

Bloxham, Susan and Boyd, Pete (2012) Accountability in grading student work: securing academic standards in a twenty-first century quality assurance context. *British Educational Research Journal*, 38 (4). pp. 615-634.

Downloaded from: <http://insight.cumbria.ac.uk/id/eprint/1180/>

Usage of any items from the University of Cumbria's institutional repository 'Insight' must conform to the following fair usage guidelines.

Any item and its associated metadata held in the University of Cumbria's institutional repository Insight (unless stated otherwise on the metadata record) may be copied, displayed or performed, and stored in line with the JISC fair dealing guidelines (available [here](#)) for educational and not-for-profit activities

provided that

- the authors, title and full bibliographic details of the item are cited clearly when any part of the work is referred to verbally or in the written form
 - a hyperlink/URL to the original Insight record of that item is included in any citations of the work
- the content is not changed in any way
- all files required for usage of the item are kept together with the main item file.

You may not

- sell any part of an item
- refer to any part of an item without citation
- amend any item or contextualise it in a way that will impugn the creator's reputation
- remove or alter the copyright statement on an item.

The full policy can be found [here](#).

Alternatively contact the University of Cumbria Repository Editor by emailing insight@cumbria.ac.uk.

Accountability in grading student work: Securing academic standards in a 21st century quality assurance context.

Sue Bloxham, Pete Boyd

University of Cumbria, UK

Abstract

This article, using a student outcomes definition of academic standards, reports on academics' sense of standards as enacted through marking practices. Twelve lecturers from two UK universities were asked to 'think aloud' as they graded written assignments followed by a semi-structured interview. The interview data was used to investigate the source of tutors' standards, their sense of accountability for their grading judgements, their use of artefacts and their attitude to internal and external moderation. The findings suggest that tutors believe there are established and shared academic standards in existence for their discipline and they endeavour to maintain them. There was no evidence of significant pressure or practice related to lowering of standards although differences in tutors' tacit 'standards' frameworks' have the potential for bias. Whilst moderation has some power to secure standards within teams, the article discusses the implications of the research for assuring standards across universities and disciplines.

Introduction

Higher education has witnessed several decades of increasing regulation and accountability regarding academic standards. In relation to assessment, concerns about standards in the UK have expressed themselves in debates about grade inflation, lack of parity of standards across Universities and weaknesses in external examining (Brown, 2010b). Largely absent in these debates is consideration of grading practices; the professional judgement which embodies our sense of academic standards. Perhaps the high stakes nature of marking combined with its low status as an academic activity have combined to inhibit significant debate on the subject. This paper reports on a study of academics' sense of standards as enacted through marking practices. It investigates the source of tutors' standards, their sense of accountability for their

grading judgements, their use of artefacts and their attitude to internal and external moderation of their grading.

Whilst research on marking in higher education is relatively limited, there has been no shortage of debate about academic standards both in the UK and elsewhere (Brown, 2010b; Vanderslice, 2004; Coates, 2010). It is a lively debate but not a new one (Brennan *et al.*, 1996) and has been prompted by the growing diversity and complexity in higher education (Woolf & Cooper, 1999). Such has been the high profile of this debate that the topic has warranted official examination (HEFCE, 2009; QAA, 2009) including a parliamentary select committee report (Innovation, Universities, Science and Skills Committee, 2009). Brown (2010b) has identified five areas of concern emerging from this debate: information, comparability, external examining, the protection of academic judgements and the role of the UK Quality Assurance Agency (QAA).

Defining academic standards

Academic standards are conceptually complex (Coates, 2010; Woolf & Cooper, 1999) and difficult to define (Coates, 2010). Middlehurst (1996) has described the term as composite including input, process and output elements. Harvey (2002) distinguishes *academic* standards from standards of competence, service standards and organisational standards (all of which are the focus of quality assurance) by centering his definition on academic attainment usually related to 'course aims and objectives, operationalised via performance on assessed pieces of work'. (p253). This definition is similar to that of Alderman (2009) who refers to academic standards as:

'discrete levels of intellectual performance, the attainment of which results in the award of academic credit' (p12)

In this vein, in the UK, The Higher Education Quality Council (HEQC) in 1995, coined the term 'graduate standards' to distinguish these output-focused academic standards from those more associated with quality management and standards of provision:

'explicit levels of academic attainment that are used to describe and measure academic requirements and achievements of individual students and groups of students' (HEQC 1995 from Brennan, 1996).

This output approach is also evident in the definition used by other higher education quality agencies, for example in Australia (Coates, 2010). It underlies the QAA's definition of academic standards which is 'the threshold level of achievement that a student has to reach to gain an academic award'. The QAA distinguishes these from *academic quality* which is defined as 'how well the learning opportunities made available to students are managed in order to help them achieve their award' (QAA, 2010, : 2). Therefore, following this consensus, for the purposes of this paper, we are using the term 'academic standards' as related to standards of student attainment.

Gold standard or social construction?

Whilst many commentators and organisations concur on an output-focused definition of academic standards, there follows less agreement about the provenance or epistemology of standards. It could be argued that there are two paradigms; the first being one in which standards are conceptualised as benchmarks (akin to a gold standard) against which student work is assessed and which are agreed across groups and over time. For example, the QAA consider that such *academic standards* are made explicit through various elements of the academic infrastructure (QAA, 2006). Such a perspective assumes a criterion referenced approach to assessment whereby students are measured against an explicit standard or a set of criteria instead of being measured in relation to each other's performance. As we shall see below, this epistemological position is central to many of the criticisms of existing quality assurance in 'safeguarding' standards. Such critiques rest at the level of our failure to set or test comparability of standards but don't challenge the notion that academic standards can be set or monitored in an objective way (Brown, 2010a; Gibbs, 2010). This is a techno-rationalist perspective with a positivist view of standards (Orr, 2007; Delandshere, 2001); an approach in which standards are viewed as untainted by values, culture or power.

