

Assessment in higher education: the potential for a Community of Practice to improve inter-marker reliability

Abstract

The design, delivery and assessment of a complete educational scheme, such as a degree programme or a professional qualification course, is a complex matter. Maintaining alignment between the stated aims of the curriculum and the scoring of student achievement is an overarching concern. The potential for drift across individual aspects of an educational scheme (teaching, learning and assessment), together with emerging criticism in extant literature of the reliability of marking processes, suggests that, in practice, maintaining alignment might be more difficult than had previously been assumed.

In this paper, the concept of a Community of Practice (CoP) is employed as an analytical lens through which the notion of a markers' standardisation meeting that focuses on maintaining alignment between the curriculum, the marking scheme and the scoring of student scripts can be critically examined. Given that the aims and subject content of management learning are both multidimensional and contextual, such meetings have the potential to develop a shared approach to the elaboration and application of the marking scheme. A further role of the CoP is in the calibration of markers to accommodate further variations in student responses as they arise in the actual marking process. In this respect, the CoP has both descriptive and prescriptive potential in terms of aiding the development of markers of professional accounting examinations and also, we suggest, within accounting education more generally.

Keywords: Assessment, Alignment, Community of Practice, Marker Calibration.

Ian Herbert - Loughborough University Business School
Professor John Joyce & Professor Trevor Hassall - Sheffield Hallam University

Correspondence details:

Ian Herbert
Loughborough University Business School
Ashby Road
Loughborough
LE11 3TU
United Kingdom
Email: i.p.herbert@lboro.ac.uk
T: + 44 (0)1509 228833

Acknowledgements

The authors are grateful to Laurie McAulay and Marjahan Begum for comments on earlier drafts of this paper and to the editor and three anonymous reviewers for their contribution.

Introduction

The design, delivery and assessment of a management learning scheme is a complex matter, not least because such schemes are generally set within the context of situated practice that is both topical and evolving. Hence, the core aims of the curriculum, together with the more detailed intended learning outcomes (ILOs), are likely to be complex because they involve developing knowledge, cognitive capacity and transferable skills. Thus, when making scoring decisions, the assessment of student achievement is likely to require the application of both objective and subjective criteria.

There is increasing unease in the literature regarding the reliability of marking practices in high stakes assessment across secondary and higher education (e.g. Bloxham, 2009). This paper is ultimately concerned with the situation in higher education (HE), in which organisational relationships are becoming looser as a result of the 'casualisation' of faculty members to improve operational flexibility in an environment of constrained financial resources (Shepherd and Bowcott, 2010). With student populations becoming more ethnically and geographically diverse, both task requirements based on real-world scenarios and marking schemes need to be sensitive to a wide range of divergent answers.

When student cohorts are large, marking is invariably a team undertaking and a 'standardisation' meeting may be held in an attempt to establish consistency of marking amongst the team. Standardisation ensures that markers' perceptions and judgements are consistent in relation to an essentially common task. In accounting education, as in other social sciences subjects, tasks are usually context-based and the answers are consequently complex and divergent (Sadler, 2013). Thus, it is important to consider how markers might negotiate a shared meaning that will allow them to reliably judge a range of student responses independently. In this sense, markers become 'calibrated' to cope with variations in student answers rather than being 'standardised' to acknowledge only an 'ideal' answer.

The notion of a Community of Practice (CoP) (Lave and Wenger, 1990) has recently featured in the assessment literature in relation to the process through which markers might develop a shared meaning before marking (Reimann and Wilson, 2012; Janke et al., 2012). The intention of this paper is to explore the ways in which the CoP framework might facilitate a working consensus amongst markers within the context of a large scale markers' meeting of the type held by professional accounting bodies in order to bring about alignment between the ILOs and student achievement. Specifically, we ask how the dynamics within a markers' standardisation meeting (MSM) might support 1) the elaboration of the marking scheme based on the published model answer and 2) calibration of the markers such that they are better able to make reliable assessments across a range of student answers and approaches.

The paper proceeds as follows. First, the nature of alignment and marking reliability are discussed. Second, the concept of a CoP is introduced as an analytical framework to aid a marking team when negotiating shared meanings. Third, empirical evidence from a survey of markers working for a professional accounting body is presented. Fourth, the findings are discussed and some parallels between the case study and university degree programmes are highlighted. The paper ends with concluding remarks, including some of the limitations of the research and suggestions for further research.

The nature of alignment and the potential for drift

The extent to which student learning corresponds with the core aims of the curriculum is dependent on both constructive alignment and operational alignment. Constructive alignment is concerned with how the design of a learning and assessment programme enables students to internalise the ILOs and then to construct their own meaning from the knowledge content (Biggs, 1996). Operational alignment is concerned with how actual learning activities and assessment tasks reflect how tutors and markers interpret and then apply the ILOs across the various stages of learning: curriculum design, teaching and assessment (Herbert, Hassall and Joyce, 2009). According to Lines and Gammie (2004, p. 4), a good assessment task 'will be *valid* in that it will test what it sets out to test and *reliable* if the result will be exactly the same across all occasions, tasks, observations and settings' [italics added]. Yet, whilst these twin aims are in essence complementary, in practice they can oppose one another (Lambert and Lines, 2000).

Herbert et al. (2009) identify examples of drift in the use of task requirement verbs premised on the Bloom Cognitive Taxonomy (Bloom et al., 1956), noting that such drift has a propensity to be cumulative rather than self-correcting over subsequent stages. They conjecture that if such misalignment can occur in the highly structured and controlled assessment regime of their case study of a professional body, then there is potential for operational drift between ILOs and scoring student achievement in higher education more generally.

Ensuring reliable assessment

In recent years, both the concept and application of learning outcomes have received much attention in the literature (Bloxham and Price, 2013; QAA and HEA, 2013), with curriculum and assessment design being the main consideration and relatively less attention being given to how ILOs might actually connect with the practical issues involved in scoring students' individual achievement. However, interest in this issue is increasing in both the academic literature and the popular press. For instance, stakeholders in secondary (tertiary) education in the UK are making direct challenges, such as the 'Facebook' campaign questioning the alignment of ILOs in an element of the Biology 'A' Level examination (Coughlan, 2010) and the *post hoc* changing of the grade boundaries in English GCSE examinations (Kershaw, 2012).

Similar concerns in the HE sector have, thus far, largely remained within the affected academic communities. Indeed, the relatively sparse scholarly literature tends to approach the matter discretely and, usually, tangentially, citing conjectured concerns rather than actual problems. For example, Imrie (1995) notes that a mismatch between curriculum outcomes and the actual learning behaviours of students *could occur* if there is not a systematic framework and that a further mismatch is *likely to occur* in the marking process. However, some empirically based concern is starting to emerge. In a study across 10 subjects in seven English universities, Bridges et al. (1999) found that the higher the qualitative content was, the narrower the distribution of marks tended to be. Thus, their main conclusion is that:

'the existence of discipline related patterns of distribution is a potential cause of inequity of outcomes for students on joint or multidisciplinary programmes' (p. 286).

