



Queensland University of Technology
Brisbane Australia

This is the author's version of a work that was submitted/accepted for publication in the following source:

[Adie, Lenore E.](#), [Klenowski, Valentina](#), & Wyatt-Smith, Claire (2011) Towards an understanding of teacher judgement in the context of social moderation. *Educational Review*.

This file was downloaded from: <http://eprints.qut.edu.au/47209/>

© Copyright 2011 Taylor & Francis

This is a preprint of an article submitted for consideration in the [Educational Review] © [2011] [copyright Taylor & Francis]; [Educational Review] is available online at: www.tandfonline.com

Notice: *Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:*

<http://dx.doi.org/10.1080/00131911.2011.598919>

Towards an understanding of teacher judgement in the context of social moderation

LENORE ADIE, VALENTINA KLENOWSKI and CLAIRE WYATT-SMITH

Lenore Adie is a lecturer in assessment and education in the School of Learning & Professional Studies, Queensland University of Technology, Victoria Park Road, Kelvin Grove, Queensland 4059, Australia; e-mail: lenore.adie@qut.edu.au. Her research interests include inclusive assessment practices, social moderation and online moderation processes.

Valentina Klenowski is a professor of education in the School of Learning & Professional Studies, Queensland University of Technology. Her research interests include standards, teacher judgement, social moderation, culture-fair assessment, digital portfolios, assessment and learning.

Claire Wyatt-Smith is a professor of education and Dean of the Faculty of Education at Griffith University. Her research interests include teacher judgement, standards and evaluative frameworks as these apply in policy and practice.

Towards an understanding of teacher judgement in the context of social moderation

Abstract

Social moderation involves teachers gathering together to discuss their judgements of the quality of student work and to reach agreement regarding the standard awarded. This qualitative study conducted over a three-year period investigated the social practice of moderation and the influence on teachers' judgements of students' work. An initial survey of teachers' understandings of moderation and standards, pre- and post-interviews of teachers who participated in the moderation meetings, observations of these meetings with a particular focus on one teacher (focus teachers) comprised the data collection methods. Data analysis involved organising, matching, coding, identifying patterns and themes using a constant comparative method. Sociocultural theories of learning and assessment underpinned the approach to data analysis and proved helpful in explaining the diverse influences on teachers' judgements beyond the task criteria, and the progressive development of shared understandings through engaging in professional discussions of students' work. The study revealed that the process is not clear and linear and is influenced by factors such as the representation of the standards and the knowledge base of the teachers.

Key words: social moderation, assessment, student evaluation, standards based assessment, teacher judgements

Introduction

This paper presents findings from a large-scale Australian Research Council Linkage project that studied how standards inform and regulate teacher judgement of student work in the middle years of schooling. The project investigated this key issue with regard to different curriculum domains, the configural properties of teacher judgements and how these are shared between teachers, the place of social moderation in developing consistency of teacher judgement and resulting in changed judgements.

The key research question addressed in this paper focuses on whether the social practice of moderation involving the application of defined standards results in changed judgements about students' work. The processes and configurations of social moderation are described in this paper. Insights gained from the study revealed the extent to which teachers' involvement in social moderation positioned them to develop a shared understanding of achievement standards. The findings were viewed through a sociocultural theoretical approach to learning to help explain the development of teachers' assessment and judgement-making skills through their participation in moderation practice.

Social moderation is considered to be a quality assurance process that supports the development of a common understanding of standards (Harlen 1994; Matters 2006; Maxwell 2006). This is related to increased reliability of grading decisions as judgements are compared across teachers and sites. While there is a substantial research base in the area of interrater reliability, such studies have focussed in the main on trained examiners (Baird, Greatorex and Bell 2004; Johnson, Penny and Gordon 2001; Suto and Greatorex 2008). One of these studies found through a quantitative study of examiner's marking ability that neither the provision of exemplars nor the coordination of meetings improved marker reliability. However, the authors did note that, "standardization meetings may be particularly important for new examiners" (Baird, Greatorex and Bell 2004, 345).

A different context of assessment and judgement making from these previous studies pertains to this study. In this context of Queensland, Australia classroom teachers were responsible for the judgements and were not trained examiners. These teachers were working in a new assessment context and moderating a common assessment task that they had implemented and marked for their own classes. These teachers met to determine how their judgements align with those of their colleagues. There exists limited empirical evidence in support of claims that this form of moderation meeting, organised to assist classroom teachers

develop consistency in their judgements of student work, will achieve such a shared understanding of a standard. The purpose of this study was to contribute to the gap in the empirical evidence in this particular context.

