



This is the post-print of Shay, S. 2005. The assessment of complex tasks: a double reading.  
*Studies in Higher Education*. 30(6): 663-679. DOI: 10.1080/03075070500339988.

It is made available according to the terms of agreement between the author and the journal, and  
in accordance with UCT's open access policy available:  
<http://www.openuct.uct.ac.za/sites/default/files/UCTOpenAccessPolicy.pdf>, for the purposes of  
research, teaching and private study.

# The assessment of complex tasks: a double reading

Suellen Shay

*University of Cape Town, South Africa*

## Abstract

Drawing on Bourdieu's theory of social practice, the author challenges common-sense notions of objectivity and subjectivity which inform assessment practice, and argues for assessment as a socially situated interpretive act. A case study of an engineering community of practice at a South African university illustrates the multiple subjectivities that shape assessors' interpretations of student performance. This case study contributes to an understanding of academic professional judgment as a 'double reading' – an iterative movement between different modes of knowledge which comprise the objective and the subjective. The author concludes with a brief discussion of the theoretical and practical implications of this for how academic communities of practice come to judge and how these judgments are validated.

Of all the oppositions that artificially divide social science, the most fundamental, and the most ruinous, is the one that is set up between subjectivism and objectivism. The very fact that this division constantly reappears in virtually the same form would suffice to indicate that the modes of knowledge which it distinguishes are equally indispensable to a science of the social world. (Bourdieu, 1990, p. 25)

## Introduction

It has become increasingly common practice in higher education to require students to submit a project or thesis as the culmination of their undergraduate programme. These are sometimes referred to as capstone or senior projects. In this article I refer to them as final-year projects. These projects are instances of complex performance (Linn *et al.*, 1991; Moss, 1992). They allow students substantial latitude in their interpretation of and response to the task, often requiring a wide range of conceptual, theoretical and practical skills, and the collection, analysis and synthesis of data from multiple sources. The final product usually takes the form of a written research report, but students may also be required to make an oral and/or graphic presentation.

These final-year projects serve a range of assessment purposes. In addition to their formative role as a final integrative assessment opportunity, they also serve a summative role in that the successful completion of a final-year project is typically a requirement for graduation. An additional purpose that is gaining momentum in South Africa is the use of final-year projects for evaluative purposes, that is, as a measure of the exit-level quality of graduates for purposes of state, institutional and professional accountability. For example, professional accreditation bodies may look at final-year projects for evidence that graduates have achieved the exit-level outcomes.

Complex tasks such as final-year projects present a variety of assessment challenges. In particular the problem of consistency between markers (or intermarker reliability) has received a great deal of attention in the assessment literature (Barritt *et al.*, 1986; Huot, 1990a, 1990b; Broad, 2000; Johnston, 2004), and increasingly so as complex forms of assessment are promoted for evaluative purposes (Moss, 1992; Koretz, 1998; Broadfoot, 1999; Moss & Schultz, 2001). In response to this problem there is a growing body of literature on ways to strengthen the reliability of assessment which includes the advocacy of strategies such as, clear specification of assessment outcomes and criteria, assessor training and rigorous internal and external moderation practices (Nightingale *et al.*, 1996; Biggs, 1999; Luckett & Sutherland, 2000). These are considered standard features of good assessment practice. Given the features of complex performance (as noted above), there are however some validity scholars who question the privilege accorded to traditional validity criteria such as reliability and generalizability (Linn *et al.*, 1991; Moss, 1992; Knight 2002a, 2002b). Moss (1996) goes so far as to question whether reliability is a necessary criterion for validity. Underlying these debates is a more substantive argument about the assumptions which underlie our approaches to assessment, in particular how we validate our interpretations of student performance. Psychometric notions of validity have evolved out of mainstream conceptions of social science. Consistent with this tradition, the goal of validity inquiry is 'not to explain any single isolated event, behaviour, or item response ... But rather the intent is to

account for *consistency* in behaviours or item responses ...' (Messick, 1989, p.14; emphasis in original). Thus the goal of validity from a psychometric perspective is the construction of generalizations across individuals, settings and tasks, that is, context-independent meanings.

Moss (1996) contrasts the goals of science which underlie psychometrics with those of the interpretive tradition. The goal of interpretation within this tradition is 'not to uncover universals or laws' but to 'explicate context' (Rabinow & Sullivan, 1987, p. 14). The goal of interpretation is to understand the actions of individuals in context, 'to understand what humans mean and intend by what they say and do, and to locate those understandings within the historical, cultural, institutional and immediate context that shape them' (Moss *et al.*, 2002, p.7). From this perspective interpretations are contextually constituted and cannot be divorced from the value-bases which interpreters bring with them. Differences between markers are not 'error', but rather the inescapable outcome of the multiplicity of perspectives that assessors bring with them. In contrasting these two scientific traditions, Moss (1996) is not advocating one perspective over the other. She argues for a dialectic between 'the contextualized understanding of local meanings and the distancing analysis of regularities' (p. 22). It is such a dialectic perspective that 'can provide a more textured and productive view of the social phenomena we seek to understand' (p. 22).