An alternative, socio-cultural, paradigm recognises assessment as a context-dependent, socially situated, interpretive activity (Shay, 2004) with historical analysis illustrating its 'key role in cultural reproduction and social stratification' (Gipps, 1999). Thus, the 'gold standard' approach is critiqued for failing to problematise academic standards in terms of power and values: 'Values, and therefore politics of preference, are inevitably at play in any acts of e/valuation' (Morgan & Wyatt-Smith, 2000: p139). In addition, Delandshere (2001) critiques assessment

practice which is based on assumptions that 'knowledge is monolithic, static and universal' (p127), a view echoed by Shay (2004). Such a fixed view also assumes that staff are acculturated into 'guild knowledge' (Sadler, 1989) regarding assessment rather than active agents in co-creating standards in their individual and local marking practices (Knight 2006). The socio-cultural approach would recognise 'guild knowledge' as 'the ability to make sound qualitative judgements' (Sadler, 1989, : p126) gained from collaboration with other assessors, but it would recognise it as a shifting terrain rather than merely 'insider knowledge' of predetermined standards. Thus, from this perspective, the simplistic and fixed notions of standards as portrayed in public debate deny the necessarily elusive and dynamic nature of academic standards which are continuously co-constructed by academic communities and ferociously difficult to explain to a lay audience.

Securing academic standards

The official discourse of quality assurance (QA) regimes in many higher education systems rests largely on the techno-rationalist perspective and sits in clear contrast to the socio-cultural paradigm. Indeed, the development of QA in relation to academic standards has been focused on procedures for making them more explicit. Alderman (2009) argues that historically academic standards were sustained over time by an oral tradition through contact between Universities and subject communities, what Brennan (1996) refers to as a private and implicit approach to communicating standards. The oral tradition was clearly evidenced in the lack of any explicit criteria for assessment or documentation of other aspects of academic standards and Brennan argued in 1996 that such 'shared understandings of an academic elite are insufficient as a basis for standards in a mass system of higher education' (p16), a point reinforced more recently by Salmi (2009). Importantly, whilst both Alderman (2009) and Bellingham (2008) argue that this oral tradition existed at the level of individuals and subjects, Bellingham also emphasises the reliance on institutional responsibility for standards prior to the move toward greater external accountability.

The introduction of more explicit efforts to record and monitor standards began in the UK with the establishment of the Council for National Academic Awards (CNAA) in 1964. In particular, the last twenty years have witnessed considerable work attempting to make academic standards more explicit and consistent. In the UK this has included development of an

'Academic Infrastructure', a code of practice for assessment, a drive towards transparency in assessment including specifying learning outcomes and assessment criteria. These developments have had a focus on both the scope and level of HE programmes. They have brought into the public domain greater information about the stated course content, aims, assessment methods and criteria of programmes of study.

These efforts to make academic standards more explicit have met with two critiques. Firstly, it is recognised that what may have been referred to as oral, private and implicit knowledge may not just reflect the nature of academic structures in the past but the problematic nature of communicating standards. Achieving construct validity in assessing complex knowledge creates difficulty in establishing comparability across institutions and over time because the nature of what is being assessed makes objective, explicit judgement difficult. Judgement draws on 'tacit' knowledge, difficult to express, in order to make assessment decisions (O'Donovan *et al.*, 2008). Price and Rust (1999) and Brown (2010a) reinforce this in pointing to the failure of documentary statements of standards to guide marking in a reliable way. For this reason, Price (2005) points out that the notion of agreed standards maintained by a small and close knit community of scholars, as described by Alderman, is probably a myth.

On the basis of this critique, Brennan's prediction that 'private' would cede to more collective authority for standards is questionable. Whilst the stated curriculum and learning outcomes for programmes are now in the public domain, one could argue that the judgement of student achievement remains largely unchanged; 'the private preserve of teaching staff' (Coates, 2010, : 10). Indeed, Middlehurst (1996) found that 'standards referents' aren't made explicit by academics except through marks. Whilst quality assurance frameworks emerging from concern about the massification of higher education (Coates, 2010) have led to external checks, these checks do not concern themselves with the level of performance and learning demonstrated, with the exception of the external examiner system.

The second, and in some ways more profound, critique is embedded in the socio-cultural and interpretive analysis of assessment standards as discussed above. The notion that we can establish standards independently from the individuals who are using them is challenged:

‘We are social beings who construe the world according to our values and perceptions; thus, our biographies are central to what we see and how we interpret it. Similarly in assessment, performance is not “objective”; rather, it is construed according to the perspectives and values of the assessor’. (Gipps, 1999: p370)

These critiques of efforts to make academic standards more explicit may explain why quality assurance regimes have retained institutional autonomy in relation to standards (QAA, 2010) and, instead, largely focus on the quality of ‘input’ and ‘process’ at the expense of ‘output’; the latter proving too difficult in practice. For example, the Academic Infrastructure established by the QAA was designed to help higher education providers in ‘setting and maintaining academic standards and the quality of provision’ (QAA, 2006). In its code of practice relating to assessment, it gives primacy to accurate and accessible assessment information regarding how students will be judged and fair mechanisms for marking and moderation (QAA, 2006). These ‘represent, but do not describe, expected and achieved levels of performance’ (p80).

This failure of quality assurance to focus on academic standards is a criticism posed by a number of writers (Harvey, 2002; Alderman, 2009; Middlehurst, 1996). Alderman (2009) argues that such an approach is underpinned by the view that high standards will be maintained by standardised procedures, a view which he considers ‘false and dangerous’ (p13) and which is supported by other research, for example on the extent to which double marking achieves consistency (Yorke, 2008). Indeed Harvey (2002) argues that, despite the fact that a key purpose of external evaluation has been to check academic standards, the external evaluation of Universities has served to legitimise the status quo, focusing more on the process of evaluation rather than the substance of what is being evaluated. This critique has led some to advocate the use of ‘validated assessment tasks’ for use across different universities and the use of ‘objective’ tests of student achievement such as the Graduate Skills Assessment Test (Coates, 2010).

Academic standards in practice

In practice, the chief processes for monitoring academic standards in the UK are local moderation and external examining. Indeed Silver & Williams (1996) refer to external

examining as 'a symbol of accepted academic standards across higher education' (p29). However, there has been considerable recent critique of the potential of external examiners to secure consistent academic standards across the sector (QAA, 2007) and Brown (2010b) points to the long term problems associated with the external examining system which has been the subject of successive critical reports. Nevertheless external examiners continue to be vested with a key responsibility for securing academic standards across the sector.