Social moderation and the development of an assessment community

Our understanding of the social moderation practice is discussed in this section to clarify the assumptions on which the paper is based within the sociocultural framework informing this work. The social moderation process is understood in relation to notions of developing professional communities and promoting teachers' learning of assessment and judgement practices. While other studies have explored the notion of teachers understanding themselves as an assessor through exploration of their tacit beliefs about teaching and learning (Yung 2010), we propose that social moderation is another way that teachers may be involved in understanding themselves as an assessor.

The purpose of social moderation is to produce valid and reliable judgements that are consistent with one another and with stated achievement standards (James and Conner 1992; Wilson 2004). In this paper, standards are understood as fuzzy (Sadler 1986) and so lack precise points of demarcation. Assessments that involve teachers in making subjective decisions of the quality of a response are usually considered to have a low reliability. To reach consistency in judgement formation involves assessors developing a common understanding of the standards on which the assessment is based (Sadler 1986) as well as "similar recognition of performances that demonstrate those standards" (Maxwell 2001, 6).

Social moderation is espoused as one practice that can increase the reliability of assessments, thus ensuring the consistency of standards across assessors and across schools (Maughan 2009). It is considered that involvement in social moderation practices can develop teachers' judgement capabilities by "situating them in a community of judgement" (Wilson

2004, 4). While authors such as Sadler (2008, 2) discuss “the person’s brain [as] both the source and the instrument for the appraisal”, in this paper we are considering the negotiation of these ‘appraisals’, which involves interactions between participants. Our understanding is based on the mind as situated ‘in the action’ (Bakhurst 1997; Cobb 1994), that is, in the nexus of these interactions, so that meaning is developed as a shared understanding. The sociocultural view of learning, as adopted in this study differs from constructivist and social constructivist views in the understanding of ‘mind’ and the construction of meaning as being in the interactions of the group rather than occurring in the ‘person’ (Elwood 2006).

A sociocultural perspective views thoughts as constructed in a process of interaction which positions the focus of the investigation on the interactions of the multiple elements involved. For example, prior to a moderation meeting, the teacher has interacted with the assessment task and the standards to develop an understanding of what denotes evidence of a particular standard. This event is connected to other historical practices and interactions in which the teacher has taken part, that has informed a particular interpretation and use of the standards. Once this teacher meets with other teachers in the moderation meeting, this interpretation and use of the standards may be challenged or confirmed. New understandings develop through negotiation of the standards with the other teachers involved in the meeting. Such occurrences are understood by socioculturists as locating the mind in the interactions as they occur (Bredo 1994). When learning is regarded as enculturation into a practice, mind is situated amidst the action (Cobb 1994). This positioning of the mind in the action continues even after the participant is no longer engaged in the action such that there is always a relationship with some other idea, context or participant that evokes an action.

Through a sociocultural lens, the focus of investigation is on the relationship between the social, cultural and historic elements that come together in the negotiation and construction of shared understandings, and shared ways of performing as an assessor in a particular assessment context. The moderation practice involves relations between various elements such as the teachers, the assessment task, the standards and the policy which are understood through the social, cultural and historic contexts in which they exist and in which they have meaning. The development of shared meaning is reliant on participation in processes of negotiation. Bruner (1990, 12) maintains that through participation “meaning is rendered *public* and *shared*”.

Rogoff (1995) further develops the idea of learning through participation by considering three planes of analysis: community, interpersonal and personal. Analysis of the community plane involves understanding the role of the institutional structures. In the context of this study, this is visible in the policy framework, and the associated artefacts developed to support teachers’ judgement making, and how each of these elements determines what it means to act within this particular community of judgement.

Analysis of the interpersonal plane involves “the processes and systems of involvement between people as they communicate and coordinate efforts while participating in culturally valued activity” (Rogoff 1995, 142). The social moderation meeting is considered to be a socioculturally structured collective activity that involves teachers in a coordinated effort to develop shared meaning. Participation in such meetings however, requires a de-privatisation of teachers’ practice and a willingness to share and negotiate understandings about quality, evidence, standards and their relationship to judgement formation.

Finally, analysis of the personal plane involves the process by which individuals come to understand a practice which then influences their future actions in a practice. Teachers' participation in the moderation meeting may affect future participation as well as their future judgement making. The discussion in this paper focuses on an analysis of the interpersonal plane as this is evidenced in the social moderation meeting while remaining conscious that the other planes are always connected to the discussion.