Baume et al. (2004) found differences between first and second marking occurring in degree programmes and suggest that:

'What the [task] requirement actually meant was not as transparent as the course designers thought it was....' (p. 456)

As might be expected, the Quality Assurance Agency [QAA] (2011) prescribe greater transparency in marking and assessment accountability, stating:

'Higher education providers should ensure that the assessment of students is robust, valid and reliable and that the award of qualifications and credit are based on the achievement of the intended learning outcomes.' (p.1)

Bloxham (2009) argues that the relative lack of discourse in HE should not be taken as a sign that all is well. She cites Price's (2005) observation that:

'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 and complex quality assurance procedures exist to support their application.' (p. 216)

The tradition of confidentiality regarding actual scoring decisions within institutions impedes more open discourse. In developing her argument, Bloxham (2009) draws on some emerging findings (Stowell, 2004; Shay, 2004, 2005 and Sadler, 2009) to describe the 'inherent frailty of marking practices and variability of standards', arguing that:

'...in developing rigorous moderation procedures, we have created a huge burden for markers, which adds little to accuracy and reliability but creates additional work for staff, constrains assessment choices and slows down feedback to students.' (p. 209).

Periodically, both student bodies and university management have demanded more objectivity in marking processes. This viewpoint has a certain intuitive appeal. If assessment and marking procedures are made explicit, then it follows that grade descriptors can both guide student learning and manage their expectations. Such demands reflect the widely held belief that marking is essentially based on both interpretivist perspectives and pragmatism (Bloxham, Boyd and Orr, 2011). According to this argument, task standardisation should enable more reliable marking. Put simply, the aim is to make the 'black box' of academic judgement and discretion transparent. However, Arnal and Burwood (2003, p. 379) counsel against the replacement of holistically fashioned, professional practice with 'cheaper, atomised, lower order activity' (see also Sadler, 2007).

It could be argued that at a broader level actual assessment strategy may involve markers taking a holistic approach and, subsequently, using marking criteria as a way of justifying (internally) and legitimising (externally) the actual score (Arnal and Burwood, 2003). Such an approach would likely include a combination of prerequisite technical knowledge and the marker's personal view of the inherent quality of the work. However, 'inherent quality' can be a slippery concept, as indicated by the much coined saying 'I know a 2:1 when I see it' (Ecclestone, 2001). More specifically, Hand and Clewes (2000) found evidence of the use of different criteria and approaches to marking by faculty staff when assessing the dissertation module within a final-year degree programme. Their study raised questions about the tension between individual tutor freedom and institutional demands, such as the pressure to meet marking deadlines and the demonstration of consistent assessment practice. They also drew attention to the opportunities for staff development through discussions of marking practice.

Maintaining alignment - Learning frameworks and marking procedures

Various conceptual frameworks have been proposed to both inform the process of curriculum design and to guide tutors and markers. A key feature of such approaches is the use of ILOs (cf. Eraut, 1989 p. 341-2). However, the ability of ILOs to feed into the overall aims of programmes has been challenged. For instance, Hussey and Smith (2008) argue that whilst ILOs are useful in guiding individual teaching sessions, their use in 'a whole programme of study leading to a qualification, such as a degree, constitutes a misuse.' (p. 114).

Yet, irrespective of the design framework, maintaining alignment between the ILOs and the scoring of student achievement requires a shared understanding of each element (questions, answers and marking scheme) by the entire marking team. When teams are large and come together only periodically, a standardisation meeting can ensure that the marking scheme actually used for marking (the elaborated scheme) can accommodate a sufficiently wide range of valid student approaches that extend beyond the scope of the official model answer.

However, 'standardisation' is perhaps not the most apposite word to describe a process that enables markers to accommodate divergent but valid answers and valid approaches to answering. An alternative term is 'calibration', which implies a shared understanding amongst tutors that enables them to score complex responses and effectively judge variations that might occur during the remote marking of the bulk of the scripts as well as carry out more holistic categorisation of marginal pass/fail scripts 'without constant engagement in moderation' (Sadler, 2013, p. 5). In other words, 'standardisation' infers constancy in marking around a common ideal response, whereas 'calibration' suggests that a common understanding is embodied within all the markers to enable them to judge divergent responses to complex tasks in a consistent manner. This nuance might appear superficial, but it is significant in practice. Nonetheless, the term 'standardisation' has a common currency in some assessment circles and, thus, it will be used herein as a label for markers' meetings seeking to achieve *reliability* in scoring student achievement.

The next section introduces the notion of a CoP in the context of a markers' standardisation meeting (MSM).

The markers' meeting as a Community of Practice

Rather than learning being essentially grounded in static, mainly explicit knowledge, learning in a CoP is a dynamic and developmental social activity grounded in situated practice. Lave and Wenger (1991) describe a process of 'legitimate peripheral participation', whereby an apprentice learns through first observing a master and then through engaging in supervised practice. Thus, the apprentice learns not only the technical skills of a trade but also those behavioural traits that encapsulate deeper, shared meanings that bond members of the group to each other whilst individual identities are respected. The authors argue that the process is mutually beneficial in that a master also refines his/her own learning through interaction with his/her apprentice(s).

Brown and Duguid (1991) explain how the evolving tacit knowledge of photocopier engineers, derived from their work experiences, was systematically codified and added to the explicit knowledge in the official field manual. They also note a further manifestation of group learning:

'Learners are acquiring not [only] explicit formal "expert knowledge" but the embodied ability to behave as community members.....The central issue in learning is becoming a practitioner not learning about practice.' (p. 48 – emphasis as original).

These ideas fit well with the notion that new markers learn from both the examiner and the experienced markers within a marking team. Likewise, the examiner, as 'the master', learns from the diversity of marker expertise, and this new knowledge feeds into the design of future questions, answers and marking schemes.

Wenger (1998) elaborated the CoP concept to create a framework of subsidiary components in order to explain how engagement in social practice is 'the fundamental process by which we learn and become who we are' (p. 1). This process of internalisation involves markers not only creating meaning

for a body of knowledge but, moreover, through participation, also creating an identity for themselves within the community through four constructs: *community* (learning as belonging), *practice* (learning as doing), *meaning* (learning as experience) and *identity* (learning as becoming) (*ibid.* p. 5).

In accountancy, the core aims of the curriculum are likely to be multi-dimensional and complex. It is necessary to apply theoretical constructs and technical knowledge to a real-world scenario before evaluating the results. In the case of professional examinations, the core aims relate very clearly to becoming a practitioner. Indeed, the professional body as a whole might be seen as a CoP. For example, the core aims stated in the syllabus of the Chartered Institute of Management Accountants (CIMA, 2010) are as follows:

'...assuring society that those admitted to membership are competent to act as management accountants for entities.... have adequate knowledge, understanding and mastery of the stated body of knowledge and skills... [and] have completed initial professional development and acquired the necessary work-based practical experience and skills (p. 6).