In a discussion of the operationalisation of academic standards in the discipline of history, Woolf and Cooper (1999) argue that academic standards are not fixed across institutions, subjects or time. They found that staff felt colleagues could be 'trusted to know how to assess and that it was this trust which provided the ultimate reliability of the departments' assessment processes' (p150) rather than formal monitoring mechanisms. However, they found staff bemoaning the reduction in the informal opportunities that are available to set common standards and acculturate new staff such that they can be trusted. However, the opportunity to be an external examiner was considered to contribute to learning about and disseminating standards. They conclude that members of their study group felt strongly that 'the whole process [of establishing standards] is subtle, organic and opaque' (p151 brackets from the original).

It is argued that this opaque and subtle collegial process allows a level of flexibility which might put standards at risk (Coates, 2010) and certainly a number of submissions to the recent Select Committee (Innovation, Universities, Science and Skills Committee, 2009) did suggest that there might be a failure to protect academic judgement experienced as pressure on staff to lower academic standards (paras 93 - 95). Undoubtedly, there is evidence of differences in standards across markers (Bloxham, 2009) including differences based on spurious factors such as the font size students use (Kangis, 2001; Hartley *et al.*, 2006). The local and tacit nature of academic standards as outlined above would, of necessity, indicate the importance of the acculturation process for staff. Currently, the research on how staff specifically acquire a sense of academic standards is fairly limited (Reimann *et al.*, 2010) although it has been theorised (Wolf, 1995; Shay, 2005; Jawitz, 2009). Nevertheless Reimann *et al.* (2010) identified the importance of informal learning about standards from others through moderation and debate with colleagues. They also focused on the importance of power relationships in moderation decisions and the need for staff to develop confidence in their assessment knowledge if they are to challenge

more senior or experienced staff. Staff learn marking standards through marking student work and through the response of colleagues; second marking and moderation is particularly important in this.

Summary

This brief review of the meaning and operation of academic standards highlights the separation that exists between the explicit face of quality standards as inputs and processes maintained through institutional procedures and external review, and the private and implicit notion of academic standards as reflected in individual and group judgements of student attainment. Whilst we have increasing documentation and evidence to use in examining the former, our knowledge of the latter remains slim. Standards of student attainment necessarily put a focus on staff judgements regarding achievement. In this context, Bellingham (2008) makes the useful point that reputation in both local and global markets is vital to Universities and this makes them very risk averse. This analysis may well be a reason why open debate about academic standards and marking is very difficult to have and why vice chancellors were considered defensive by the Select Committee (Hinsliff, 2009) from (Satchell & Pratt, 2010). As Price (2005, : 216) argues, 'To question the existence of common assessment standards is very uncomfortable for the HE community because these standards lie at the foundation of award giving powers'.

Consequently, there is a tension for tutors. UK Higher education operates within a QA framework which posits academic standards as something fixed, capable of explicit expression and of consistence in use; something which needs to be comparable across programmes, disciplines, institutions and time (Brown, 2010a); something which is essential to our picture of the UK as a world ranking university system; something which staff need to grasp in order to carry out their assessment work effectively; something which institutions must respect in spite of drives for widening participation, retention and achieving good rankings in league tables during a period of financial pressure. Yet, in the context of grading work, tutors face a much more intangible picture of standards in use. The concrete words on the page of learning outcomes, assessment criteria or grade descriptors become slippery and opaque concepts once they have to be interpreted in relation to complex and diverse student work and reduced to a single mark or grade.

It could be argued that marking student work is at the interstices between this institutional quality assurance environment and the notion of standards as co-constructed in the everyday activity of assessment practice. Academic staff have to negotiate the tension between a technological view of standards as portrayed through QA bodies and institutional quality procedures and the private and tacit experience of embedding standards in their marking. This paper reports on research which is designed to explore how this tension is experienced by staff, if indeed, they experience it as a tension at all. It will focus specifically on the social construction of academic standards as enacted through marking practices. How do academic standards play out in the private process of marking? How do tutors experience academic standards; where and how does the 'social' experience of shared marking and moderation play into tutors' views of academic standards? Are there pressures on staff to lower standards?

Methodology and methods

The investigation uses a socio-cultural theoretical framework, recognising marking as a socially situated activity (Delandshere, 2001). The results set out here were part of a larger mixed method study which combined observation of staff marking practices through the use of think aloud and researcher's field notes with follow up interviews. The sample included twelve academic tutors from two post 1992 universities in England although only 11 of these tutors were able to participate in the interview. The two universities have considerable numbers of students on professional programmes. Six of the tutors were in the professional field of teacher education and the other six came from a range of other subject disciplines: history (2), English literature (2), business studies and performing arts. The tutors were volunteers, recruited through open advertisement in both Universities. This article draws largely on the interview material as this focuses on staff views about standards and their espoused behavior in that regard. However, the 'think aloud' results and analysis, which have been reported elsewhere (Bloxham *et al*, 2011), provide an important backdrop for this aspect of the research. Whilst the interviews explored espoused rather than actual practices in grading (Orrell, 2003), it was hoped that as they followed directly after a marking episode, the responses would be more likely to be grounded in actual grading experience rather than a more disembodied reporting of espoused behaviour which may 'reveal assessors' ideals and rationalisations' rather than actual practices (Orrel 2008: 253). Nonetheless, the latter was more likely as the questions moved away from discussing the specific marking activity recorded to more general questions regarding standards.

The findings presented here focus particularly on aspects of the research which relate to the question: What influences staff marking practices and judgements? The interview questions investigated tutors engagement with academic standards as experienced through marking the scripts and other marking and moderation. In the interview, respondents were asked a number of questions on their style, confidence about and approach to marking, how they know the required standards, their experience of moderation, their use of supporting documents and how they decide what feedback to provide.