The context

The article is based on a research project that investigated standards-driven reform in assessment in the middle years of schooling. The project was developed in response to the Australian Government's focus on the use of standards for accountability and reporting in schools, and the resultant requirement on teachers to provide standards-referenced evidence of student achievement. The aim of the project was to provide new knowledge about how standards inform and possibly regulate teacher judgement of student work in the middle years, and the place of social moderation as a means of assisting teachers to develop common understandings of the standards.

The middle school teachers involved in this project were from the state of Queensland, Australia, and were working in a new assessment context. The introduction of a curriculum, assessment and reporting framework into Queensland schools provided an opportune time to study the social practice of moderation within a standards-referenced assessment system. The Queensland Curriculum, Assessment and Reporting (QCAR) Framework focuses on the alignment of curriculum, assessment and reporting practices (Department of Education and Training 2008a). The aim of the framework is to improve the quality of student learning through developing teachers' assessment practices. This is

supported through the provision of an online Assessment Bank of resources, and the Queensland Comparable Assessment Tasks (QCATs).

The assessment tasks

The QCATs are state-wide assessment tasks that appear to be quite unique to Queensland in their design, purpose and scope of implementation (Darling-Hammond 2010; Stobart 2008; Klenowski 2011). One of the purposes of introducing these tasks was to exemplify for teachers how assessments that measure higher order thinking processes as well as content knowledge are designed and assessed. Another purpose of the QCATs was to support teachers to develop skills in making and moderating judgements aligned with stated achievement standards (Queensland Studies Authority 2010). The QCATs were intended to be undertaken annually by students in Years 4, 6 and 9 in the disciplines of English, Mathematics and Science.

Initially, these tasks were introduced to a selected group of Queensland middle school teachers through a trial program that commenced in 2006 and concluded in 2008. After this time it was anticipated that all teachers in Queensland state schools in Years 4, 6 and 9 would implement this mode of assessment. To support the implementation of the QCATs and the judgement making process, teachers were provided with a Teacher's Guide, annotated student work samples, and a Guide to Making Judgments (the criteria sheet) henceforth called the Guide.

Social moderation using the common assessment tasks

It was suggested to teachers through the Teacher's Guide associated with the QCATs that involvement in social moderation of the tasks would enhance consistency of judgements and support the development of a shared understanding of the standards. The Queensland

Studies Authority (QSA) was given responsibility for the implementation of the QCATs and organised a series of moderation training days introducing middle school teachers to the different ways that moderation could be conducted.

Modes of moderation available through the research project

Schools in the study had the option of participating in the moderation meetings either as face-to-face meetings or online. Face-to-face meetings occurred within school or at local cluster or regional levels. The school level moderation sessions predominately involved small groups of two to six teachers coming together to discuss samples of graded assessment tasks in classrooms, offices or meeting rooms. Most meetings were conducted at the end of the school day or during non-teaching periods in the secondary schools. A limited number of schools provided release time for teachers for moderation sessions during the school day as well as for teachers to be involved in local cluster or regional meetings. These meetings ranged in size from large groups of teachers (for example, nineteen groups of four teachers from fifteen schools) from one region, to local cluster meetings of five teachers from small regional schools.

Online moderation meetings were also trialled using the WebEx[®] meeting centre. To take part in such a meeting, participants needed access to a computer with internet capabilities and a phone located next to the computer. A hands-free/speaker phone function was desirable as this freed up participants to type, annotate and refer to other hard copies of documents if necessary. Participants were invited to attend an online meeting via an email. This provided a link to join the meeting, as well as toll free numbers to phone if difficulties were experienced in connecting to the meeting. The assessment tasks and the Guide were uploaded onto the meeting site.

Different procedures for conducting moderation

Schools were presented with three different procedures for conducting moderation: the conferencing model, the calibration model, and the expert model (Queensland Studies Authority 2007). The explanation of each of these models, as provided by QSA through the QCAT Teacher guidelines, is presented in figure 1. Each of these models of moderation offers opportunities for teachers to develop as ‘experts’ in a system of standards-based assessment. If we view these models through a sociocultural lens, each offers a different pathway to development as an assessor and judge who shares an understanding of a standard with others. In the expert model, the ‘apprentice’ is guided in this practice by the ‘master’ who provides feedback. Through reflective engagement with this feedback and perhaps dialogue with the ‘master’ the apprentice moves closer into the practice. The conferencing and calibration models as described by the QSA offer opportunities for development of shared understanding through negotiated practice. In the calibration model the teachers’ discussion focuses on a limited number of common marked samples of student work. The purpose is to establish a shared understanding of the qualities that denote different standards before the marking of a class set of tasks is commenced. This process renders the apprentice as expert and implies a shared confidence that comparable marking between teachers will occur. The conferencing model does not entail such assumptions as the negotiated practice occurs after judgements have been made by teachers on class sets of work. The outcome is that once a shared understanding of the standards is reached through the moderation of selected samples of student work, teachers may need to revisit and revise the judgements that they have made on other samples.