There is a growing body of research which approaches assessment from this interpretive and more socio-cultural perspective (see Gipps, 1999, for an overview). Broadfoot (2002) highlights the tension between 'the scientific aspirations of assessment technologies to represent an objective reality and the unavoidable subjectivities injected by the human focus of these technologies' (p. 157). While there is an increasing acknowledgment of assessment as a social practice, few studies have explored this practice within academic communities. Some notable examples include the earlier work of composition theorists (Barritt *et al.*, 1986; Phelps, 1989; Huot, 1990a; Schwegler, 1991) and more recently Broad (2000, 2003), and in South Africa Starfield (2001) and Reed *et al.* (2003). These studies all point to interpretive frameworks that assessors bring to bear in their judgments of student performance – frameworks that are constituted in part as a result of membership within particular communities of practice (Wenger, 1998).

My central argument is that the assessment of complex tasks is a socially situated interpretive act (Shay, 2004). This article draws on a case study of the assessment of final-year projects in an engineering department to explore and illustrate this interpretive act. The particular focus of the exploration is how the positionality of assessors constitutes their interpretations of student performance. The argument is developed in four parts. In part one, I present a case of an academic community struggling to achieve consensus. This case raises questions about why markers agree and why they disagree, and what these agreements and disagreements reveal about the nature of academics' judgements of student performance. In part two, drawing on Bourdieu, I construct a theoretical and methodological approach which illumines assessment as a social practice. Bourdieu's concepts of field and habitus provide generative 'thinking tools' for exploring the phenomenon of consistency and inconsistency in academic professional judgement. In part three, I 'drill down' into these inconsistencies to explore assessors' interpretive processes, in particular how the supervisor's involvement with the student and the project influences their interpretations of student performance. Finally, I turn to consider the theoretical and practical implications of assessment as a socially situated interpretive act for academic communities' validation practices.

### **A problem in practice**

Consistent with many undergraduate engineering programmes around the world (Jawitz *et al.*, 2002), students in the Engineering Faculty at a South African university are required to submit a project as the culmination of their undergraduate programme. In contrast to the assessment of postgraduate theses, supervisors of undergraduate projects mark their own students. In the interest of strengthening the validity of its assessment, one of the engineering departments introduced a system of double marking. In keeping with university policy, marks are awarded out of 100%. The percentage allocations fall into four possible class awards: 75% and above is a first-class award; 60 to 74% is a second-class award (divided into lower-second, 60–69% and upper-second, 70–74%); 50 to 59% is a third-class award; and below 50% is a fail. The department requires the markers to record grades and comments on an official marking memorandum. The memorandum consists of a

list of headings (e.g., problem identification, theoretical background, design), against which markers allocate points out of 100.

The supervisor assesses the project first. It is then independently assessed by another member of the department. Given the variety of disciplinary specializations within the department, there is an attempt to allocate a second marker who has related expertise on the topic. In addition to assessing the thesis, the second marker is expected to attend the student’s oral presentation. If there is a significant discrepancy between the supervisor and second marker’s score, the project is passed on to a third internal independent marker. Finally, all projects are assessed by a team of external examiners – academics from other institutions – who are responsible for awarding a final recommended mark. All the assessors, both internal and external, come together at a departmental examination meeting, where these recommended marks are discussed with particular attention to the projects that have gone to three markers. At this meeting final awards are made.

An analysis of the 1999 and 2000 final-year projects marks reveals relatively high levels of reliability between the marks awarded by the first and second markers, with correlation coefficients of .71 and .72 respectively. However, a different kind of analysis highlights the inconsistencies (see Table 1). Table 1 shows the difference between first and second markers by class awards (upper- and lowersecond class have been conflated for the purposes of the analysis). ‘Same class’ refers to the number of projects which were given the same class award by both markers, 38% and 49% in 1999 and 2000 respectively. Thus 62% (in 1999) and 51% (in 2000) of the projects have class differences of one or more. In 2000 one thesis was awarded a first-class pass by one marker and failed by another. It could be argued that these instances of significant difference between the marks are ‘outliers’, and that the rigorous double-marking system is intended to ensure reliability in the final outcome. While this may be the case, the head of department’s reflections on the 1999 departmental meeting suggested that these differences were not, in fact, insignificant. They mattered to individuals and to this community of practice.

Table 1. Final-year project differences between first and second markers (1999–2000)

Year	Same class	One class	Two classes	Three classes	Four classes
1999 T =102	38% (39)	44% (45)	17% (17)	1% (1)	0
2000 T=84	49% (41)	39% (33)	7% (6)	2% (2)	1% (1)

The head of department, the one responsible for implementing the double-marking system, noted that while the system had run fairly well for a number of years, in the 1999 departmental meeting they had run into some significant problems. In this particular meeting a heated debate ensued between two of the internal examiners and the external examiners, the latter objecting that the marks for some of the projects were unjustifiably high. One of the internal examiners in the end conceded to the external examiner’s recommended marks, but the other did not. The meeting was unable to come to consensus, and a decision was made to send the contentious projects to another external examiner. While this procedural solution eventually resulted in the award of final marks, the meeting had been a very unsettling experience for both the individuals involved and the department as a whole.

Thus while the correlational analysis points to acceptable levels of consistency between the internal markers (Nitko, 2001), the failure of this community to come to consensus signals that there are differences – differences which matter to individuals and to this academic community. This case prompted questions for further exploration: why do markers agree? Why do they disagree? What can these disagreements reveal about the nature of academics’ judgments of student performance? Drawing on Bourdieu, I now construct a theoretical and methodological approach which illumines assessment as a social practice.