Both think aloud activity and interview were recorded, transcribed and analysed using a qualitative thematic approach (Ritchie & Lewis, 2003). An initial coding framework drawn from literature and discussion with volunteer 'co-researchers' (other tutors in a research network concerned with assessment practice) was amended during the process of coding the data as new themes emerged. Following this, the 'markers' and co-researchers were offered an opportunity to comment on the analysis. Their feedback was used to refine the analysis and subsequent discussion of the findings.

Findings

Three main themes emerged in the interview material of relevance to the research questions outlined above. These themes are:

- Accountability
- Concepts and texts as artefacts
- Shared standards and moderation

Accountability

The theme of accountability emerged in the analysis of interview data. This included three key elements, who the tutors feel accountable to, how that accountability mediates their grading and feedback activity, and the tension they experience between maintenance of standards and retention of students.

The analysis suggests that tutors feel most accountable to their immediate colleagues and external examiners for their academic standards. This accountability appears to be viewed by some in a competitive or anxious way; concerned that their marking decisions and feedback will reveal them to be out of alignment with their colleagues. The comparison with peers that tutors make is sometimes expressed in terms of how the quality of their feedback compares with others' but also, in some cases, at a more fundamental level of misjudging grading standards. Tutors express the desire to be accurate markers in the sense of awarding grades that are close to their colleagues:

It's blind marked so you suddenly go into a room with the second marker and you sit down and you sit down and say 57 and they say 78 and you say oh my god I've done it completely wrong. (T2)

Tutors especially do not wish to be seen as too generous with their grades:

I think that this business of average marks is very tricky because you see your average marks in comparison to everybody else's. If yours are much higher then you think oh am I a soft touch. (T10)

Tutors are particularly focused on the pass mark but they may experience a tension between the quality of their teaching and the number of students failing:

I'm trying to think what else might appear as a pressure. Er I don't want too many of them to fail. That sounds an awful thing doesn't it but I do say that to myself? It means I haven't taught them very well, in a sense. (T9)

A competing accountability is expressed by the tutors towards their students as a responsibility for them to take grading seriously and to be 'fair' in their grading.

It's something to do with fairness.....to the students. If you're fair to the students then you are providing them with an entitlement. Yeah, something like that. It's something

to do with what education at this level could be and what we would like it all to be which is something...um.....fairness to the process of enquiry for its own sake. (T1)

This fairness is focused primarily towards the students but within the context of the university as an institution:

I would say I was a generous marker but I'm not generous about allowing things [that] are manifestly inadequate to cross the border because I don't think that does the student a favour, and I don't think it's fair to the other students, and I don't think to be honest it's good for a University to do. (T5)

Tutors appear to reify their 'university' as an expression of the constraining culture felt to exist, rather than report anything more direct in relation to these matters:

I think there's an implicit understanding in this University, like most Universities today, that if you can possibly get the students through you do. In other words if you're hovering over is it just a fail, is it just a pass then be charitable and get the student through. I wouldn't say that's a pressure but certainly a consideration. (T11)

In relation to possible fails, tutors appear to be using their discretion in borderline cases to identify attainment wherever possible:

But [it is a pressure] only because obviously we know the students and like them mostly and want to do well by them. Also because we're well aware of you know targets, income, don't want to lose students all that sort of thing so I'm sure they're both there at the back of my mind. That on the whole, if it comes to pass/fail particularly that we would try quite hard to give a pass rather than a fail. It doesn't stop us from giving fails if we have to but it means that you know that where there's an uncertainty we would err on the side of generosity.(T11)

However, their comments retain a sense of fixed minimum standards which they retain in marking decisions. In this sense, the evidence suggests that tutors did not perceive, or engage in, any explicit lowering of standards.

The analysis of tutors responses about accountability reveals some subtle pressures, for example to achieve an overall average comparable with other tutors, not to fail too many students and not to be too harsh or too soft a marker. Indeed some tutors used the word 'subtle' to refer to these constraints.

The tutors also demonstrate a sense of accountability to the University, the external examiner and external quality review bodies and that particularly relates to feedback. This tends to support the points made in the introduction that quality measures focus on processes (for example, the amount and nature of feedback) rather than the actual standards of student attainment. In this way quality assurance bodies and university procedures appear to be putting considerable pressure on some tutors' time and approach to feedback:

..in terms of constraints it's making sure, am I giving the right kind of feedback that if I'm interrogated by HEFCE or a QAA audit or whatever comes round next or you know what's my external examiner looking for and actually what's the student looking for? It's trying to cover all those things. (T4)

Within the limitations of a very small sample, the teacher educators seem most aware of the external examiner and review bodies as authorities to whom they are accountable for both upholding standards and promoting learning whilst tutors in other subject disciplines were more likely to see colleagues as an audience and positioned the external examiner as a peer.

I think the problem with it is that people can mark very much in isolation, away from each other so there's an awful lot of thinking you know, what everybody else does in order to arrive at their decisions. It's no different from teachers in school really and some of this is putting on an act in a sense. Who are you writing for just as much your colleagues? Do you want to ensure it seems just as astute... (T8)

In relation to external examiners, there were frequent references to the need to meet their expectations regarding academic standards:

...it goes to the Board and the external examiner comes to talk to us and that can be really nerve-wracking really I think - have we done it right, have we got it right and I suppose actually initially, through the support of other colleagues who've been here much longer and I think that was very instrumental I think in giving me the confidence at the start. (T12)

Tutors expressed only limited recognition of the potential fallibility of external examiners. The teacher educators in particular generally appear to experience the role and contribution of external examiners as authoritative rather than informative.

Concepts and Texts as Artefacts

It has become common practice for Universities, departments or staff teams to generate statements of criteria or standards for use during assessment. Sadler has distinguished between the concepts of 'standards' and 'criteria' (Sadler, 2005; Sadler, 2009a; Sadler, 2009b) where 'criteria' represent the characteristics of the work by which its quality will be judged (e.g. structure, evidence, presentation) and 'standards' express the level of attainment (e.g. A; B; C; or 2.i; 2.ii). Marking schemes (or rubrics) generally combine criteria and standards in grid form providing statements of the achievement needed for each criterion to achieve different standards. In the activity of grading, the tutors use a range of artefacts; concepts or texts, through which the grading activity is mediated. They referred to all of the following text artefacts: *assessment criteria, success criteria grid, grade descriptors, rubric, teaching standards, module guides, and learning outcomes*. Whilst some of these constitute criteria, other describe standards or a combination of both.