Within this broader framework, whilst aspiring to mastery within the profession and focusing on the specifics of the examination in hand, markers in a MSM are able to *practise* marking, give *meaning* to the marking scheme, *learn* from each other as a community and create an individual *identity*. The outcome of the MSM is an elaborated marking scheme and an agreed set of principles to be applied to further divergent answers (cf. Ardichvili, Page and Wentling, 2003). Bloxham et al. (2011) argue that the actual process of assessment is a social practice informed by situational judgements and that expertise develops through participation in a CoP. (Note: based on findings by Stowell, 2004; Shay, 2004, 2005 and Sadler, 2009).

Wenger (1998) describes three modes of belonging within the CoP: *engagement*, *imagination* and *alignment* (p. 173-4). Mutch (2003) argues that an individual's social and educational background will generate a disposition to act in similar ways in different contexts and that in order to resolve any conflicts, a process of negotiation and reconciliation is necessary. Lindkvist (2005) disagrees, seeing CoPs as tight-knit groups that share an understanding rather than give preference to dominant individual positions. However, he also acknowledges the potential for tension since at any one time individuals may be participants in several groups, each with different practices and identity structures; individual HE tutors play different roles in different assessment groups, acting as team markers, responsible examiners and maybe as external examiners in other institutions' departments. Wenger (1998) sees CoPs as relatively loosely based and acknowledges the possibility of conflict due to a lack of a shared understanding. Handley et al. (2006) conclude that tensions are likely to be continually renegotiated but never fully resolved.

Wenger (1998) makes the point that the continuity of *meaning* is likely to be problematic if *participation* prevails at the expense of *reification* or vice versa. The examiner's official answers are intended for publication so students can view them and show a single 'ideal' answer, sometimes noting alternative approaches. Within the MSM, the examiner's answers are elaborated to cover a wider range of possible solutions and approaches for the benefit of markers (i.e. the discussions are reified). However, at some stage, the lead marker has to decide if further elaboration will make the marking scheme more helpful or, alternatively, too granular and unwieldy for use in practice (cf. Sadler, 2009). If the marking scheme is too comprehensive, it may suggest to markers that other approaches (that might yet emerge) are not valid. Thus, an ideal scheme will guide markers in scoring diverse responses but will also allow them to adopt a more holistic perspective through the shared view of what is valid and what is not.

In the context of professional accounting examinations, markers use their tacit knowledge in conjunction with the explicit subject knowledge set out in the elaborated marking scheme (EMS) to

assess the extent to which students have demonstrated mastery in the application of financial techniques to a simulated real-world scenario. For example, in the financial evaluation of a business, the calculation of, say, a liquidity ratio is based upon explicit knowledge. Yet, the significance of the result will depend on cross references to other financial ratios *within* the overall business context. Consequently, in addition to explicit technical knowledge, markers will develop their own tacit understanding and marking skills by interacting with, observing, copying and imitating a master (Arnal and Burwood, 2003). Put simply, the process of marking is not dissimilar to other complex cognitive activities that benefit from 'learning by doing' in a CoP. Jawitz (2009) introduced the concept of 'learning to judge trajectories' (p. 604) through which new academic staff members grasp the implicit norms (i.e. criteria and benchmarks) of assessment through participating in a CoP.

Baird, Greatorex and Bell (2004) addressed the issue of inter-marker reliability in terms of 1) setting the right standard and 2) marking to that standard. They posited that a CoP could describe how a marking team can increase reliability through a standardisation meeting. They then conducted controlled experiments in relation to the grading of GCSE examinations using a sample of authentic previously marked scripts and two methods. In the first, markers were guided by different styles of exemplar script that had been marked by senior examiners, whilst in the second, markers were guided by discussions held within a 'co-ordination' meeting. The authors conclude that although neither approach had a significant effect on marking reliability, Hand and Clew's (2000) suggestion that a markers' meeting can play a role in training new markers was supported. Price (2005) suggests that the CoP has a role to play in harnessing the joint efforts of markers but that its effect depends upon the extent to which individuals are prepared to engage and share knowledge.

It should be noted that there is no reference in the literature to the role a marking CoP plays in maintaining alignment by challenging the validity of the examiner's answers and marking scheme or, indeed, by recognising alternative valid approaches in divergent student answers. There seems to be an inherent assumption that the questions, answers and marking scheme are correct and that the only challenge is to ensure inter-marker reliability. This may be the case in HE, where one person or a small team undertake the design, teaching and assessment stages, but in a professional body with many people involved in each stage, there is likely to be a significant level of challenge to validity, resulting in new ideas of what is valid and what is not.

Roberts (2006) and Contu and Willmott (2003) both note that negotiations within the CoP will be influenced by power that depends upon factors such as an individual member's age, expertise, personality and authority. In HE, power within the CoP may also be the result of an individual's line management position. Whilst the notion of the CoP may infer a sense of holistic group thinking, a management theorist might delineate the decision process into democratic, autocratic and technical expert approaches. Although, Roberts (2006) emphasises the importance of trust between colleagues in the process of negotiation within a CoP, and notes that such trust indicates a high degree of mutual understanding between the individuals concerned. To conclude this section, Baird et al. (2004) assert that 'the CoP literature has great descriptive utility, but its prescriptive utility has yet to be established' (p. 346). A case study is now introduced to demonstrate how a MSM might operate and how the CoP might both explain and guide its operation.

Case study

A professional institute of accounting makes an interesting case study because in many ways the assessment process within this environment is a microcosm of the assessment process in higher education, albeit that a professional body tends only to be concerned with designing the curriculum and assessment and not with teaching. In the case study organisation, student numbers are relatively high (typically around 5,000 students for each subject, with two exam diets a year). Thus, the marking team for each subject (typically 15-30 markers from a variety of backgrounds) only convenes on two

occasions each year. Furthermore, the turn-around period is relatively short, with bulk marking taking place remotely and individually.

Typically, a MSM will take an entire day, thus bringing into sharp focus how both individual expertise and team cohesion can be rapidly forged and maintained within the framework of the CoP. Whilst not all the features of a professional MSM may not be present in an HE marking MSM, the present trend in HE is towards larger cohorts, greater use of casual staff, more home/remote working and tighter deadlines, reflecting the situation in the professional environment. Moreover, most HE institutions that offer accounting and business programmes seek to maximise the number of exemptions from professional qualifications. Thus, programme directors are keen to assimilate the subject content, core aims and assessment styles of the various professional bodies into their degree programmes. The participants and processes of the MSM in the case study are tabulated in Appendix 1.