**The assessment of complex tasks: a socially situated interpretive act**

Bourdieu’s theory of social practice offers a dialectic perspective, a ‘textured and productive view’ of assessment such as Moss (1996) calls for. To apply Bourdieu’s ‘logic of practice’ (1990) to an exploration of assessment as a social practice suggests, firstly, that academic assessment practices only make sense within the logic of the field, that is, within the logic of the academic profession and its disciplinary sub-fields. For it is the

institutional and professional field that determines the epistemic ‘principles of vision and division’ (Bourdieu, 1996, p. 265) which shape and inform the legitimacy of assessment-based interpretations: what is perceived to be legitimate knowledge, the legitimate criteria for assessing this knowledge and the legitimate assessors of this knowledge.

Secondly, Bourdieu’s concept of habitus, and specifically *class* of habituses (Bourdieu, 1998), suggests that by definition communities of practice to a large extent share a common set of ‘principles of vision and division’ (Bourdieu 1996, p. 265), a common classificatory framework. This means that, as members of the same field, academics share a practical sense, a ‘feel for the game’ (Bourdieu, cited in Bourdieu & Wacquant, 1992, p. 128). With very few explicit codes, rules or criteria, they know how to get on with the myriad of classificatory tasks required of them as members of the field. In reference to this practical sense Goodwin (1994) uses the term ‘professional vision’ to encapsulate the ‘socially organized ways of seeing and understanding ... that are answerable to the distinctive interests of a particular social group’ (p. 606). Members of a community, to the extent that they have been subjected to the same social conditions imposed by the field, share a relatively common ‘perceptual framework’ (p. 616). Thus the ability of academics to interpret student performance as meaningful is not a ‘transparent, psychological process but instead a socially situated activity’ (p. 606). This vision is ‘lodged not in the individual mind but instead within a community of competent practitioners’ (p. 606). Out of this collective, ongoing practice of producing and re-producing these interpretations emerges structure, that is, the conditions or constraints which become internalized again into new acts of construction. This professional vision, or what Taylor (1987) refers to as intersubjectivity, is not synonymous with consensus; it is merely a condition for consensus. Only those with common points of reference can hold one another accountable for, or can contest, the legitimacy of these reference points.

Thirdly, while Bourdieu’s concept of habitus is useful for explaining common frames of reference, his notion of ‘positions’ is crucial for understanding difference and contestation within the field. The reality is that members of the same community can and do disagree (as the case material illustrates); *different* interpretations are as much a feature of communities of practice as consensus. For Bourdieu, individual positions (and thus interpretations) within the field are defined by differential access to resources, or what Bourdieu refers to as ‘capital’. One of the primary forms of capital in academe is knowledge; academic positions are determined to a large extent on the basis of that specialist knowledge. Thus, while two members of an academic community might share a common professional vision of what constitutes research in their field, different positions within that disciplinary field, constituted by for example a commitment to a particular epistemological approach, might result in different interpretations.

Drawing on Bourdieu’s theory of social practice, on the one hand, and the methods of critical discourse analysis and ethnography on the other, I construct a theoretical and methodological approach for illuminating the ‘logic’ of assessment as a social practice. Such an approach requires the analysis of four constitutive elements of social practice. Firstly, it entails the analysis of social structure (or what Bourdieu, 1990, refers to as ‘field’), that is, the particular macro-social conditions of constraint and possibility constituted by the immediate, institutional and societal contexts in which the assessment practices are located. Secondly, social practice is constituted by more micro-level networks of practices within which it is located. This acknowledges that assessment is not an isolated act or event, but a practice located within a network of practices, for example, teaching and research. Thirdly, social practice is constituted by a particular instance of practice, event or occasion. I was interested specifically in the discourse of events where colleagues are required to justify the basis of their assessments to each other; for example, the departmental examination meeting. This discourse, particularly when there are disputes, provides a rare window into the multiple factors which influence academics’ interpretive acts. Fourthly, social practice is constituted by text; this entails the analysis of actual discourse and the ways in which the very choices of language made by speakers (consciously and unconsciously) serve to constitute and are constituted themselves by the practice.

What this meant in terms of my specific methodological approach is a ‘drilling down’ into differences of interpretation – using the disputes which arose during the departmental examination meeting as points of access into the tacit processes of assessment. I began by exploring with members of these academic

communities their own theories for the consistencies and inconsistencies which arise in the interpretation of student performance. Out of a total of 20 academic staff in the department involved in the assessment of final year projects, I interviewed 15 of them once, 11 of them twice, and two of them three times. Out of their explanations emerged common themes which were followed up and fed back into a further round of interviews. The interview findings were then triangulated with a discourse analysis of the moderation events. The discourse analysis focused on the whole departmental examination meeting text, with particular attention to the disputes. 'Dispute' refers to a situation where one of the assessors contests the external examiner's final recommended mark. Out of the 16 disputed cases in the meeting, six of them concerned disputes over the award of first-class passes – projects where the supervisor awarded a first-class pass, the second marker awarded a lower mark and the external examiner supported the second marker. The dispute analysed below is one such example.

Critical discourse analysis is concerned with the relational and ideational functions which discourse serves (Halliday & Hasan, 1985). The former refers to how discourse serves to preserve or challenge interpersonal relations between participants, in this case collegiality. The latter refers to how discourse serves to sustain or challenge systems of belief and knowledge held by members of the group. This article focuses only on the ideational work of the departmental examination meeting discourse, by exploring the different kinds of evidence that assessors draw on in support of their mark. The relational functions of this discourse, for example the way authority is exerted without threatening collegiality are analysed elsewhere (Shay, 2003, 2004) and supported by other studies (Gumpez & Gumpez, 1996; Starfield, 2001).

