use. It posits a future research agenda that draws on the interplay between existing approaches to develop a more coherent model. A theory is required that bridges the limitations of explicit standards and the invisibility and variability of tacit standards to clearly demonstrate realistic and robust ways to achieve more effective security and fairness of standards for students and the academy in general. A framework is offered here, which focuses on systematising a range of processes for learning about and safeguarding standards particularly for new staff, reviewing the role and potential of documented standards, building staff awareness and assessment literacy and establishing trust in standards by students and other stakeholders. This framework is purposively at an early stage of debate and warrants close scrutiny by others working in the field of assessment, standards and quality assurance. However, it does correspond with and build on the work of existing researchers in the field such as O'Donovan *et al.* (2008) and Carless (2009) in recognising the importance of new approaches to building consistency and trust through 'community' approaches and assessment dialogues.

References

- Alderman, G., 2009, 'Defining and measuring academic standards: a British perspective', *Higher Education Management and Policy*, 21(3), pp. 11–22.
- Ashworth, M., Bloxham, S. & Pearce, L., 2010, 'Examining the tension between academic standards and inclusion for disabled students: the impact on marking of individual academics' frameworks for assessment', *Studies in Higher Education*, 35(2), pp. 209–23.
- Bloxham, S. & Boyd, P., 2011, 'Accountability in grading student work: securing academic standards in a twenty-first century quality assurance context', *British Educational Research Journal*, available on line at <http://dx.doi.org/10.1080/01411926.2011.569007>, accessed 5 July 2012.

- Bloxham, S., Boyd, P. & Orr, S., 2011, 'Mark my words: the role of assessment criteria in UK higher education grading practices', *Studies in Higher Education*, 36(6), pp. 655–70.
- Boyd, P. & Bloxham, S., 2011, 'Grading student work: the workplace learning of university lecturers in their role as assessors of student coursework', paper presented at the Society for Research in Higher Education Conference, 7–9 December, Newport, Wales.
- Brennan, J., 1996, 'Introduction: the standards debate' in Brennan, J. (Ed.), *Changing Conceptions of Academic Standards* (London, Open University, Quality Support Centre) pp. 9–26.
- Broad, B., 2000, 'Pulling your hair out: crises of standardization in communal writing', *Research in the Teaching of English*, 35, pp. 213–60.
- Broad, B., 2003, *What We Really Value: Beyond rubrics in teaching and assessing writing* (Logan, Utah, Utah State University Press)
- Bruner, J.S., 1996, *The Culture of Education* (Cambridge, MA, Harvard University Press).
- Carless, D., 2009, 'Trust, distrust and their impact on assessment reform', *Assessment and Evaluation in Higher Education*, 34(1), pp. 78–89.
- Coates, H., 2010, 'Defining and monitoring academic standards in Australian higher education', *Higher Education Management and Policy*, 22(1), pp. 41–58.
- Crisp, V., 2008, 'Exploring the nature of examiner thinking during the process of examination marking', *Cambridge Journal of Education*, 38(2), pp. 247–64.
- Delandshere, G., 2001, 'Implicit theories, unexamined assumptions and the *status quo* of educational assessment', *Assessment in Education*, 8(2), pp. 113–33.
- Delaney, C. M., 2005, 'The evaluation of university students' written work', paper presented to the *British Educational Research Association Annual Conference*, 14–17 September 2005, Glamorgan, Wales.
- den Outer, B., & Handley, K., 2010. 'Standards, situation and self-criticality: exploring situational analysis, grounded theory and the postmodern turn, to enhance reflexive practice in higher education research', paper presented at the *Higher Education Close Up 5 Conference*, Lancaster, UK, 20–22 July 2010.
- Dobson, S., 2008a, 'Applying a validity argument model to three examples of the viva', *Nordisk Pedagogik*, 28, pp. 332–44.
- Dobson, S., 2008b, 'Theorising the viva—a qualitative approach', *Assessment and Evaluation in Higher Education*, 33(3), pp. 277–88.
- Elander, J. & Hardman, D., 2002, 'An application of judgement analysis to examination marking in psychology', *British Journal of Psychology*, 93, pp. 303–28.
- Gaunt, D., 1999, 'The practitioner as external examiner', *Quality in Higher Education*, 5(1), pp. 81–90.