Method

The authors have gained significant cumulative experience through holding various positions in the MSMs of professional bodies; however, in order to compare and supplement their own experiences, semi-structured interviews were used to capture and clarify views across the range of roles in a selected marking team. A mid-level management accounting paper was selected for study because this involved a mix between quantitative, mainly technical tasks (one right answer but maybe different approaches) and complex, qualitative, mainly discursive tasks (many right answers and multiple approaches).

Five markers were chosen at random from a team of 20, together with the examiner, the two lead markers (LM1 and LM2) and a monitor marker (MM), making a total of nine interviewees from the MSM. The participants came from a variety of employment backgrounds: public university, private sector college and commercial organisation employees (one from the latter), self-employed tutors and retirees (ex-HE). Because of their geographic dispersion, interviews were conducted by telephone. Two researchers conducted the first four interviews to establish a consistent approach, whilst a single researcher interviewed the remaining participants. Two interview formats were used: one for markers (see appendix 2) and a second, containing supplementary questions about the process of managing the marking team, for the examiner and the two lead markers. An additional interview (semi-structured) was subsequently conducted with the director of examinations to explore the extent to which the subject findings resonated with the wider curriculum within the institute.

The telephone interviews typically lasted between 30 and 40 minutes and were recorded (with permission). A full transcript was prepared by a professional typist and then sent to the respective interviewees for verification. This often resulted in some helpful clarification and further insights. All names were disguised. A summary was produced by one of the research team, and this was then discussed with the other two researchers to arrive at an agreed interpretation of the findings.

Participants were asked, through a series of structured questions and more open ended prompts, to describe how an elaborated marking scheme is negotiated by markers, and how the dynamics of the deliberations foster a shared meaning that can be used to interpret divergent answers. Thus, alignment across 1) the core aims of the curriculum, 2) technical knowledge (subsumed in a specific subject syllabus) and 3) situated practice is maintained by ensuring that the examiner's answers and EMS are valid and that inter-marker reliability is satisfactory.

Findings

Coping with answer diversity

The divergence of students' answers and approaches in the initial batches of scripts received along with conjectured permutations that might be encountered in the future forms the basis of the discussions in the MSM. The following comments reflect the diversity of students' real-world approaches and how alternative valid interpretations, approaches and solutions need to be accommodated.

"The LMs put in a lot of preparation beforehand in selecting suitable scripts that will identify issues".

LM 1

"You see different ways of approaching questions, of approaching areas of the syllabus that perhaps I hadn't thought about previously. There is more than one way of 'skinning a cat' [in management accounting] as opposed, say, to the rather formulaic and programmed methodology of financial accounting."

Marker 4

"People are very knowledgeable. When things are discussed, people often say 'you could look at this from this perspective or you could look at it from that perspective or it could be interpreted like this'. So there tends to be a lot of examination of different ways of doing things, different business issues that could relate to a particular scenario. There's a lot of discussion on all of those things."

Marker 1

Marker diversity

Markers negotiate a shared meaning from their pooled world experience. Previous experience of marking and the personal background of each individual team member are relevant factors when interpreting a) the marking scheme, b) students' answers and c) the validity of further approaches. The experienced team members felt that knowledge of the subject area was more critical than the current vocation (technical) of the marker.

"There's quite a few people who I consider to be much more knowledgeable on some aspects of the syllabus. [For instance] when it's something relatively obscure, say, a finer point of financial accounting for which I don't have the background. I do find it helpful to hear the views of others on that. I think that the sample scripts give us a really good focus. You get discussion about each of the questions led by the people who know how the process works."

Marker 3

"I haven't detected any differences between where they come from [vocational background]. As long as they've had plenty of experience in the subject area, that's the main thing really."

LM 1

Marker development

All the markers in the survey felt that the MSM was a valuable learning experience.

"Yes, I usually learn something. Maybe not masses, but you get different people there. Some of those with a very strong maths background can develop my understanding. My technical understanding has also developed somewhat."

Marker 5

"I think [I learn] more if it's a topic that's got real-world relevance because I've got twenty-odd years teaching experience but very little or no real-world experience. So someone who has got some relevant real-life experience to bring to the table is useful for me. Technically I'd like to think that I'm probably fairly okay because I've done it for so long. ... all that I do is teach and mark."

The comments above indicate that different levels of practical experience and subject specialism can enrich the MSM. The senior members of the team (as masters) also learn from hearing the views of other participants at the meeting, confirming Lave and Wenger's (1991) claim that masters learn from their apprentices.

"There are some very experienced markers. A lot of them deliver [teach] this paper or the equivalent [for another professional body]. Especially one who's spot on with maths. He can run around me in circles on maths, and I always pick up something from him. I've got no embarrassment saying 'X, talk me through that. I don't know that'. You've also got people like Y., He gives everybody a bad ride, and I expect that from him. But 99% of what he says is excellent I've learnt something every time to be honest."

LM 1

Certainly, practitioners who are out there doing things [in the real world] often add [something extra] to [the discussion], and inform me, in certain areas.

Examiner

"I think the MSM is really important ... if people go away saying, 'I've got this clearly in my head now and I know what I'm doing', [then] I think that's what the objective is. I've a feeling that after the meeting most people feel really confident, and I don't think they arrived at the meeting [feeling] confident."

LM 2

Elaboration of the marking scheme

The appropriate level of detail necessary for the elaborated marking scheme is a moot point. If there is not enough detail, less experienced markers will remain uninformed. However, if there is too much detail, it could deter markers from rewarding alternative approaches. The MSM is also a vehicle for the markers' pool of wisdom that can be transported into future meetings.

"I think they [answer and marking scheme] are about right, but I have to say there are people on the marking team who would prefer much more detail on exactly how marks should be awarded."

LM 1

"I think they (answers and marking scheme) are about right at the end of the meeting. Beforehand, they are reasonably detailed, but they don't include fine tuning points, such as how many decimal places we expect an answer to be to and how we're going to deal with roundings."

Marker 1

"I would say [it's] marginally under elaborated. I think just having seen other marking schemes, perhaps they don't cover as many options as they could possibly around the question. I would like to see differing approaches being documented."

Marker 2

"I think most people are so experienced that they know what works, and you don't have to discuss exactly how to round every single number and such things. [There are] a lot of principles that we've developed over the years on how to approach certain things."

LM 1

Achieving Reliability

The senior members of the team all felt confident about the reliability of the marking and believed that the MSM was vital to ensure reliability. It appeared to play a number of roles; encouraging diversity was considered and also moderating the views of individuals to cohere with the consensus view.

"... generally, I'm quite confident about what goes on in this institute, having had experience of other exam boards. I have a lot of confidence in the result that comes out."

LM 1

This was confirmed by other markers:

"At the meeting, I think it (the marking scheme) becomes just right. Prior to the meeting, I think it is under elaborated, and I don't think you could possibly mark without the standardisation meeting."