The Queensland Studies Authority, charged with developing the assessment tasks, introduced the conferencing model and the calibration model to clusters of teachers through full day moderation meetings organised around the state in 2007. Although many schools

used these different processes for conducting moderation, many others used combinations of these models. Further support for development of skills in the moderation process was provided by Education Queensland, the government department responsible for state education, through its website and distributed materials (for example, a DVD on moderation practices). This material promoted the conferencing model of moderation (see <http://education.qld.gov.au/qcar/social-mod.html>). On this webpage, the practice of social moderation is defined, as is the role of a facilitator and protocols for moderation.

Configurations of moderation

Moderation meetings were organised at different systemic levels. Local meetings occurred either formally or informally amongst teachers located within a school. This most often occurred with teachers from the same year level. At times, moderation activities were organised by the administrative team that involved teachers from different year levels or across subject disciplines. Although not a common practice, those teachers who moderated across year levels and disciplines believed that this led to a greater understanding of the standards and a greater consistency of interpretation across their school. Observations of intra-school moderation meetings usually involved teachers bringing all of their class samples to a meeting, and then selecting some to share as the conversations unfolded. Quite often, teachers came with borderline decisions that they wanted to moderate.

Cluster moderation or interschool moderation involved a greater degree of preparation for the organisers than intra-school moderation. Teachers were required to select five graded student work samples representing A to E standards, though sometimes only a selection from within this range was presented. At these meetings, teachers were organised into groups that considered each sample. In some cases teachers were able to speak to their own samples; or

Figure 1: Three approaches for consistency of teacher judgement

Calibration model

A facilitator selects samples deemed to be of a certain standard to be used in the calibration process. Teachers individually grade the samples and then compare their judgement with the grade nominated for the sample. Task-specific descriptors are used as the basis for common and explicit language for teachers to use in their discussions about the quality of student performance. These discussions are based on evidence provided in student responses.

Through this professional dialogue, teachers aim to adjust their interpretation and application of the *Standards* to reach consensus about the quality of the sample.

This process is repeated for all the student samples. Teachers then individually grade all student responses, applying the shared understanding achieved through this calibration process.

Advantage — Saves time because it focuses on establishing a common understanding of the *Standards* in context, before marking all of the student responses.

Disadvantage — Making the initial quality judgements in isolation can be difficult.

Conferencing model

Teachers grade student responses individually and then select student samples representative of their application or understanding of the A to E qualities. A meeting is convened in which a conferencing process is employed to enable teachers to share samples and discuss their judgements.

Task-specific descriptors are used as the basis for a common and explicit language for teachers to use during discussions about the quality of student performance. These discussions are based on the evidence provided in student responses.

Through professional dialogue, teachers aim to reach consensus on the interpretation and application of the *Standards*. Teachers review judgements about their previously graded student responses, applying the shared understanding achieved through this conferencing process.

Advantage — Teachers are involved in professional dialogue with other teachers to reach consensus.

Disadvantage — Establishes a common interpretation and application of the *Standards* after student work has been allocated a grade. Extra time is needed to review and adjust previously graded work.

Expert model

Teachers grade all student responses and then submit selected samples representative of their application or understanding of the A to E qualities to an expert. Advice is provided by the expert confirming whether there is consistency in the way the *Standards* are interpreted and applied, or whether teachers need to adjust their understanding, and why. This advice is used by teachers when reviewing judgements about their previously graded student responses.

Advantage — Imposes a common schoolbased view of the interpretation and application of the *Standards*.

Disadvantage — Teachers are not involved in the rich professional dialogue of reaching consensus with other teachers. This model can be used to reach consistency within a school, but does not best support consistency of teacher judgements across the state.

(Queensland Studies Authority. (2009). Queensland comparable assessment tasks: Year 6 Mathematics: Walk the line. Teachers Guide. Brisbane, Australia: The State of Queensland, p. 16)

the samples were collated and teachers moderated samples other than their own and provided written feedback.