To re-cap, a socially situated interpretive approach to assessment argues that assessors' interpretive frameworks are constituted, in part, by the objective conditions of the field and of the community of practice. These are objective because they are to a large extent independent of the individual assessor; they are conditions which apply as a result of being a member of the field and sub-disciplinary fields. At the same time these interpretations are constituted by the particular context of the assessment event. This is a highly subjective terrain; that is, it is significantly dependent on the assessor. From this perspective assessment is situationally contingent, rooted in local cultures and reliable and robust only in terms of sets of assumptions, attitudes and values which are, in part at least, localized (Trowler & Cooper, 2002).

### **Illuminating assessors' interpretive frameworks**

As noted above, the dispute which arose in the engineering departmental examination meeting served as a useful springboard to explore with staff in interviews their own explanations for why differences emerge between markers. One of the explanations which emerged was markers' differing levels of experience. New members of staff need time to become socialized into a community's particular 'ways of seeing', to acquire a 'feel for the game' in Bourdieu's terms (Bourdieu, cited in Wacquant & Bourdieu, 1992, p. 18). Another explanation was disciplinary specializations. Departments are organizational units characterized by a wide variety of sub-specializations (Becher & Trowler, 2001). These disciplinary specializations tackle research problems from different methodological and epistemological standpoints. The third explanation for differences between markers which emerged from the interviews is the assessor's involvement with the thesis, the student and their research process. This last feature of assessors' interpretive frameworks is the focus of this article.

Assessors' interpretive frameworks are highly complex, and the particular explanations which emerged from my study were strongly influenced by the context of the data, in particular the difficulties surrounding the 1999 departmental examination meeting. One of the members of staff who was at the centre of the controversy in 1999 was relatively new to the department. The other member of staff who contested the external examiners' marks was from a specialization in the department which employed methodological approaches that were considered unconventional by other members of the department. Thus, I make no claim that these explanations offer a comprehensive picture of assessors' interpretive frameworks, nor am I suggesting that they are necessarily generalizable to other contexts. This would need to be tested in further study.

The most commonly cited explanation for inconsistencies in marks was the different positions that assessors occupy in relation to the topic, the research project and the students. The interviewees were particularly

conscious of their position as supervisor; as one noted, the supervisor is 'tied up in the work'. For example, the research topic may be an extension of the supervisor's own research, as is the case when students join existing projects. The supervisor thus has some idea of the objectives, scope, possibilities and limitations of the research (often based on the previous attempts of students on similar projects), which contribute to a particular set of expectations. There are also particular expectations about the students. One interviewee noted, 'I know for the student [with] whom I was familiar ... I know when I came to read his thesis, I already had a picture in my mind [of] what I thought he was capable of before I read that thesis'.

One interviewee commented on the supervisor's position with this insight:

You read the acknowledgments. On virtually the first page of the document the writer acknowledges the contribution of the first marker. Because that's the supervisor, immediately the supervisor is tied up in the work, so the first marker can't stand right back and say: 'I am objectively assessing this report into which I put a whole lot of effort for the past three months'. You have got to be coming at it from a different direction ... It would be foolish to think that the first and second marker can both be independent. The first marker is deeply involved in the project, every time.

Does this 'involvement' impact on the supervisor's assessment of the project? 'It should affect the marking because I, as the first marker, will give a final mark which depends on what I saw of the student, how he performed. I know how many of the ideas were mine and how many, if any, were his, for example'. Another marker noted, 'The supervisor will have spent two to three months agonizing with the student over various problems, and during that time you get to understand how well the student deals with technical issues, how well they formulate them, how well they look at solving them, how quick they are at actually implementing the solutions and so on. And that definitely has a bearing [on the assessment]'. What emerges is that the supervisor is assessing not only a product, but a person, a research process and to some extent themselves.

Interviewees also problematized the independent and objective positions of the second marker, perspectives which were often referred to in interviews as 'cold' or 'blind'. As noted earlier, the rationale underlying the existing assessment system was to achieve greater consistency by introducing multiple independent assessments of performance. However, many interviewees questioned the extent to which second markers' perspectives were independent, and the extent to which this independence was a good thing. One interviewee, David, noted that for him part of the psychology of being a second marker was to 'avoid conflict'. He explained that if he knew that a particular first marker had a reputation of being generous, he would assign his own marks accordingly: 'In general I wouldn't want to pick up a fight, I have enough work to do'. In contrast, another interviewee, Stephen, described a reverse adjustment: 'Some of my colleagues are known to be generous in their marks, and certainly if I get a thick one to mark from one of them, I think of pulling it down and vice versa'. David concludes, 'Basically what I am trying to get at [is] it's not as independent as – it's not like I go and look at this thesis and I mark it independently of who the student is or who the first marker is and what topic it is'. Stephen concludes, 'I think there is a whole complex thing that goes on'.

Thus, the interviews offer rare accounts of academics' reflections on the subjectivities of their position as supervisors. This includes candid acknowledgments of 'having a picture in my mind ... before I read the thesis', of being 'deeply involved in the project every time', of adjustments made based on 'stuff outside of the thesis'. These reflections include a critique of the other independent internal markers. Some question whether these judgments can ever be independent of student and supervisor identities and the topic. These interpretations too appear to be subjectively constituted. Indeed there is a 'whole complex thing that goes on', leading some members of this community to question the validity of the interpretations which arise from the double-marking process. I turn now to explore these multiple subjectivities in the context of the departmental examination meeting.