Marker 1

"The markers' meeting is important because I think it's creating a [common] culture. If that's written on a piece of paper, that's fine, but if there is peer pressure within the group that you are the person that's out of sync, ... then that is a powerful message and it changes attitudes and thinking. Also, the way that things are clarified; if people do not understand something, then it's raised [as an issue] and people appreciate the clarification that comes through the meeting."

LM 2

Reaching agreement - power and democracy

Given the divergent nature of this subject, it is likely that there will not be ready agreement on every issue. We enquired how decisions were actually made and whether things were resolved through democratic, expert or hierarchical means.

"I think it's fairly democratic. Occasionally, when there's a complete disagreement, the examiner will have to give the final say, but that's quite rare. Most of the time, it's a fairly consensual agreement. I would describe it as democratic."

Marker 3

"Sometimes there is a lot of discussion and then somebody has to take a lead and say 'right this is the ruling because the discussion could go on forever'. I think sometimes if it's a simpler matter, then it is just hierarchical and we're told what to do. I think there's always someone [an expert] who does say 'this is what will happen', and when it does, it tends to come from the examiner via the LMs. I tend to think that's the way it goes, but discussions are definitely heard, most definitely."

Marker 1

"... the markers' meetings are based on a consultative dictatorship. There is a full and frank exchange of views, but ultimately the examiner has the final say and [he/she] stands or falls by that decision. I think usually though it's fair to say that the strength of the arguments raised determines the outcome. I don't think it's very often that we get a situation where there is a major outstanding issue of principle and we just have to say 'well OK, let's adjudicate via the examiner'. Discussions have to be terminated at times, but ultimately it is a consultative dictatorship; it is not a straight democracy."

Marker 5

On the basis of some of the answers given by participants, it is interesting to try and identify the underlying power dynamics. The first comment below suggests (as one might expect given the technical orientation of the team) that expert power is deferred to. The second demonstrates the role of leadership in mediating and moderating the meeting.

"There is a fair amount of respect in that room in both directions, and there are some very sharp individuals in the marking team who could easily be sitting at the top table [senior executive level]."

Marker 4

"I think it's the way the meetings are chaired, such that nobody feels that they can't raise points, and we do have a good debate. But equally the debate is kept relevant, and isn't allowed to carry on too long. Eventually we have to say 'Well what is the consensus view on this? Right, okay, is everybody happy we're going to go down this route?' So it's getting that balance, and I think LM1 in particular does a very good job on that."

Marker 2

As might have been anticipated, the director of examinations tended to give a more 'managed' view of the MSM in terms of how decisions are actually reached.

"To ensure fairness, it might be the case that when certain unusual answers have cropped up or students have actually tackled a question in a different way than was envisaged by the examiner, this might need to be addressed."

"The markers have an opportunity to discuss the issues as they go through. When I say 'discuss', it's really led by the lead marker for consistency, so it's controlled. It's not a sort of free-for-all situation. The aim is that at the end of the day everybody goes away able to mark to the same standard."

"It is not acceptable for the lead marker to try and operate a democracy at a meeting and change things as a result of a lot of pressure. The position can be difficult sometimes."

"I think the dynamics [of the team] and the personalities of the markers make it interesting, but, predominantly, if you have a good lead marker, they will apply their authority."

We asked the director of examinations about the extent to which markers might refer back to the core aims of the curriculum.

"...particularly looking at border line scripts, the holistic view is that if this person were to pass this paper, would they be a danger to the profession? And those actual words are used at the meetings. So the answer is 'yes'."

"One of the concerns that may be raised is that students can actually tot up marks [for technical points] from the marking scheme and get very close to a pass mark. The question then is, would you say, as a board member [of an organisation], if this was placed in front of you as a report, what would you think of it? That's where a more holistic approach is applied."

Overseas scripts

The questionnaire also probed the interpretation of overseas scripts. The issues involved, such as use of English and local business practices/examples, are complex. In the anonymous marking regime operated by professional bodies, the very definition of an overseas script (ethnic background, vocational situation or actual assessment domain), and thus the principle of whether to allow wider interpretation of responses, is contested territory. However, limited space here prevents discussion of the issues or presentation of the markers' views.

Summary

Given that markers must allocate a whole day to attending the MSM, plus spend time travelling, for which they receive no direct remuneration, one might have expected there to have been comments along the lines of “I know what I’m doing and would prefer just to get on with the marking!” However, that view was absent in the interviews. All participants felt that the MSM played a vital role in ensuring reliability and that the discussions developed their own learning. Given that professional accounting examinations might be seen by many as more technical and less conceptual than is the case with examinations in undergraduate degree programmes, it is worrying that MSMs of the type described here are deemed necessary when in universities there tends to be relatively little such deliberation before marking commences. The next section discusses the findings in the context of the assessment and CoP literature.

Discussion

Maintaining alignment

The case findings suggest that the challenge of achieving efficacy in marking is multi-faceted. Primarily, there is a need for consistency between markers (inter-marker reliability) and between scripts across the bulk marking process (individual marker reliability). However, achieving reliability suggests ‘standardisation’, and this raises the question: standardisation against what?

Negotiating shared meanings within a MSM is not just about coercing outlying markers towards ‘the middle’ but also involves challenging at a more fundamental level what marks should be awarded for. In other words, it is necessary to first establish what alternative approaches and answers are valid. For instance, discussions within the MSM might highlight that the question may have been open to interpretation by students in ways that had not previously occurred to the examiner. Maybe the examiner’s model answer was incorrect or incomplete in some way or the original mark allocation was inappropriate in relation to the difficulty or ease of actually answering the question. This concurs with the point made by the director of examinations that students might be able to ‘tot up’ marks for technical aspects whilst displaying little holistic understanding. Indeed, it is often the case that an ‘outlying’ score is successfully justified by the individual marker and, thus, the marking ‘standard’ moves to a new level.

The comments made by the participants in the case study MSM indicate that there is considerable discussion on the principles according to which marks should be awarded and that this involves referring back to the core aims and the intermediate elements of the assessment process in order to develop the final elaborated marking scheme. The steps in the assessment process and the negotiation within the CoP are illustrated in Figure 1. References back to the published curriculum and examination documentation are termed ‘feedback control’ because the intention is to align marking with the stated intentions and principles. Within the CoP, the process is dynamic and acknowledges new valid responses; we use the term ‘feed forward control’ for this situation.