Research design

The data collected in this project consists of survey responses collected at the beginning of the project in 2006, pre- and post-moderation interviews of the middle school teachers who participated in the moderation meetings with a particular focus on one teacher (focus teacher), as well as audio and observation data of the moderation meetings, and the artefacts discussed in these meetings where permission was granted. The focus teachers volunteered to be interviewed about their understandings of the process before and after the moderation meeting.

The conferencing model of moderation was the procedure most often observed during this study. There were two main reasons for this occurrence. The first was the interest of the project team in observing and understanding how shared understandings of standards may be supported through participation in social moderation meetings. The second was that the conferencing model of moderation was the preferred procedure from the State employment body, Education Queensland.

Table 1 is a summary of the corpus of data that was collected in Semester 1 and 2, 2007 and Semester 2, 2008 when moderation meetings of the QCATs were held across the state. Some of the teachers who participated in the 2007 data collection also participated in the 2008 data collection and so were interviewed twice. Additionally not all teachers had the time available for post moderation interviews. In some instances two or more of the focus teachers were involved in the moderation sessions, and in other cases focus teachers were followed through different levels (school, cluster and regional) of moderation.

Table 1

Summary of Data Collection (2007 & 2008)

					TOTAL
Focus teachers - gender	Female	66	Male	23	89
	(19 secondary level)				
School numbers	Primary	26	Secondary	20	49
	Special Schools	3			
Sector	State	33	Catholic	10	49
			Independent	6	
Interviews	Pre-moderation	90	Post-moderation	74	164
Moderation sessions	Face-to-face meeting	63	Synchronous online meeting	12	75

A sociocultural theoretical perspective provided the lens for the analysis and a constant comparative method was used to identify the emergent categories and themes. First, the research team divided the task of analysis amongst the members so that each data set was analysed as a separate set by one or more of the researchers. After this phase of analysis of the interview data had been completed, and emergent issues were identified and discussed, a cross comparative analysis occurred. The focus during this next phase of analysis was on the negotiation of judgements in the context of social moderation. The data from the observations and the interviews were analysed with particular attention on any resultant changes in teacher judgements of students' work. The authors acknowledge that relying on interviews and observations limits the research design to what can be observed and the information participants are able and willing to discuss.

The following discussion analyses the different interpretations of moderation as these relate to understandings of standards that were observed throughout the duration of the study and the influence of different factors on these interpretations

Different interpretations of moderation

The Queensland middle school teachers involved in this trial came with a wide variety of prior experience with moderation practices. This caused variation in the way they viewed moderation and approached the moderation process. For example:

- In secondary schools, Year 9 teachers may also have been involved in the senior moderation process (Queensland Studies Authority 2011). This process has been widely documented (for example, Maxwell 2004, 2006; Rolph and Jordan 2010).
- In primary schools teachers may also have been involved in the Year 2 diagnostic net¹ moderation which involved Year 2 teachers meeting once yearly to moderate student portfolios of work aligned with phases of development which included key indicators of performance (Queensland Department of Education and Training 1998)
- Some teachers had experience of moderation from teaching in other states of Australia, and
- Teachers who were involved in the New Basics² trial (Queensland Department of Education and Training 2004) had received professional development in assessing and moderating authentic assessment tasks using standards and a continuum form of criteria sheet.

As could be expected, at the beginning of this trial in 2006, the survey [n = 186] revealed a wide variation in teachers' understanding of the term 'moderation'. For example, table 2 shows the range of terms respondents used when asked to describe moderation.

¹ The Year 2 Diagnostic Net was developed by the QSA to support learning and development in literacy and numeracy in the early years of schooling. Specially designed assessment tasks are given to children in Year 2 to help identify those who need additional support. (<http://www.qsa.qld.edu.au/584.html>)

² The New Basics Project aimed to engage students in a futures-oriented curriculum. More information can be accessed on <http://education.qld.gov.au/corporate/newbasics/> .

Table 2

Common Terms Used to Describe Moderation

Common terms used in descriptions of moderation:	
• Consistency of teacher judgement	• Collaboration
• Comparability of judgements	• Collaborative understanding of the standards ascribed to student work
• Confirming a professional judgement	• Confirming school policy, practice and procedure
• Ratifying judgements	• Process of accountability
• Reaching consensus / agreement	• Building of knowledge
• Across school consistency	• Standardise
• Across schools consistency	• Developing similar standards and expectations
• Consistency in grades	• Moderation of student assessment and performance
• Confirming own ideas	• Supportive environment
• Process for ensuring comparability	• Quality assurance process
• Sharing of student work	
• Sharing pedagogy	
• Sharing assessment strategies	

Amongst these descriptions of moderation that were provided in the initial stages of the trial are variations in notions of consistency, sharing aspects of practice, different ways to facilitate collaboration, and the outcomes of the process. As the project progressed, it was evident that teachers were changing their perceptions of the moderation process. In the second and third years of the project, moderation was being considered as a process that started with task design when teachers shared their understanding of the standards and the quality of responses expected through completion of the assessment task.