As noted above the departmental examination meeting is a gathering of both the internal and external examiners, the purpose of which is to resolve discrepancies between the internal markers. Prior to the meeting the external examiners – who have access to all the internal examiners' marks – have gone through all the theses, paying particular attention to those theses where the discrepancy between internal markers is high. They come to the examination meeting with a recommended mark. The analysis focuses on one of these

disputes in which the supervisor awarded a high first-class pass and the external examiners recommended a lower mark.

The analysis of the dispute (see Table 2) focuses on five units of text which correspond to each of the speakers' inputs in the dispute: 5–7, 8–16, 17–21, 22–25, 26–32. The chair begins the discussion (1–2). In the first text unit (5–7) the supervisor, who has awarded the student a high firstclass pass (88%), seems to justify his high mark on the basis of the interdisciplinary nature of the task and the fact that the student 'seemed to get around and sort out information from a number of difference sources' (6). (The analysis of the supervisor's justification is limited due to poor quality of the audio-recording of this particular text unit.)

Table 2. Excerpt from Engineering departmental examination meeting

Code	
(X)	To protect anonymity of speakers, all of the names used are pseudonyms and references to specialist discourse have been omitted and replaced with an X.
(*)	unclear word
K	Kenneth
D	Donald
R	Robert
A	Andre
N	Nicholas
K	Mr V [the student], the supervisor gave it 88, second marker gave it 72, third marker gave it 69, and the consensus mark was 75, and external examiners recommend a 70.
D	(**)
K	You wished to make a remark?
D	It was about the (X), it was vague on the (*). It's a multi-disciplinary subject and he seemed to get around and sort out information from a number of different sources but you're recommending 70 (**)
R	Could I say that I gave it 72, and it's actually very useful to speak to the student and to see how the student approached the problem. I was convinced it should be a first after discussing the project with the student and hearing the student's story of how he went through this project. It was entirely self-initiated, had almost no support from his supervisor whatsoever, and from the report, for instance, Prof. R. gave it 69, and I think if I came into it cold, and I read the report I would have given it a 69 or 70, but I was really very impressed with the fact that the student had worked entirely on his own, and I was convinced to raise the mark from 72 to 75, and I think that that it's important to take that into account.
A	This brings out an important aspect is how much the candidate is able to communicate what he's doing to the outside world and how important that is. We've had in past many excellent guys, but as part of an engineer you must be able to communicate to the outside world. That you can't do this, I think this must weigh against that, but I think both needs to be taken into consideration.
K	I think looking at the write up itself, the thesis, is not as good as it should be, the references are not and the bibliography is not properly documented, and on that account I gave it (*) but taking everything into account, yes, he did quite a lot of work,
N	I think there are a few reasons not to recommend a first. One of them is that some of the parts of the project didn't work. There was an (X) that was supposed to operate every six hours and he couldn't get that to work, a fairly simple (X) process of (X). In addition he couldn't (X), and really for a distinction candidate you'd like to have these things anticipated in advance and then in the project report itself, really there wasn't great multi-disciplinary detail, for example, (X). There was no additional investigation as to the process. So I think the 70% is a reasonable mark.
K	70?
X	70.

In the second text unit (8–16), the second marker, Robert, makes a strong case for the award of a first-class pass, though his mark is not as high as the supervisor's. He had initially given the thesis a 72%, a mark which is quite close to that of the external examiners. But in the consensus meeting – a gathering of first, second and third markers prior to the examination meeting – he agreed to raise his mark to a 75%. The evidence which he offers in support of the higher award is that 'it [the report] is entirely self-initiated, had almost no support from



the supervisor whatsoever' (11–12), 'the student had worked entirely on his own' (14–15). This evidence appears to have been gathered from a discussion with this student where he heard 'the student's story of how he went through this project' (10–11). Robert further supports his award of the higher mark by contrasting his position/reading (as one who has interacted with the student) with the 'cold' position/reading of his colleague. He argues, 'It's actually very useful to speak to the student and to see how the student approached the problem' (8–9). He notes that had he read the report 'cold' (13), he might have also given it a 69 or 70 as the third marker had. But having talked to the student and seen what the student had actually done, 'I was convinced to raise the mark from 72 to 75 ... and I think that it's important to take that into account' (16).

The third text unit (17–21) provides further insight on yet another position. Andre was not an assessor of this project and it is not likely that he had even seen the thesis. His input was to 'bring out an important aspect' (17) about engineers being 'able to communicate to the outside world' (18, 20). The inference is that independent work does not constitute sufficient evidence to justify a first if the student cannot 'communicate to the outside world' (18). He qualifies, however, by stating, 'but I think both need to be taken into consideration' (21). Presumably 'both' refers to the student's ability to work independently, as well as his ability to communicate.

The fourth text unit (22–25) is that of the third marker, who has given the thesis a 69. The evidence he offers in support of his award is based on the 'write-up itself, the thesis', which he argues 'is not as good as it should be' (22). He points to the 'references' and notes that the 'bibliography is not properly documented' (23). But again he concedes in closing, 'taking everything into account, yes, he did quite a lot of work' (24).