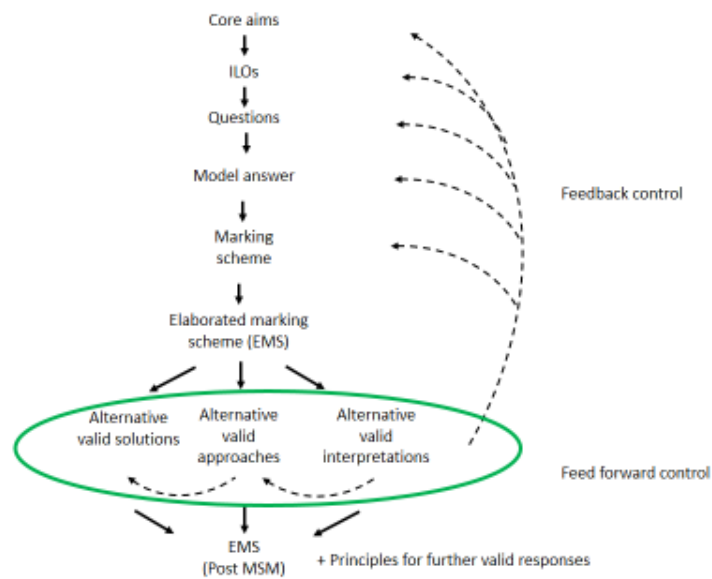


Figure 1 Negotiation of the Elaborated Marking Scheme

Lambert and Lines (2000) argue that whilst the twin aims of validity and reliability are in essence complementary, in practice they can oppose each other. The findings from the case study MSM suggest that the two constructs are intertwined in a dynamic process within which answer validity is negotiated backwards to the core aims and through the individual assessment elements and forwards into divergences yet to emerge. The objective is to produce an (explicit) elaborated marking scheme (EMS) that represents a practical, albeit not exhaustive, instrument to guide reliable marking in conjunction with the (tacit) embedded principles shared by the calibrated markers.

Further causes of drift that might occur over the course of marking that were not specifically probed in the case study are: marker fatigue, forgetting the agreed principles or over reaction to a run of good or poor scripts. In the experience of the authors, it is often the case that in the MSM itself markers who may be high or low in one set of control script comparisons then over compensate when marking the next control script. Thus, alignment can drift in the bulk marking phase and, hence, the professional body continues to sample mark throughout the process.

CoP processes and power

The case study reflects the broad components of Wenger's (1998) social theory of learning, which are learning as *belonging, doing, becoming* and *experience*. The published answers and the examiner's marking scheme are interpreted and negotiated through a process of participation. The EMS is a reification of the markers' collective knowledge derived from situated practice and the potential range of student responses. Formal knowledge is supplemented with tacit knowledge in the meeting and codified into the EMS (Brown and Duguid, 1991). Understanding divergent answers and anticipating further permutations requires *imagination, alignment* and *engagement* on the part of team members.

However, there is also some evidence of variation and intra-community conflict (Hardy et al., 2006), which appears to be resolved by the process of negotiation and reconciliation (Mutch, 2003). The case study MSM benefits from a settled team, and it is worth noting that the curriculum is also settled (mid-way through the five-year revision cycle). According to the participants, any conflict is resolved through the exercise of power based upon the authority of the executive (lead) markers or the

expertise of the examiner or markers with specialist technical/vocational backgrounds, e.g. in maths (Contu and Wilmot, 2003; Roberts, 2006).

The relevance to higher education

Whilst noting that evidence on the actual efficacy of marking practices in higher education is somewhat limited due to confidentiality issues, nonetheless, concern is emerging in the literature that there may be instances of drift in the alignment between the stated aims of the curriculum and actual student achievement. There are also questions about inter-marker reliability (see Sadler, 2009 and 2013), specifically in the context of accounting (Watty et al., 2013). When investigating the somewhat intensive MSM associated with professional examinations, the CoP provides an analytical framework to aid an understanding of the way in which issues are discussed and resolved within the marking team and, thus, how response validity might be tested and inter-marker reliability might be improved through the EBS (Wenger, 1998). Further benefits associated with the MSM are 1) new markers are inculcated into the team, 2) markers practise marking together and 3) shared meanings and marking principles are developed to guide individuals in the bulk marking phase.

Whilst the various marking processes, together with the various quality checks included in the wider control system in a professional body, may not be practical/cost effective for, say, a degree programme of 100 students, nonetheless, some first year modules in HE now have cohorts of 1,000+ students. Even for such relatively modest numbers, the importance of ensuring marker reliability in high stakes assessment is growing as the proportion of casual to tenured staff continues to rise and, hence, the ongoing CoP effect between colleagues diminishes. Whilst noting that Price (2005) found that the engagement of individuals in CoPs was rather disappointing, we believe that the findings in this enquiry are very positive, both in terms of marker comments and the fact that the professional body continues to allocate what amounts to significant resources for MSMs. We believe that the CoP model and the processes of the MSM should be of interest to the accounting education community in degree programmes. One way forward might be for faculty management to encourage staff to engage in external marking activities as a way of improving their technical knowledge and broadening their experience of peer group moderation. In other words, it may be beneficial to bring the experience of the professional MSMs into HE, even where it may not be practical or cost effective to operate 'standardisation' events in-house.

The counter argument put forward by HE institutions might be that large-scale, faculty-wide ratification boards are already held after assessment scoring and such boards are able to view an individual student's performance holistically in terms of their personal background, latent ability and study trajectory, whilst the facility to take into account a student's background and character is absent in professional assessment. However, HE ratification boards can do little directly to achieve alignment within each subject.

Conclusion

The CoP described by Lave and Wenger (1991) is essentially a group of people developing shared meanings but also individual skills and identity. The findings suggest that there is evidence of both aspects of a CoP, with new members developing through participation in the community and experienced members refining their own learning through interaction with their peers.

Even in the highly structured context of a professional body, a significant number of answer variations arise in each examination that require negotiation and agreement. The solution is to elaborate the marking scheme to provide for a sufficient range of alternative answers, embedding in markers a set of principles that enable them to handle emerging variations as the marking proceeds. The informal processes that occur around and within the MSM are, we argue, central to the aim of maintaining alignment between the stated aims of the curriculum and the scoring of student achievement.

The CoP within the MSM provides a framework for the processes by which a marking scheme for each examination is made valid (aligned with the core aims) and more reliable through inter-marker reliability. The MSM is also helpful in that it can play a part in inculcating new markers into the ethos of a marking team and can enable the examiner, lead markers and experienced markers (as masters) to learn from students as apprentices in a learning process (prescriptive). We suggest that due to the calibration and development of markers that occurs within the MSM, it is both a significant quality mechanism and an important vehicle for tutor development.

Some colleagues may not accept our view that the findings have applicability to the HE sector. However, we argue that the principle of challenge to the examiner's materials by the variety of technical and vocational expertise in the MSM, together with the way in which the CoP calibrates and develops markers, should at least serve as a benchmark for HE. Moreover, there is an opportunity for HE to see the value in providing markers to professional MSMs and to research the processes and results in depth.

Given that this study and previous studies in HE have had to stop short of revealing sensitive results data, it is suggested that more detailed empirical evidence could be publishable if a number of professional bodies and/or HE organisations could be persuaded to take part in a collective survey that would allow institutional anonymity to be preserved. It would also be useful to explore in depth the principles and practical issues involved in the marking of scripts in an organisation that has students from culturally diverse backgrounds.