The importance of professional skills such as clarifying, listening, and negotiating were made apparent as teachers were confronted with other interpretations of assessment and standards. At this point it is worthwhile considering two factors regarding the professional culture in which these moderation practices occurred. First, the moderation process was new for middle school teachers, though these processes had been a part of the senior school in Queensland for almost forty years (Clarke 1987). Second, the moderation practices that were introduced to the middle school teachers differed to the senior school model in that the focus

was on establishing shared understandings of the qualities that provide evidence of a standard, rather than a quality assurance process necessary in high-stakes assessment. The confluence of these factors was observed in the teachers' active yet respectful and professional approach to the moderation meetings, with minimal observances of power-based issues, particularly as the project progressed.

The promotion of a culture of professional respect was further established in a set of protocols provided to teachers and reinforced at the commencement of each meeting, as well as through the appointment of a meeting facilitator. Throughout all of the meetings observed that involved a facilitator, teachers showed appreciation and respect for this role. When agreement of the standard of an assessment response was not reached, and the moderation meeting was conducted without a facilitator, teachers often cited the lack of this appointed role as the contributing factor.

Negotiating judgements in the context of moderation

Involvement in moderation requires teachers to explicate the reasoning underpinning their judgement decisions. We shadowed focus teachers during these moderation meetings and analysed their explanations of how and why they decided on the particular grades for the student samples under discussion. We were interested from a sociocultural perspective in the teachers' participation in the moderating process and how their understanding of the use of standards for reporting achievement was being developed. We were keen to know when and why teachers changed their judgement acknowledging the importance of context, the role of language, speech and cultural tools (Vygotsky 1978) in this situated learning (Lave and Wenger 1991) of teachers' use of achievement standards for the grading of centrally devised tasks.

We understand the importance of studying the individual teacher in the sociocultural environment of the moderation meeting to fully appreciate the contributions from the individual teachers, their colleagues and the materials. Rogoff (1990) expresses participation in cultural activities as a form of appropriation where the participant both contributes to the event and changes as a result of this participation. Thus, shared understandings developed through negotiation with others are appropriated into one's own understandings influenced by other histories of interactions.

To be able to negotiate judgements, teachers first needed to develop shared understanding of the terminology used in the descriptions of standards. In the following extract from a Year 9 Mathematics moderation meeting the teachers are trying to develop a common understanding of the qualifiers 'appropriate' and 'efficient'. Clarifying and substantiating are key activities here.

Teacher 1: So what is the difference between 'appropriate' and 'efficient' [strategies]?...

Teacher 2: ... I took 'efficient' as being what I would normally use in setting out to do the calculation. If the student had gone around about, so instead of multiplying they just added it up multiple times, as some of the kids do, then that wouldn't have been the efficient strategy, but it would be appropriate. They are the sorts of things that if we see them, we will have to discuss later down the track.

Teacher 3: I guess you could say... oh no, you're right... an appropriate strategy would be one that works but is not the most efficient. (Year 9 Maths moderation meeting, 2007)

The importance of participation through questioning and active listening in the social activity of moderation became apparent.

Rogoff's (1995) concept of participatory appropriation was useful in explaining the development of shared understandings of the standards. As articulated by Rogoff (1990, 195-

196), “It is not a matter of bringing to the internal plane a product that was produced externally. It is a matter of social engagement that leaves the individual changed”. The following extract from a post-moderation interview is indicative of teachers’ comments when their judgements were verified by their peers.

I was really surprised because I was thinking, “I’ve given him a C, I really should have given him a D”, but to hear other people talking about the work, and then saying very clearly a C, I thought, “Oh, oh well, okay, I will take that on board” because, obviously, my interpretation of things is different to theirs and I might be a little more strict or tight on my interpretations and that four other people at the table said, “Oh no, it’s a, it’s definitely a C”. (Post-moderation interview, Special School Cluster, 2007)

Rogoff (1995) suggests that individuals change through engagement in an activity and become prepared for involvement in subsequent activities; this is understood as a process of becoming. Through a process of negotiation, teachers learn how another has understood the standards and may then apply this in their future assessment and judgement making decisions. Teachers stated that it was during times of disagreement on judgement decisions when opportunities arose for learning but these instances required teachers to skilfully participate in the act of negotiation.