- Grainger, P., Purnell, K. & Zipf, R., 2008, 'Judging quality through substantive conversations between markers', *Assessment and Evaluation in Higher Education*, 33(2), pp. 133–42.
- Greatorex, J., 2000, 'Is the glass half full or half empty? What examiners really think of candidates' achievement', paper presented at the *British Educational Research Association Conference*, 7–10 September 2000, Cardiff.
- Hartley, J., Trueman, M., Betts, L. & Brodie, L., 2006, 'What price presentation? The effects of typographic variables on essay grades', *Assessment and Evaluation in Higher Education*, 31(5), pp. 523–34.
- Harvey, L., 2002, 'Evaluation for what?', *Teaching in Higher Education*, 7(3), pp. 246–63.
- Hawe, E., 2002, 'Assessment in a pre-service teacher education programme: the rhetoric and the practice of standards-based assessment', *Asia Pacific Journal of Teacher Education*, 30, pp. 93–106.
- Hunter, K. & Docherty, P., 2011, 'Reducing variation in the assessment of student writing', *Assessment and Evaluation in Higher Education*, 36(1), pp. 109–24.
- Jackson, N., 1999, 'Foreward' in Smith, H., Armstrong M & Brown, S. (Eds.) *Benchmarking and Threshold Standards in Higher Education* (London, Kogan Page).
- Jawitz, J., 2009, 'Learning in the academic workplace: the harmonization of the collective and individual habitus', *Studies in Higher Education*, 34(6), pp. 601–14.
- Kangis, P., 2001, 'Presentational dimensions and marks awarded to assignments', *Quality in Higher Education*, 7(3), pp. 199–206.
- Kappler, K., 2004. 'NCATE: Wolf in shepherd's clothes' in Poetter, T., Goodney, T. & Bird, J. (Eds.) *Critical Perspectives on The Curriculum of Teacher Education* (Lanham, MD, University Press of America).
- Knight, P., 2006, 'The local practices of assessment', *Assessment and Evaluation in Higher Education*, 31(4), pp. 435–52.
- Leach, L., Neutze, G. & Zepke, N., 2001, 'Assessment and empowerment: some critical questions', *Assessment and Evaluation in Higher Education*, 26(4), pp. 293–305.
- Martin, S. & Cloke, C., 2000, 'Standards for the award of qualified teacher status: reflections on assessment implications', *Assessment and Evaluation in Higher Education*, 25(2), pp. 183–90.
- Middlehurst, R., 1996, 'Degree standards and quality assurance: a discussion' in Brennan, J. (Ed.) *Changing Conceptions of Academic Standards* (London, Open University, Quality Support Centre).
- Milanovic, M., Saville, N. & Shuhong, S., 1996, 'A study of the decision-making behaviour of composition markers' in Milanovic, M. & Saville, N (Eds.) *Performance Testing, Cognition and assessment* (Cambridge, Cambridge University Press).

- Morgan, W. & Wyatt-Smith, C. M., 2000, 'Im/proper accountability: towards a theory of critical literacy and assessment', *Assessment in Education*, 7(1), pp. 123–42.
- Moss, P. A. & Schutz, A., 2001, 'Educational standards, assessment and the search for consensus', *American Educational Research Journal*, 38(1), pp. 37–70.
- Mullins, G. & Kiley, M., 2002, 'It's a Ph. D, not a nobel prize'; how experienced examiners assess research theses', *Studies in Higher Education*, 27(4), pp. 369–86.
- Neil, D. T. & Wadley, D. A., 1999, 'A generic framework for criterion-referenced assessment of undergraduate essays', *Journal of Geography in Higher Education*, 23(3), pp. 303–19.
- Norcini, J. J. & Shea, J. A., 1997, 'The credibility and comparability of standards', *Applied Measurement in Education*, 10(1), pp. 39–59.
- O'Donovan, B., Price, M. & Rust, C., 2004, 'Know what I mean? Enhancing student understanding of assessment standards and criteria', *Teaching in Higher Education*, 9(3), pp. 325–35.
- O'Donovan, B., Price, M. & Rust, C., 2008, 'Developing student understanding of assessment standards: a nested hierarchy of approaches', *Teaching in Higher Education*, 13(2), pp. 205–17.
- Orr, S., 2008, 'Real or imagined? The shift from norm referencing to criterion referencing', in Havnes, A. & McDowell, L. (Eds.) *Balancing Dilemmas in Assessment and Learning in Contemporary Education* (London, Routledge).
- Orr, S., 2010, 'We kind of try to merge our own experience with the objectivity of the criteria': The role of connoisseurship and tacit practice in undergraduate fine art assessment', *Art, Design and Communication in Higher Education*, 9(1), pp. 5–19.
- Orrell, J., 2008, 'Assessment beyond belief: the cognitive process of grading', in Havnes, A. & McDowell, L. (Eds.) *Balancing Dilemmas in Assessment and Learning in Contemporary Education* (London, Routledge).
- Orrell, J., 2003, 'Congruence and disjunctions between academics' thinking when assessing and their beliefs about assessment practice', in Rust, C. (Ed.) *Improving Student Learning: Theory, research and scholarship* (Oxford, OCSLD).
- Popham, W. J. 2009 *Unlearned Lessons: Six stumbling blocks to our schools' success* (Cambridge, MA, Harvard Educational Press).
- Price, M., O'Donovan, B. & Rust, C. 2001, 'The student experience of criterion-referenced assessment (through the introduction of a common criteria assessment grid)', *Innovations in Education and Teaching International*, 38(1), pp. 74–85.
- Quality Assurance Agency for Higher Education (QAA), 2010, *Evaluation of the Academic Infrastructure: Final report* (Gloucester, QAA)
- Ross, V., 2009, 'External music examiners: micro-macro tasks in quality assurance practices', *Music Education Research*, 11(4), pp. 473–84.