References

- Anderson, L.W. and Krathwohl, D.R. (Eds.) (2001) *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives* (New York: Longman).
- Ardichvili, A., Page, V. and Wentling, T. (2003) Motivation and barriers to participation in virtual knowledge-sharing communities of practice, *Journal of Knowledge Management*, 7(1), pp. 64-77.
- Arnal, S.G. and S. Burwood (2003) Tacit knowledge and public accounts. *Journal of Philosophy of Education*, 37(3), pp. 377-91.
- Baird, J., Greatorex, J. and Bell J.F. (2004) 'What makes marking reliable? Experiments with UK examinations', *Assessment in Education*, 11(3), pp. 331-348.
- Barrie, S. C. (2007) A conceptual framework for the teaching and learning of generic graduate attributes, *Studies in Higher Education*, 32(4), pp. 439-458.
- Baume, D. and Yorke, M. with Coffey, M. (2004) What is happening when we assess and how can we use our understanding of this to improve assessment? *Assessment & Evaluation in Higher Education*, 29(4), pp. 451-477.
- Biggs, J. B. (1996) Enhancing teaching through constructive alignment, *Higher Education*, 32, pp. 347-364.
- Biggs, J. B. (1999) *Teaching for quality learning at university: what the student does* (Buckingham: Society for Research into Higher Education).
- Bloom, B. S. (Ed.), Englehart, M. D., Furst, E. J., Hill, W. H. and Krathwohl, D. R. (1956) *Taxonomy of educational objectives. The classification of educational goals. Handbook 1: Cognitive domain* (New York: David McKay).
- Bloxham, S. (2009) Marking and moderation in the UK: false assumptions and wasted resources, *Assessment and Evaluation in Higher Education*, 34(2), pp. 209-220.
- Bloxham, S., Boyd, P. and Orr, S. (2011). Mark my words: the role of assessment criteria in UK higher education grading practices. *Studies in Higher Education*, 36(6), pp. 655-670.
- Bloxham, S. and Price, M. (2013) External examining: fit for purpose, *Studies in Higher Education*, i-first DOI 10.1080/03075079.2013.823931

- Bridges, P., Bourdillon, B., Collymore, D., Cooper, A., Fox, W., Haines, C., Turner, D., Woolf, H. and Yorke, M. (1999) Discipline-relayed marking Behaviour Using Percentages: a potential cause of inequity in assessment, *Assessment and Evaluation in Higher Education*, 24(3), pp. 285-300.
- Brown, J.S. and Duguid, P. (1991) Organizational Learning and Communities of Practice: toward a Unified View of working, Learning, and Innovation, *Organization Science*, 2(1), pp. 40-57.
- Chartered Institute of Management Accountants (2005) *Qualification and Syllabus Structure* (London: CIMA).
- Coughlan, S. (2010) Facebook campaign against A-level exam paper, BBC Education Online, 26.1.14, <http://newsvote.bbc.co.uk/1/hi/education/8480563.stm>
- Contu, A. and Willmott, H. (2003) 'Re-embedding situatedness: the importance of power relations in learning theory'. *Organization Science*, 14(3), pp. 283-96.
- Delandshere, G. (2001). Implicit theories, unexamined assumptions and the status quo of educational assessment. *Assessment in Education*, 8(2), pp. 113-33.
- Ecclestone, K. (2001) 'I know a 2:1 when I see it': understanding criteria for degree classifications in franchised university programmes, *Journal of Further and Higher Education*, 25(3), pp. 301-313.
- Eraut, M. (1989) *The International Encyclopaedia of Educational Technology*, pp. 338-352 (London: Pergammon Press).
- Eraut, M. and Cole, G. (1993) Assessment of Competence in Higher Level Occupations, *Competence and Assessment*, 21, pp. 10-14.
- Hall, C. and Johnson, A. (1987) Planning a test or examination, Module A5 in Imrie, B. W. and Hall, C., *Assessing Student Performance* (Wellington, Authority for Advanced Vocational Awards).
- Hand, L. and Clewes, D. (2000) Marking the difference: an investigation of the criteria used for assessing undergraduate dissertations in a business school, *Assessment & Evaluation in Higher Education*, 25(1), pp. 5-21.
- Handley, K., Sturdy, A., Fincham, R. and Clark, T. (2006) Within and Beyond Communities of Practice: Making Sense of Learning Through Participation, Identity and Practice, *Journal of Management Studies*, 43(3), pp. 641-653.
- Herbert, I.P., Joyce, J. and Hassall, T. (2009) Maintaining alignment in management education: the potential for drift in assessment, *International Journal of Management Education*, 7(2), pp. 21-31.
- Hornby, W. (2003) Maintaining Standards? An Analysis of Assessment Practices in a Business School, *The International Journal of Management Education*, 3(3), pp. 3-17.
- Hussey, T. and Smith, P. (2008) Learning outcomes: a conceptual analysis, *Teaching in Higher Education*, 13(1), pp. 107-115.
- Imrie, B. W. (1995) Assessment for Learning: quality and taxonomies, *Assessment & Evaluation in Higher Education*, 20(2), pp. 175-189.
- James, M. and Brown, S. (2005) Grasping the TLRP nettle: preliminary analysis and some enduring issues surrounding the improvement of learning outcomes, *The Curriculum Journal*, 16(1), pp. 7-30.
- Janke, K.K., Seaba, H.H., Welage, L.S., Scott, S.A., Rabi, S.M., Kelley, K.A. and Mason H.L. (2012) Building a multi-institutional community of practice to foster assessment, *American Journal of Pharmaceutical Education*, 76(4), pp. 1-7.
- 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.
- Kershaw, A. (2012) 'Tens of thousands' affected by GCSE grade boundary changes say heads, *Independent*, 29.8.12. <http://www.independent.co.uk/news/education/education-news/tens-of-thousands-affected-by-gcse-grade-boundary-changes-say-heads-8092539.html>
- Lambert, D. and Lines, D. (2000) *Understanding assessment: Purposes, perceptions, practice*. Key issues in teaching and learning (London: Routledge/Falmer).