Finally, the fifth text unit (26–32) is that of the external examiner who has recommended a mark of 70. His evidence in support of 'not recommend[ing] a first' (26) appears to be primarily based on the point that 'some of the parts of the project didn't work' (27). From his perspective a 'distinction candidate' (29) should have been able to anticipate certain aspects of the functioning of his project, and this student did not. Thus, we see from the analysis of these text units that different interpretations of the student's performance are being justified on the basis of different kinds of evidence, offered from the perspective of the different positions which the assessors hold in relation to the performance. Evidence presented by the supervisor and the second marker supports a first-class pass. This is evidence from the nature of the task, the student's independent tackling of the task, the student's 'story' of what he did, as well as a demonstration of the actual equipment. Evidence presented by the third marker and the external examiner supports a lower- and upper-second respectively. This is evidence on the basis of the report itself, the manner in which the references were documented, and that aspects of the project did not 'work'.

An analysis of all the disputes which occurred during the examination meeting presented a consistent pattern – that different assessor positions powerfully influence assessors' interpretations. What the analysis exposes are the multiple and varied subjectivities which shape assessors' interpretations of complex performances. But where does this leave us? Reminiscent of Bernstein's Cartesian anxiety, it may appear that we are left with the haunting spectre of relativism – no universally fixed standards, no clear-cut criteria to which we can appeal, no escape from subjectivity (Bernstein, 1983). A retreat to relativism is, however, not an option. There is too much at stake for students, the academic community and society more broadly. Bourdieu's logic of practice (as well as Bernstein's work) calls for a move beyond this Cartesian anxiety to explore alternative explanations.

### **Implications for academic validation practices: theoretical and practical**

This article has argued that the assessment of complex tasks is a socially situated interpretive act; that these interpretations are, at least in part, constituted by assessors' positions. There are many dimensions to assessor positionality, each of which may signal different priorities for what constitutes ideal performances. The focus of this analysis has been on one dimension of positionality, that is, the assessors' relation to and investment in the learner and the performance. Other dimensions of assessor positionality and their respective notions of 'ideal' performance point to generative directions for future study. In this final section, I explore both the theoretical and practical implications of assessment as socially situated practice for how academic communities of practice come to judge and how these judgments are validated.

A central theme of Bourdieu's project is to challenge the opposition between objectivism and subjectivism – an opposition which is deeply rooted in common-sense understandings of assessment. As illustrated in the engineering case, validation systems rely on consistency between independent judgments and privilege (at least officially) the perspective of the external examiner who is appointed guardian of the 'standards'. It is this common sense that Bourdieu challenges. For Bourdieu there is no privileged objective perspective. There is no mere personal subjective perspective. The logic of social practice lies in its double constitution; all judgment is both objectively and subjectively constituted. What light does this shed on our understanding of academic professional judgment in the context of the assessment of complex tasks?

The case material offers a close-up view of academics' interpretive processes and the ways in which they as a community negotiate the tensions between multiple modes of knowledge. This exposes patterns of rationality which underlie academics' professional judgment: a rationality which I argue is neither objectivist nor relativist, to use Bernstein's terminology (1983). Such a rationality might usefully be characterized as a 'double reading' (Wacquant, cited in Bourdieu & Wacquant, 1992, p. 7), as an iterative movement between two modes of knowledge which comprise the objective and the subjective. One is the objective reading, that is, the attempt to 'grasp from the outside' (Wacquant, cited in Bourdieu & Wacquant, 1992, p. 8), to observe, measure and map reality independently of the representations of those who live in it (both the assessed and the assessor). This is an important task; student performances do exhibit regularities and these matter to the public. These regularities in turn powerfully shape the classificatory schemes which are re-applied in the interpretation of this objective reality. This is the manner in which the field sustains itself. As one of the interviewees argued, 'The university is a self-sustaining institution. The belief that the university matters as an important place ... is constructed by the university. It constructs a lot of things which allow it to continue its existence'. These objective regularities also explain how, without reference to explicit criteria, standards or benchmarks, academics are able to get on with the myriad of classificatory acts required of them as members of their profession. Their intuitive judgments are internalizations of the objective regularities of the field they inhabit. Thus academic professional judgment entails an objectivist reading.

The other mode of knowledge which comprises this rationality is the subjective reading. This is the reading which is deeply invested with the self, the reading which acknowledges professional judgment as inescapably (in part) an embodiment of the assessor. This is an insight which assessors are acutely conscious of, illustrated by supervisors' testimonies of how deeply implicated they are in their students' performances. This subjective reading is also relational; it is a communicative exchange between the assessor and the assessed. It thus, inevitably, has to acknowledge and contend with power, that is, the force of relations (Bourdieu, in Bourdieu & Wacquant, 1992, p. 229). This subjective reading is situational, for not only are judgments invested with the assessor, but they are invested with the particularities of the assessment event. This subjective reading is also profoundly pragmatic since, in their judgments, academics are cognizant of the multiple functions which their assessments are playing. They are also extremely sensitive to the consequences of their assessments both for themselves as professionals as well as for the students. Therefore, far from being *mere* personal opinion or an *arbitrary* 'taste' or 'gut-feel', this subjective reading is a socially constituted, practical mastery. This is an aspect of rationality which has been greatly undervalued (Flyvbjerg, 2001), and yet it is essential to an understanding of assessment as an interpretive process and to professional judgment more generally.