However, the interview data indicated some confusion regarding the difference between the two terms which were used in very loose and overlapping ways. This may be a reflection of the wide range of different artefacts linked to grading judgements. The interview schedule included a question that referred to standards, with a planned clarification prompt that asked how tutors distinguished between grades. In response tutors tended to discuss grade boundaries, particularly whether work should pass and for most, their answers suggest that standards and criteria are almost inextricable:

Interviewer: 'how would you say you know the standards required in your marking?'

I think that's why the criteria – it's a question of matching – fitness of purpose and fitness for matching the criteria to start with (T3)

I'd look at the benchmarking in the published criteria. Um and judge on those. (T11)

Some tutors did not really express their sense of the meaning of the term standards at all through their answer. As a result, in the analysis it was not possible to distinguish between tutors' conceptions and use of 'standards' and 'criteria' and the difference between the two concepts in the academic standards debate warrants further exploration.

Analysis of the think aloud data (Bloxham *et al.*, 2011) suggests that the tutors did not physically refer to text artefacts whilst marking and used them, if at all, at the end of the marking process to check or confirm their holistic judgement. This raises questions as to what they were drawing on to make these judgements and the data suggests that the answer lies in internalised or embedded sets of standards and criteria which they have developed. They used the words or phrases 'internalised'(T1), 'absorbed'(T5), 'instinctively'(T4), 'got a sense of'(T1), 'in my mind'(T5,T11), 'subliminal'(T5), 'rooted in my mind', 'got a mind set'(T9), 'implicit'(T9), 'have things in our heads'(T9), 'feel' (T12), 'familiar' and 'an understanding' (T6).

For some tutors, this internalisation was reported as memorisation of published criteria (T4, T8, T12); they have marked the same module so many times that they know the criteria 'off by heart' (T8). On the other hand many tutors describe their development of a more personalised set of criteria and standards which does not just involve 'memorising' a set of institutional, discipline or module criteria. These personalised criteria appear loosely linked to explicit criteria, learning outcomes, and other statements of standards, but have been blended into a 'personal standards' framework':

I don't have a very rigorous yardstick here for measuring this paragraph does this, this paragraph does that with learning outcomes but I do have a generic sense of what those learning outcomes are aiming for and what, for example this module [title of module] is supposed to be doing and that's in my mind when I'm marking the papers so I'd say it's

a kind of you know almost subliminal level I've absorbed the outcomes and aims and I am using them. (T5)

Five of the eleven tutors appeared to be in this latter group with a looser, more tacit grasp of their standards. For example, one tutor described how they had learned these standards:

I think probably in reality, the published criteria have to be interpreted in terms of past experience because they're so vaguely written. So I suppose it is more a question of what I know in the past, from colleagues and especially the external is regarded as a 40, 50, 60 whatever. (T11)

In general, there was fairly strong support for the perception by seven of the respondents that, even if they didn't refer to explicit criteria or standards as they marked, they have developed a satisfactory knowledge of what the standards or expectations are in their department. There wasn't generally a sense of rejection of explicit assessment information, although one tutor (T1) challenges aspects of their institution's artefacts. There was more a sense of having absorbed the wide range of relevant criteria, grade descriptors, professional standards and so on sufficiently to support grading activity and to feel confident in making judgements.

.....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 andI feel I can judge now myself without referring to any kind of written standards but we do operate in accordance with those standards. (T5)

Nevertheless, as one tutor pointed out, this internalised approach to assessment knowledge does not easily align with institutional policy.

I think that's true for a lot of people and a lot of people wouldn't want to admit it to managers, to Higher Education Managers, that that's actually the way that it works but I think that is the case a lot of the time and it amazes me I think it's probably very, very accurate a lot of the time. (T4)

However, there was some evidence that the standards staff draw on to mark do vary and might influence their grading decisions. Five of the eleven interviewees indicated a perception that judgement was, in some way, affected by individual differences. For example, a number of tutors highlighted what might be described as 'trigger' qualities in student work which helped them identify the appropriate grade or which side of a grade boundary work should be assigned. Other research (Gilovich et al 2002; Crisp 2008) points to the potential bias that can occur from assessors basing judgments on such 'simplifying heuristics' (Gilovich et al, xv). In this study, trigger characteristics encompassed the inclusion of strong introductions and conclusions, particular source material, the evaluative quality of the work, sources, content and expression or the professional standards demonstrated.

Over and above perhaps you know, there's certain things that tick boxes for certain people aren't there? And some people very much a standard of English - now that is important. With the QTS [Qualified Teacher Status] standards that is important (T8).

And whilst some tutors identified limited trigger qualities, others expressed the complexity of the factors they were taking into account: 'It's so multi-factorial you see' (T1); 'I have a set of criteria in my mind which I can list' (T5).

Three respondents indicated that their informal guidance to students reflects these trigger characteristics. It is as though how the tutor has taught and what guidance they have given signals their own perception of standards; the benchmarks against which they judge the work. This, in turn, affects their expectations of the work or their interpretation of the textual artefacts that the students are provided with.

Interviewer: How do you decide what marks to give?

Um there's um, the students have a success criteria grid and so according to that, and what I tell them, you know there are certain things they have to put in so there's certain descriptive information that has to go in (T10).

These responses, both in terms of the internalisation of standards and the notion of individual trigger characteristics tend to reinforce the idea of tutors' 'personal standards' frameworks' as identified in Ashworth *et al.* (2010). Similar notions of this concept have been identified as 'Teachers conceptions of quality' (Sadler 1989: 127), 'assessors' interpretive frameworks' (Shay, 2005: p 665) and 'pre-formed knowledge structures' (Crisp 2008: 250). Such a framework for judgement could be perceived as a personal expression or understanding of broader academic standards in the relevant academic community. It is constructed and reconstructed over time by reference, amongst other things, to specialist knowledge, other tutors' marking, students' work, involvement in moderation and examiners' feedback. As we see below in the discussion of shared standards, this framework is not whimsical (Sadler, 1987) but clearly informed by their university and subject community, appearing to incorporate and be informed by shared textual and conceptual artefacts. Whilst the idea is not explored in this research, it is likely that there is sequence of mutual influence between tutors' personal standards frameworks and the textual artefacts such as learning outcomes and assessment criteria which they are involved in creating and refining. Shay (2005) sees this as an iterative movement between the two elements. For the purposes of this discussion, their personal framework appears to provide the lens through which tutors read and judge student work.