It wasn’t always unanimous agreement, but what it made us do was if there were perhaps two or three people who thought it was an A and one or two who thought it was a B then we kind of went backwards and forwards to justify why each of us thought that, and in some cases we went one way and in some cases we went the other way...and the most valuable part was when we actually disagreed on it, not when we agreed on it. (Post-moderation interview, Year 6 English, 2007)

Through discussion and negotiation, teachers came to a shared understanding of hitherto perplexing subtleties in the qualities that differentiate the standards. Sadler (1986, 5) states that the objectivity of judgements based on standards is increased when there is “agreement

among competent assessors”. This involvement with ‘competent assessors’ is particularly important for teachers new to the practice of standards-based assessment.

The role of artefacts and changes in judgements

Standards, described as fuzzy and lacking sharp boundaries of differentiation, are open to different interpretations by teachers (Sadler 2009b). Sadler (2009a) describes four necessary elements to support an understanding of the standards: exemplars, explanations of the judgement decisions based on the evident qualities in the work aligned with the standards, conversations between participants, and tacit knowing which is the elusive and often difficult-to-articulate understandings that develop through shared experiences. While Sadler (2008) advocates the use of standards in a holistic grading system in which multiple criteria are considered simultaneously, the system described in this paper could be considered closer to analytic grading in that separate judgements were made on a selected range of criteria which were then combined for an on-balance overall grade. The difference between these grading systems is significant but not the purpose of this paper and will not be dealt with here. However, we consider that the elements to support an understanding of the standards as advanced by Sadler are also necessary to support teachers working with defined standards within selected criteria.

To support the judgement making process of the QCATs, QSA provided teachers with the Guide (criteria sheet), a Teacher’s set of Guidelines which involved instructions on the preparation, administration, marking and moderation of the tasks, and annotated student work samples for the different A – E standards. These artefacts and the practices promoted within them equate to some degree with the four elements to support an understanding of the standards proposed by Sadler (2009a). While the annotated examples were not exemplars of each standard, they provided teachers with annotations of decision making on authentic

student work for each of the standards. The moderation meetings provided organised opportunities for conversations amongst teachers. Through the use of these artefacts, teachers were able to access a common way of speaking about the assessment tasks and their judgement-making process.

In the early meetings conducted in 2007, teachers relied most heavily on the annotated samples and used them as exemplars of a standard rather than one example of how competence may be demonstrated at that standard. During the moderation meeting the collaborative response of the teachers to the various resources resulted in some teachers becoming aware of the value of using the Guide, drawing on the annotated samples as support, if required, in their decision-making. The following segment from a post-moderation interview in 2007 illustrates how the moderation discussion supported teachers to use the cultural tools of the practice as represented in the Guide.

Interviewer: ...following your moderation session today what insights have you gained about how the standards are used in making judgements that are consistent with other teachers' judgements?

Teacher: Possibly, I used ... the 'Guide to Making Judgements', probably a bit more today and actually got into actually reading them, than before. Before I had used the student sample responses and, um, based it on that and gone, "Oh, yeah, that's an A" and today I probably did a little more reading within the grid [sic] (the criteria sheet/Guide). (Post-moderation interview, 2007)

In contrast to this statement is that by another teacher later in 2008.

We used that [the Guide] as our...that's what we were using to mark so we referred back to that for every question. That was our guide to marking...So that was...the final thing was based on the 'Guide to Making Judgements'. (Post-moderation interview, 2008)

As the trial progressed, teachers, now with some experience of judgement-making using standards came to understand that the annotated samples only provided one possible response and that there were other ways to demonstrate evidence of a standard. With this understanding, teachers predominantly matched the qualities within the student response against the standards-descriptors in the Guide.

The Teacher Guidelines also provided necessary information for teachers to help in deciding a grade by providing a general view of a standard. For example, this teacher described how, in their moderation meeting, the teachers had changed a student's grade from an E to an N (not assessable) based on the general definition for an E standard.