Thus, the case study contributes to an understanding of professional judgment as a double reading, as an iterative movement. As Bourdieu argues, 'It is this *double truth*, objective and subjective, which constitutes the whole truth of the social world' (Bourdieu, cited in Bourdieu & Wacquant, 1992, p. 255; emphasis in original). These insights into the nature of academic professional judgment and its underlying forms of rationality are not simply theoretical issues; they have implications for our practical tasks as academic communities of interpreters. I turn briefly to consider some of these.

This article began with the description of a practical assessment problem – a community of practice struggling to reach consensus. This event temporarily shook this community's confidence in its validation system. In the aftermath of this contentious examination meeting, a number of proposals were made for how to review the system. Some called for clearer marking criteria, others called for review of the role of external examiners. The

final recommendation was the constitution of a committee of internal assessors responsible for a final decision in the event of irresolvable disagreement between internal and external examiners. The question remains, if such a committee was called upon to meet, how would its discourse strengthen the validity of the final decisions which were made?

If assessment as a socially situated interpretive act is accepted, then one of the implications is that validation, that is, the ongoing process of evaluating the soundness of our interpretations, is a community process. In contrast to 'objectivism', Bernstein (1976) argues that objectivity is a communal process of striving for 'intersubjective standards of rationality or norms of inquiry by which we attempt to distinguish personal bias, superstition, or false beliefs from objective claims' (p. 111). In the absence of a 'single univocal set of criteria', the crucial issue, he argues, is 'the existence of a community of inquirers who are able, willing and committed to engage in the argumentation' (p. 111).

This highlights what is perhaps one of the great failings in our academic communities of practice, in which the typical technologies of our assessment and moderation systems – marking memorandum, double-marking, external examiners – privilege reliability. These technologies are not in themselves problematic. The problem is our failing to use these technologies as opportunities for dialogue about what we really value as assessors, individually and as communities of practice. In a critique of marking memorandum (or what he calls 'rubrics'), Broad (2000) argues that such lists of criteria support the mistaken notion that assessment consists of identifying characteristics in the text and assigning them relative values. Ultimately, he argues, these rubrics are misleading for they are completely silent about the value system which underlies the interpretive acts. Broad (2003) suggests that a commitment to exploring 'what we really value' requires the pursuit of several related questions: how do we discover what we really value? How do we negotiate differences and shifts in what we value? How do we represent what we have agreed to value? And what difference do our answers to these questions make? Implied in these questions is a model of rationality which is *communal* in character (Bernstein, 1983), where there is 'choice, deliberation, interpretation, judicious weighing and application of "universal criteria" and even rational disagreement about which criteria are relevant and most important' (p. 172).

What about our students? I recall one interviewee's confession that in relation to students' perceptions of their assessments he often found himself in collusion with the myth of objectivity. To an extent, this collusion is something in which all educational assessors are trapped. And yet Bourdieu argues that it is a 'double truth, objective and subjective, which constitutes the whole truth of the social world' (Bourdieu, cited in Bourdieu & Wacquant, 1992, p. 7; emphasis in original). The question is, how do we expose our students to this 'double truth', to assessment as 'double reading'? Assessment as 'double reading' suggests that our interpretations of student performance are both objectively and subjectively grounded. These interpretations are grounded in the objective structures of society which value and privilege particular kinds of performance, and we are responsible for making these expectations as explicit as we can to our students. And yet we do them a disservice if we pretend that these 'standards' have an independent fixed status. Challenging as it may be, the spin-off is that by exposing students to and modelling for them the contextually complex, communal character of professional judgment, we prepare them for the kinds of rational thinking which their future professional contexts will require of them – decision-making which is relational, situational, pragmatic and value-based. This might be one of the most crucial life-skills we have to offer our students.

## References

- Barritt, L., Stock, P. & Clark, F. (1986) Researching practice: evaluating assessment essays, *College, Composition and Communication*, 37(3), 315–327.
- Becher, T., & Trowler, P. (2001) *Academic Tribes and Territories* (2nd edn) (Buckingham, The Society for Research into Higher Education & Open University Press).
- Bernstein, R. (1976) *The restructuring of social and political theory* (New York, Harcourt Brace Jovanovich).
- Bernstein, R. (1983) *Beyond objectivism and relativism: science, hermeneutics, and praxis* (Philadelphia, PA, University of Pennsylvania).
- Biggs, J. (1999) *Teaching for quality learning at university* (Buckingham, Open University Press).
- Bourdieu, P. (1990) *The logic of practice* (Stanford, CA, Stanford University Press).
- Bourdieu, P. (1996) *The state nobility: elite schools in the field of power* (Cambridge, Polity Press).