- Rust, C., O'Donovan, B. & Price, M., 2003, 'Improving students learning by developing their understanding of assessment criteria and processes', *Assessment and Evaluation in Higher Education*, 28(2), pp. 147–64.
- Rust, C., O'Donovan, B. & Price, M., 2005, 'A social constructivist assessment process model: how the research literature shows us this could be best practice', *Assessment and Evaluation in Higher Education*, 30(3), pp. 231–40.
- Sadler, D. R., 1987, 'Specifying and promulgating achievement standards', *Oxford Review of Education*, 13(2), pp. 191–209.
- Sadler, D. R., 1989, 'Formative assessment and the design of instructional systems', *Instructional Science*, 18(2), pp. 119–44.
- Sadler, D. R., 2005, 'Interpretations of criteria-based assessment and grading in higher education', *Assessment and Evaluation in Higher Education*, 30(2), pp. 175–94.
- Sadler, D. R., 2009a, 'Indeterminacy in the use of preset criteria for assessment and grading', *Assessment and Evaluation in Higher Education*, 34(2), pp. 159–79.
- Sadler, D. R., 2009b, 'Transforming holistic assessment and grading into a vehicle for complex learning' in Joughin, G (Ed.) *Assessment, Learning and Judgement in Higher Education* (New York, Springer).
- Sadler, D. R., 2011, 'Academic freedom, achievement standards and professional identity', *Quality in Higher Education*, 17(1), pp. 85–100.
- Salmi, J., 2009, *The Growing Accountability Agenda in Tertiary Education: Progress or missed blessing?* (Washington, DC, The World Bank).
- Shay, S., 2004, 'The assessment of complex performance: a socially situated interpretive act', *Harvard Educational Review*, 74(3), pp. 307–29; pp. 23.
- Shay, S., 2005, 'The assessment of complex tasks: a double reading', *Studies in Higher Education*, 30(6), pp. 663–79.
- Shay, S., 2008, 'Beyond social constructivist perspectives on assessment: the centring of knowledge', *Teaching in Higher Education*, 13(5), pp. 595–605.
- Silver, H. & Williams, S., 1996, 'Academic standards and the external examiner' in Brennan, J. (Ed.) *Changing Conceptions of Academic Standards* (London, Open University, Quality Support Centre,).
- Stiggins, R. J., 1991, 'Assessment literacy' *Phi Delta Kappan*, 72(7), pp. 534–39.
- Tang, S. Y. F., 2008, 'Issues in field experience assessment in teacher education in a standards based context', *Journal of Education for Teaching*, 34(1), pp. 17–32.
- Vaughan, C., 1991, 'Holistic assessment: what goes on in the rater's mind?', in Hamp-Lyons, L (Ed.) *Assessing Second Language Writing in Academic Contexts* (Norwood, NJ, Ablex Publishing Corp).

- Warren Piper, D., 1994, *Are Professors Professional* (London, Jessica Kingsley).
- Wolf, A., 1995, *Competence-based Assessment*, (Buckingham, Open University Press).
- Woolf, H., 2004, 'Assessment criteria: reflections on current practices', *Assessment and Evaluation in Higher Education*, 29(4), pp. 479–93.
- Woolf, H. & Cooper, A., 1999, 'Benchmarking academic standards in history: an empirical exercise', *Quality in Higher Education*, 5(2), pp. 145–54.
- Yorke, M., 2008, *Grading Student Achievement in Higher Education* (London, Routledge).
- Yorke, M., 2009, 'Faulty signals? Inadequacies of grading systems and a possible response', in Joughin, G (Ed.) *Assessment, Learning and Judgement in Higher Education* (New York, Springer).
- Yorke, M., 2011, 'Summative assessment: dealing with the 'measurement fallacy'', *Studies in Higher Education*, 36(3), pp. 251–73.