Some tutors were less secure in their own 'standards frameworks' and talked about needing to use written artefacts, particularly for borderline cases:

In terms of criteria beyond that it's occasionally, if erring on the side of one mark, if I'm not sure basically which category – is this a 2:1, is this a 2:2, I think it's that borderline stuff, I'll go back to the generic Faculty assessment criteria that we use which I find quite limiting for most things but it's a good thing just to go back to, to say hang on, is it? Which side of the border is this going to go? (T4)

And there was some recognition that their own standards might change:

Sometimes it goes up and down, as with every point it will go up and down but sometimes it's very difficult and I think sometimes if I looked at it on a different day I might possibly give it a different mark. I don't know. (T10)

Shared Standards and Moderation

Despite the foregoing discussion which focuses on individual frameworks for marking, the strongest finding emerging from the data analysis is a sense that standards are shared. Whilst there is some recognition of difference, there is strong faith in the power of moderation, written criteria and other artefacts to provide adequate consensus. This supports the work of Shay (2005) and the notion of subjectivity and objectivity combining together in marking. This consensus is particularly strong at the level of local teams and for some it also emerges in relation to their disciplinary community and University (T1, T11).

There are things that are kind of implicit and in fact sometimes difficult to articulate but which nonetheless are relatively sound, that are disciplinary. They are just shared by being in the same discipline and provide a framework for marking that might not be available to other people outside that context. (T1)

In particular, the respondents appeared to hold the view that there is a correct mark for each piece, a finding that emerged in Jawitz' (2009) study.

I make the judgement on a piece of work but it's always that niggling doubt. Am I right? And I can look at the criteria and think am I right? (T4)

I suppose with the rigorous second marking procedure and having the external examiner as well who looks at all our work so we have to be getting it right and it is quite a rigorous process really. (T12)

This expression regarding the 'right' mark is possibly the closest the respondents came to making a judgement about the extent to which their academic standards matched the shared and 'correct' (and fixed?) ones held by other colleagues and, particularly, external examiners:

I'm really, really happy I'm in a situation where I can recalibrate myself against my colleagues and I think the moderation process is hugely important for that (T1)

Overall, there is a strong (and intangible) sense of the consistency between their 'personal standards' frameworks' and locally agreed standards (as discussed by Shay (2005)). However, despite this internalised understanding, there is a sense in which markers gain a sense of reassurance in their standards by a range of processes and devices and this is important because they recognise the responsibility that they hold as markers.

It's a real responsibility actually. Well I suppose that's why I'm using these, the criteria and the marking grades and the second marking procedure to ensure that I'm being fair in the marks that I am giving. (T12)

All eleven tutors expressed a strong sense of the value of moderation with positive comments about its contribution to marking processes and accuracy of grades. For five of the tutors, moderation has an important quality assurance function as a safety net or comfort, getting a second opinion, providing a fairer result for students because people have different approaches to marking, with one tutor describing it as 'insurance' against poor markers. Moderation is seen as particularly useful for lengthy assignments and borderline decisions. Three tutors referred to using pre-marking moderation where tutors all marked the same pieces in order to agree standards before commencing marking. Tutors mentioned the usefulness of moderation for thinking about their marking, for reflecting on what they value in student work and identifying if they were out of line with others' marking standards:

it's a very, very useful process because then you find yourself, if you come out, you know, in the rare occasions I discovered myself more generous than a colleague you think oh that's ... you know what does that tell me about what I value? Within a piece of work. (T8)

although one tutor would appreciate a greater opportunity to test tutors' practices in debate with others:

this is why I agreed to take part in this, only I think it's an interesting and valuable exercise because I'm very much aware it's the sort of thing that we don't always discuss in detailI think it is interesting to step back and think why is that 59? Why is that

61? And to have to justify it to some extent to somebody else. I think that's a healthy thing to have to do. (T11)

The generally positive attitude to moderation activities expressed in this study is important and suggests it has an essential role to play as a site of workplace learning and in managing academic standards. The traditional purpose of moderation activities (e.g. agreeing marks between multiple markers) has been challenged as a basis for assuring 'accurate' or best marks (Bloxham, 2009; Satchell & Pratt, 2010), but in common with Reimann et al's (2010) work, these findings strongly support the view that it does have an important role to play in staff learning about standards in the private, autonomous act of marking. Indeed, it could be argued that shared marking and moderation is probably a major element in the continuity of standards (Jawitz, 2007, 2009) because it is the only real opportunity to see the marking of experienced staff and engage in discussion about academic standards, albeit with limitations created by power relationships in moderation discussions (Orr, 2007; Reimann *et al.*, 2010; Jawitz, 2007). Thus, Rust *et al's* (2005) 'social constructivist assessment process model' places moderation as one opportunity for the active engagement and participation by tutors that fosters a shared interpretation of standards. Moderation, whether formal or informal, provides a context to expose the tacit knowledge embodied in our personal 'standards' frameworks' with the potential for this sharing of perspectives to lead to clearer agreement about those standards. For new markers, it offers legitimate peripheral participation (Lave & Wenger, 1999) in the community of assessors; to acquire 'guild knowledge' (Sadler, 1989). In this way, moderation should be perceived as fundamentally a quality enhancement activity and should be used accordingly. Following Satchell and Pratt (Satchell & Pratt, 2010), spurious mathematical manipulation of marks (e.g. averaging two markers grades) should be dispensed with.