- Lave, J. and Wenger, E. (1991) *Situated Learning: Legitimate peripheral participation* (Cambridge: Cambridge University Press).
- Lindkvist, L. (2005) Knowledge communities and knowledge collectivities: a typology of knowledge work in groups', *Journal of Management Review*, 42(6), pp. 189–210.
- Lines, D. and Gammie, E. (2004) *Assessment Methods Report* (New York: IFAC).
- Mutch, A. (2003) Communities of practice and habitus: a critique, *Organization Studies*, 24(30), pp. 383–401.
- Price, M. and Rust, C. (1999) The Experience of Introducing a Common Criteria Assessment Grid Across an Academic Department, *Quality in Higher Education*, 5(2), pp. 133-144.
- Price, M. (2005) Assessment standards: the role of communities of practice and the scholarship of assessment, *Assessment & Evaluation in Higher Education*, 30(3), pp. 215–230.
- Quality Assurance Agency (QAA), (2011) Chapter A6: Assessment of intended learning outcomes, *QAA. Amending Code of Practice for the assurance of academic quality and standards in higher education* (2006).
<http://www.qaa.ac.uk/Publications/InformationAndGuidance/Documents/Quality-Code-Chapter-A6.pdf> accessed 8.11.12.
- Quality Assurance Agency for Higher Education (QAA) and the Higher Education Academy (HEA) (2013) External examiners' understanding and use of academic standards, <http://www.qaa.ac.uk/Publications/InformationAndGuidance/Documents/External-Examiners-Report.pdf> accessed 16.6.14
- Ramsden, P. (1992) *Learning to Teach in Higher Education* (London: Routledge).
- Reimann, N. and Angelina Wilson, A. (2012) Academic development in 'assessment for learning': the value of a concept and communities of assessment practice, *International Journal for Academic Development*, 17(1), pp. 71-83.
- Roberts, J. (2006) Limits to Communities of Practice, *Journal of Management Studies*, 43(3), pp. 624-639.
- Shay, S.B. (2004) The assessment of complex performance: A socially situated interpretive act, *Harvard Educational Review*, 74(3), pp. 307–29.
- Shay, S. (2005) The assessment of complex tasks: A double reading, *Studies in Higher Education*, 30(6), pp. 663–79.
- Sadler, D. R. (2007) Perils in the meticulous specification of goals and assessment criteria, *Assessment in Education: Policy & Practice*, 14(3), pp. 387-392.
- Sadler, D. R. (2009) Indeterminacy in the use of preset criteria for assessment and grading, *Assessment & Evaluation in Higher Education*, 34(2), pp. 159-179.
- Sadler, D.R. (2013) Assuring academic achievement standards: from moderation to calibration, *Assessment in Education: Principles, Policy & Practice*, 20(1), pp. 5-19.
- Shepherd, J. and Bowcott, O. (2010) *Thousands to lose jobs as universities prepare to cope with cuts*, (London: The Guardian) <http://www.guardian.co.uk/education/2010/feb/07/job-losses-universities-cuts> accessed 9.2.10.
- Stowell, M. (2004) Equity, justice and standards: assessment decision making in higher education, *Assessment & Evaluation in Higher Education*, 29(4), pp. 495-510.
- Watty, K., Freeman, M., Howieson, B., Hancock, P., O'Connell, B., de Lange, P. and Abraham, A. (2013) Assessment & Evaluation in Higher Education Social moderation, assessment and assuring standards for accounting graduates, *Assessment & Evaluation in Higher Education*, DOI: 10.1080/02602938.2013.848336. <http://dx.doi.org/10.1080/02602938.2013.848336> In press.
- Wenger, E. (1998) *Communities of Action, Learning, meaning, and identity*, (Cambridge: Cambridge University Press).

Appendix 1

Markers' Standardisation Meeting – participants and process

The examiner (E) has previously written the examination paper and answers together with a marking scheme for subsequent publication to students. Various technical experts and an assessment panel have considered the appropriateness (validity) and technical accuracy of the assessment through various drafts over an 18-month period. Examiners may also choose to mark some scripts.

Two lead markers (LM1 and LM2) each supervise a team of around ten markers. One LM will chair the MSM.

A monitor marker (MM) attends the meeting as an observer and then marks around 10-12 scripts sampled at random from across the time period of marking, i.e. sensitive to both inter-marker reliability and drift over time. The MM then writes a summative reflection for the director of examinations, who has overall responsibility for the governance of assessment processes.

Phase	Action
Preparation	One batch of scripts (25) is sent from exam centres to each marker.
	LM1 selects four control scripts representing a range of student achievement and emails as pdfs to each marker.
	Markers score two of the control scripts and review but do not mark the first batch of scripts.
	LMs and the examiner (E) meet to elaborate the marking scheme in the light of early scripts. The examiner then explains each question, answers and mark allocation to the meeting.
Standardisation	Individual scores from the first of the two marked control scripts are written up on a flip chart. Markers at the top and bottom of the range are asked to justify their individual positions. At this stage, the overall score range can easily be around 10 percentage points, say on a score of 50%. Following a discussion (maybe of around one hour for this first script), a mark is agreed and the marking scheme is further clarified and/or elaborated.
	Second script marks are put up and discussed.
	A third control script is marked in pairs (to increase calibration) and scores are discussed.
	A fourth control script is marked. If there is already a consensus on some questions, these may be omitted.
	Up to two further scripts/questions may be marked until a consensus has been reached on all questions.
Ratification	Markers then return home and mark the first batch of scripts and send them to their designated LM.
	LMs check the marking, make recommendations to markers and then give clearance to continue. There may be further, usually minor, clarifications and amendments.
	Bulk marking. Further batches are sent to LMs for checking.
	The MM second marks a sample of scripts from each marker, including the LMs and the examiner. The MM then prepares a report direct to the director of examinations on the MSM process and the reliability of marking achieved.
	Sample checks on arithmetic additions of marks plus comparisons of inter-marker reliability are made using various statistical methods and the results of exam centres that have been split between markers, etc.

Appendix 2

Semi-Structured Questionnaire for markers

Background

- a) Age b) Employment c) Years in education
d) Experience in professional exam marking

1. Standardisation meetings

- 1.1 Do you feel marking schemes are over or under elaborated?
1.2 To what extent might impression (holistic) marking be used to supplement analytical marking?
1.3 Do you recognize terms such as
'..but would this person be a danger to the Institute?' Or..
'..would you employ this person?'
1.4 To what extent do marking teams reference the core aims of the syllabus?
1.5 How are problems generally resolved?
- democratic/ expert/ hierarchical
1.6 Are there any particular issues with overseas scripts?
1.7 Can you give examples of where consensus has been reached?
1.8 Can you give examples of where consensus has been difficult to achieve?
1.9 To what extent might the validity and reliability be compromised if markers' meeting were not held?

2 Actual marking

- 2.1 To what extent does the marking scheme or the agreed sample scripts set the standard?
2.2 Do you feel that markers have sufficient discretion to react to approaches emerging as marking proceeds?
2.3 To what extent do you feel that markers are influenced by the preceding clutch of scripts when marking. Say, treating a poor script more harshly after a run of good scripts?

3 Feedback

- 3.1 Feedback into further diets - marker performance
- assessment strategy

4 Participant perspective

- 4.1 Do you gain much from hearing the views of others
- in terms of marking the immediate diet
- in terms of developing you understanding of technical matters and the assessment regime in general?
4.2 Do you think the markers' meeting achieves its objectives?
4.3 What is it about the process that leads to a successful outcome?
4.4 Can you think of any ways in which the process could be improved?

Thank you