- Bourdieu, P. (1998) *Practical reason: on theory of action* (Stanford, CT, Stanford University Press).
- Bourdieu, P., & Wacquant, L. (1992) *An invitation to reflexive sociology* (Cambridge, Polity Press).
- Broad, B. (2000) Pulling your hair out: crises of standardization in communal writing assessment, *Research in the Teaching of English*, 35, 213–260.
- Broad, B. (2003) *What we really value: beyond rubrics in teaching and assessing writing* (Logan, UT, Utah State University Press).
- Broadfoot, P. (1999) Empowerment or performativity? English assessment policy in the late twentieth century, paper presented as part of the Assessment Reform Group Symposium on Assessment Policy presented at the *British Educational Research Association Annual Conference*, University of Sussex at Brighton.
- Broadfoot, (2002) Editorial. Dynamic versus arbitrary standards: recognising the human factor in assessment, *Assessment in Education* 9(2), 157–159.
- Flyvbjerg, B. (2001) *Making social science matter: why social science inquiry fails and how it can succeed again* (Cambridge, Cambridge University Press).
- Gipps, C. (1999) Socio-cultural aspects of assessment, *Review of Research in Education*, 24, 355–392.
- Goodwin, C. (1994) Professional vision, *American Anthropologist*, 96(3), 606–633.
- Gumpe, J. C. & Gumpe, J. J. (1996) Treacherous words: gender and power in academic assessment, *Folia Linguistica*, 30(3–4), 167–188.
- Halliday, M. A. K. & Hasan, R. (1985) *Language, context and texts: aspects of language in a socialsemiotic perspective* (Geelong, Deakin University Press).
- Huot, B. (1990a) The literature of direct writing assessment: major concerns and prevailing trends, *Review of Educational Research*, 60(2), 237–263.
- Huot, B. (1990b) Reliability, validity, and holistic scoring: what we know and what we need to know, *College, Composition and Communication*, 41(2), 201–213.
- Jawitz, J., Moore, R. & Shay, S. (2002) Management and assessment of final year projects in engineering, *The International Journal of Engineering Education*, 18(4), 472–478.
- Johnston, B. (2004) Summative assessment of portfolios: an examination of different approaches to agreement over outcomes, *Studies in Higher Education*, 29(3), 393–412.
- Knight, P. T. (2002a) The Achilles' heel of quality: the assessment of student learning, *Quality in Higher Education* 8(1), 107–115.
- Knight, P. T. (2002b) Summative assessment in higher education: practices in disarray, *Studies in Higher Education* 27(3), 275–286.
- Koretz, D. (1998) Large-scale portfolio assessments in the US: evidence pertaining to the quality of measurement, *Assessment in Education*, 5(3), 309–334.
- Linn, R. L., Baker, E. L. & Dunbar, S. (1991) Complex, performance-based assessment: expectations and validation criteria, *Educational Researcher*, 20(8), 15–21.
- Luckett, K. & Sutherland, L. (2000) *in improving teaching and learning in higher education: a handbook for Southern Africa* (Johannesburg, Witwatersrand University Press).
- Messick, S. (1989) Validity, in: R. L. Linn (Ed.) *Educational Measurement* (3rd edn) (New York, Macmillan), 13–103.
- Moss, P. A. (1992) Shifting conceptions of validity in educational measurement: implications for performance assessment, *Review of Educational Research*, 62(2), 229–258.
- Moss, P. A. (1996) Enlarging the dialogue in educational measurement: voices from interpretive research traditions, *Educational Researcher*, 25(1), 20–28.
- Moss, P. A. & Schutz, A. (2001) Educational standards, assessment, and the search for consensus, *American Educational Research Journal*, 38(1), 37–70.
- Moss, P. A., Pullin, D., Gee, J. & Haertel, E. (2002) The idea of testing: expanding the foundations of educational assessment, draft paper.
- Nightingale, P., Te Wiata, I., Toohey, S., Ryan, G., Hughers, C. & Magin, D. (1996) *Assessing learning in universities* (Kensington, Committee for the Advancement of University Teaching/Professional Development Centre, University of New South Wales).
- Nitko, A. (2001) *Educational assessment of students* (3rd edn) (Upper Saddle River, NJ, Merrill Prentice Hall).
- Phelps, L. (1989) Images of student writing: a deep structure of teacher response, in: C. Anson (Ed.) *Writing and response: theory, practice, research* (Urbana, IL, National Council for the Teacher's of English), 37–67.
- Rabinow, P. & Sullivan, W. (1987) The interpretive turn: a second look, in: P. Rabinow & W. Sullivan (Eds) *Interpretive social science: a second look* (Berkeley, CA, University of California Press), 3–30.
- Reed, Y., Grandville, S., Janks, H., Makao, P., Stein, P. & van Zyl, S. with Samuel, M. (2003) (Un) reliable assessment: a case study, *Perspectives in Education*, 21(1), 15–28.
- Schwegler, R. A. (1991) The politics of reading student papers, in: R. Bullock & J. Trimbur (Eds) *The politics of writing instruction: post-secondary* (Portsmouth, NH, Heinemann), 203–225.
- Shay, S. B. (2003) The assessment of final year projects: a study in academic professional judgment, unpublished PhD Dissertation, University of Cape Town, South Africa.
- Shay, S. B. (2004) The assessment of complex performance: a socially-situated interpretive act, *Harvard Educational Review*, 74(3), 307–329.

- Starfield, S. (2001) 'I'll go with the group': Rethinking 'discourse community' in EAP, in: J. Flowerdew & M. Peacock (Eds) *Research perspectives on English for Academic Purposes* (Cambridge, Cambridge University Press), 132–147.
- Taylor, C. (1987) Interpretation and the sciences of man, in: P. Rabinow & W. Sullivan (Eds) *Interpretive social science: a second look* (Berkeley, CA, University of California Press), 33–81.
- Trowler, P. & Cooper, A. (2002) Teaching and learning regimes: implicit theories and recurrent practices in the enhancement of teaching and learning through educational development programmes, *Higher Education Research and Development*, 21(3), 221–240.
- Wenger, E. (1998) *Communities of practice: learning, meaning and identity* (Cambridge, Cambridge University Press).