Conclusion

This work has explored how staff negotiate the potential tension between the concept of academic standards contained in quality assurance procedures and public debate, and the tacit nature of standards embedded in the private, autonomous act of marking. The introductory discussion identified growing political concerns that academic standards are at risk in an increasingly diversified and massified higher education system. This research does not support

that concern. The evidence here suggests that tutors adhere to the view that there are established and shared academic standards in existence for their discipline and they endeavor to achieve them, applying flexibility and referring back to formal artefacts in borderline situations. There was no evidence of significant pressure or practice related to a lowering of standards and to some extent, at least in the two post 1992 universities concerned, there is some evidence that tutors will tend to err on the low side as markers because they do not wish to be seen as too generous in comparison with their peers or the external examiner. In this sense, the research dismisses the notion of a tension for staff between QA and marking practice. Instead, there is a recognition that whilst much of what they do is internalised and not explicit, they gain a sense of security about their marking judgements from the various QA processes and procedures; statements of criteria and standards, moderation and cross marking, professional standards and the external examiner role.

What is less certain is whether these individuals' personal standards frameworks and local 'guild knowledge' provide any confidence in the consistency and comparability of academic standards more widely. Good moderation activities have the power to support the co-construction of standards but they are likely to be limited to the local context and determining standards across institutions and disciplines require other approaches. External examiners are endowed by tutors with the power to hold and recognise appropriate standards; perhaps an unwarranted authorisation given recent analyses of their role in practice. However, this does vest them with considerable power and therefore there is an obligation to ensure that examiners are sufficiently experienced, alert to the vagaries of professional judgement and conscious of developments in good assessment practice; in other words, that they are *assessment literate*. Biggs (2002) has written lucidly about the power of external examiners to work against quality enhancement if they lack good assessment knowledge. Nevertheless, the external examining system has a significant benefit (Harvey & Williams, 2010, : 92) in that it 'constitutes a conference, in some cases the only conference, in which academics discuss educational issues related to their discipline area'. It is important that teaching teams, with the support of their institution, have an effective balanced relationship with external examiners which positions examiners as useful in both prompting this dialogue and providing an external perspective from a peer rather than as some kind of infallible judge of standards.

A complementary approach described by Woolf and Cooper (1999) is an Assessment Benchmarking Club involving several university history departments. They argue that it proved a powerful tool for developing agreement and checking on assessment principles and strategies. Similarly, a Subject Centre of the Higher Education Academy (HSLT, 2009) set up a 'college of peers' where colleagues across several universities reviewed exemplars of student work graded at the borderlines of different grades in an effort to help the subject community establish and maintain standards. These shared activities, welcomed by participants, recognise the tacit nature of much assessment knowledge but might be more effective in building consistent 'standards frameworks' across institutions than the current reliance on either external examiners or attempts to codify academic standards in written guidance. Coates (2010) also argues for more systematised and cross institutional moderation as a way to improve the generalisability of assessment standards.

Such an approach may go some way to tackling the evidence in this study that tutors do differ in the criteria they use to make marking judgements. These findings confirm students' perceptions that success partially depends on identifying each tutor's particular values for written work (Bloxham & West, 2007). It also suggests that students are well advised to listen carefully to tutors' oral guidance as well as paying attention to the written assessment artefacts.

This research, whilst providing some important lessons for the development of thinking about marking and quality assurance in relation to standards, is based on a small study in a limited number of disciplines. Further work is clearly needed to explore the topics within a wider range of disciplines including those in more determinate fields of study such as the natural sciences, medicine and engineering.

The authors would like to thank Prof. Susan Orr for her contribution to this research and her valuable feedback on drafts of this article.

References

- Alderman, G. (2009) Defining and measuring academic standards: a British Perspective, *Higher Education Management and Policy*, 21 (3), 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), 209-223.
- Bellingham, L. (2008) Quality assurance and the use of subject level reference points in the UK, *Quality in Higher Education*, 14 (3), 265-276.
- Biggs, J. (2002) *The reflective institution: assuring and enhancing the quality of teaching and learning*. Institute for Learning & Teaching in Higher Education. Available at: http://www.heacademy.ac.uk/resources.asp?process=full_record&id=109
- Bloxham, S. (2009) Marking and moderation in the UK: false assumptions and wasted resources, *Assessment & Evaluation in Higher Education*, 34 (2), 209-220.
- 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 forthcoming.
- Bloxham, S. & West, A. (2007) Learning to write in higher education: students' perceptions of an intervention in developing understanding of assessment criteria, *Teaching in Higher Education*, 12 (1), 77-89.
- Brennan, J. (1996) Introduction: The standards debate, in J. Brennan, *Changing Conceptions of Academic Standards*. London: Quality Support Centre, Open University., 9-26.
- Brennan, J., Frazer, M., Middlehurst, R., Silver, H. & Williams, R. (1996) *Changing conceptions of academic standards*. London: Quality Support Centre, Open University,
- Brown, R. (2010a) *Comparability of degree standards?* Oxford: Higher Education Policy Institute,
- Brown, R. (2010b) The current brouhaha about standards in England, *Quality in Higher Education*, 16 (2), 129-137.
- Coates, H. (2010) Defining and monitoring academic standards in Australian higher education, *Higher Education Management and Policy*, 22 (1), 41-58.
- Crisp, V (2008) Exploring the nature of examiner thinking during the process of examination marking, *Cambridge Journal of Education*, 38 (2), 247-264.
- Delandshere, G. (2001) Implicit theories, unexamined assumptions and the status quo of educational assessment, *Assessment in Education*, 8 (2), 113-133.
- Gibbs, G. (2010) *Dimensions of Quality*. York: Higher Education Academy,

- Gipps, C. (1999) Socio-cultural aspects of assessment, *Review of Research in Education*, 24, 355-392.
- Gilovich, T., Griffin, D., & Kahneman, D. (2002) *Heuristics and Biases: the psychology of intuitive judgements*. Cambridge: Cambridge University Press.
- Hartley, J., Trueman, M., Betts, L. & Brodie, L. (2006) What Price Presentation? The effects of typographic variables on essay grades, *Assessment & Evaluation in Higher Education*, 31 (5), 523-534.
- Harvey, L. (2002) Evaluation for what?, *Teaching in Higher Education*, 7 (3), 246-263.
- Harvey, L. & Williams, J. (2010) fifteen years of quality in higher education (part two), *Quality in Higher Education*, 16 (2), 81-113.
- HEFCE. (2009) *Report of the sub-committee for Teaching, Quality, and the Student Experience: HEFCE's statutory responsibility for quality assurance*. Bristol: HEFCE, (2009/40).
- Hinsliff, G. (2009) "Dumbing down" row over value of degrees, *Observer*, 2nd August
- HSLT (2009) *Assessment Standards in Hospitality, Leisure, Sport and Tourism (HLST) Programmes*. Oxford: Higher Education Academy, Hospitality Leisure and Tourism Subject Centre, Oxford Brookes University.
- Innovation, Universities, Science and Skills Committee. (2009) *Students and Universities*. London: House of Commons.
- Jawitz, J. (2009) Learning in the academic workplace: the harmonization of the collective and individual habitus, *Studies in Higher Education*, 34 (6), 601-614.
- Jawitz, J. (2007) New academics negotiating communities of practice: learning to swim with the big fish, *Teaching in Higher Education*, 12 (2), 185-197.
- Kangis, P. (2001) Presentational Dimensions and Marks Awarded to Assignments, *Quality in Higher Education*, 7 (3), 199-206.
- Knight, P (2006) The local practices of assessment, *Assessment & Evaluation in Higher Education*, 31 (4) 435-452
- Lave, J. & Wenger, E. (1999) Legitimate peripheral participation in communities of practice, in R. McCormick & C. Paechter, *Learning and knowledge*. London: Paul Chapman.
- Middlehurst, R. (1996) Degree standards and quality assurance: a discussion, in J. Brennan, *Changing Conceptions of Academic Standards*. London: Quality Support Centre, Open University., 71-96.

- Morgan, W. & Wyatt-Smith, C.M. (2000) Im/proper accountability: towards a theory of critical literacy and assessment, *Assessment in Education*, 7 (1), 123-142.
- 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), 205-217.
- Orrell, J (2003) Congruence and disjunctions between academics' thinking when assessing and their beliefs about assessment practice, in C. Rust (Ed) *Improving Student Learning: Theory Research and Scholarship*. Oxford: OCSLD
- Orrell, J. (2008) Assessment beyond belief: The cognitive process of grading, in A. Havnes & L.McDowell (Eds) *Balancing dilemmas in assessment and learning in contemporary education*. London: Routledge.
- Orr, S. (2007) assessment moderation: constructing the marks and constructing the students, *Assessment & Evaluation in Higher Education*, 32 (6), 645-656.
- Price, M. (2005) Assessment Standards: The Role of Communities of Practice and the Scholarship of Assessment, *Assessment and Evaluation in Higher Education*, 30 (3), 215-230.
- Price, M. & Rust, C. (1999) The Experience of Introducing a Common Criteria Assessment Grid Across an Academic Department, *Quality in Higher Education*, 5 (2), 133-144.
- QAA. (2010) *Evaluation of the Academic Infrastructure: final report*. Gloucester: QAA,
- QAA. (2010) *Who we are and what we do*. Available at:
<http://www.qaa.ac.uk/aboutus/whatwedo.asp>
- QAA. (2009) *Thematic enquiries into concerns about academic standards in higher education in England*. Gloucester: QAA,
- QAA. (2007) *The Classification of Degree Awards*. Gloucester: Quality Assurance Agency,
- QAA. (2006) *About the Academic Infrastructure*. Gloucester: Quality Assurance Agency, Available at: <http://www.qaa.ac.uk/academicinfrastructure/default.asp>
- QAA. (2006) *Code of Practice for the assurance of academic quality and standards in higher education. Section 6: Assessment of students*. 2nd edn. Gloucester: Quality Assurance Agency.
- Reimann, N., Harman, K., Wilson, A., McDowell, L. (2010) *Learning to assess in higher education: a collaborative exploration of the interplay of formal and informal learning in the academic workplace*. Lancaster 20 - 22nd July 2010.

- Ritchie, J. & Lewis, J. (ed.) (2003) *Qualitative Research Practice: a guide for social science students and researchers*. London: Sage.
- 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 & Evaluation in Higher Education*, 30 (3), 231-240
- Sadler, D.R. (2009a) Indeterminacy in the use of preset criteria for assessment and grading, *Assessment & Evaluation in Higher Education*, 34 (2), 159-179.
- Sadler, D.R. (2009b) Transforming holistic assessment and grading into a vehicle for complex learning, in G. Joughin, *Assessment, learning and judgement in higher education*. New York: Springer., 45-63.
- Sadler, D.R. (2005) Interpretations of criteria-based assessment and grading in higher education, *Assessment and Evaluation in Higher Education*, 30 (2), 175-194.
- Sadler, D.R. (1989) Formative assessment and the design of instructional systems, *Instructional Science*, 18 (2), 119-144.
- Sadler, D.R. (1987) Specifying and promulgating achievement standards, *Oxford Review of Education*, 13 (2), 191-209.
- Salmi, J. (2009) *The growing accountability agenda in tertiary education: progress or missed blessing?* Washington, DC: The World Bank, (Education Working Paper Series, No 16).
- Satchell, S. & Pratt, J. (2010) The dubiety of double marking, *Higher Education Review*, 42 (2), 59-62.
- Shay, S.B. (2004) The Assessment of Complex Performance: A Socially Situated Interpretive Act., *Harvard Educational Review*, 74 (3), 307-329; 23.
- Shay, S. (2005) The Assessment of Complex Tasks: A Double Reading., *Studies in Higher Education*, 30 (6), 663-679; 17.
- Silver, H. & Williams, S. (1996) Academic Standards and the external examiner system, in Brennan, J., et.al. *Changing conceptions of academic standards*. London: Quality Support Centre, Open University., 27-48.
- Vanderslice, R. (2004) When I was young,, an A was an A: Grade inflation in higher education, *Phi Kappa Phi Forum*, 84 (4 (Fall)),
- Wolf, A. (1995) *Competence-based assessment*. Buckingham: Open University Press.
- Woolf, H. & Cooper, A. (1999) Benchmarking academic standards in history: an empirical exercise, *Quality in Higher Education*, 5 (2), 145-154.

Yorke, M. (2008) *Grading Student Achievement in Higher Education*. London: Routledge.