The other minor change was the student who we had thought would end up being an E; we actually ended up deciding on the basis of not simply the criteria and standards but also actually on the basis of the 'Guidelines to Teachers'. Page 4 of the teacher guidelines it says, "The E standard: the evidence in a student's work typically demonstrates rudimentary knowledge and understanding, typically demonstrates superficial application of mathematical processes, typically demonstrates communication using everyday language" and in fact the evidence we had was that the student very rarely demonstrated any of those things, so we decided there was actually not enough evidence to award even an E, so we awarded an N.
(Post-moderation interview, 2007)

The cumulated artefacts acted to support teachers to understand their work in a standards-based assessment system, and connected teachers' learning and participation in the language of the practice as an aspect of becoming in the practice. Vygotsky (1997) highlighted the importance of artefacts and language as inherent in a culture, and as they are developed through social interaction in a cultural system.

When changes were made to judgements they were frequently made within an assessable element (criterion) but the overall grade of the student's work was rarely changed.

At times, changes that were made were based on an easily quantifiable element of a standard descriptor. For example, this teacher describes how students were moved from a C to a B in one assessable element based on the description in the Guide that a C required one response while a B required two responses.

Interviewer: You said you might change a couple [of grades]. Why would that be?

Teacher: Actually when we got into the finer details you had to have answered two to be classified as a B whereas for a C you only had to do one. So some of mine, their explanation may not have been good, but they have actually done two explanations so consequently they are a B. (Post-moderation interview, 2007)

The student had supplied two responses and so could be awarded a B even though the teacher stated that the responses were poorly reasoned. Is this a case of criteria, in trying to be specific, actually restricting the marking by not allowing for other important dimensions of the response? The teachers' group moderation discussion is also included here and illustrates the authority of the Guide in making judgements with some reference to the annotated samples.

Teacher 1: ...if you look at the Guide it says to provide, to be a C you have to write a logical explanation for length of one life cycle where he's actually answered both so he has to be a...

Teacher 2: So it has to be a B because he's answered (b) and (d), so it has to be either A or B, because he's answered both of them.

Teacher 1: I'd say it would be a B because he's not particularly comprehensive...

Teacher 2: yeah

Teacher 1: ...but he has logically, he has logic behind what he said.

[Discussion regarding answer]

Teacher 2: Well that's why we say, because he's answered both, he's got to be either A or B and he's got to be B because he hasn't really put his point across very clearly. (Year 6 Science moderation, 2007)

The teachers are adjusting their reasoning to the Guide but do not appear to be critically analysing the validity of the Guide's standards or descriptions of standards. In this instance, the quantitative measure of the response was considered more important than the quality and validity of the response. However, observations of other moderation meetings showed that this was not the position taken by all teachers. For example, in the moderation meeting between two Year 9 English teachers, one teacher reasons through the standard so that the other also agrees that in fact a lower grade should be awarded as the student did not meet the criteria. The task required the selection and sequencing of five missing paragraphs from a choice of eight into a news article, as well as justification of why one of the remaining three paragraphs was not chosen for insertion into the article. Originally the student had been awarded a B for this section which required the correct insertion of four of the five paragraphs and a clear explanation. The student had correctly inserted four of the paragraphs but had not provided a clear explanation of why a paragraph was not chosen. The result was that the teachers decided to award a D for this section of the task based on the lack of explanation even though the description of a D response included that only "one or two paragraphs are correctly organised".

Teacher 1: Ok he's got the four... you can see where he's got the B, for that I split the criteria, but when it comes to the justification it's a D...So I guess if you average it out it'll be a C. But look at this, it's not...I just don't think, question 2 shows you have to give the reasoning behind your choice and why you didn't choose the others.

Teacher 2: Yep, didn't do it.

Teacher 1: [discussion of answer] I'm inclined to say D, just simply because he just could not justify his answers. (Year 9 English moderation, 2007)

The student's grade is marked down based on the teachers' reasoning (or valuing) that being able to justify answers is a skill needed to be awarded a higher grade. 'Justifying' is mentioned in the standard descriptors for A and B as one quality necessary to be awarded either of these standards. In their discussion, the teachers also considered other elements, some of which were not part of the criteria (for example, spelling) and they imposed a quantitative understanding of the grading process by averaging out the result (B for correct quantity + D for poor justification = C). The teachers started at B, then averaged the grades to a C, then finally through discussion awarded a D to the student based on his inability to justify his responses.

Sadler (2008) refers to the act of combining scores by addition as a linear model. It is a compensatory procedure in that higher scores compensate for lack of achievement in another area however, the resultant grade is misleading as it fails to describe the qualities that have been achieved. Further, if one of the purposes of pre-defined standards is to inform students of "the bases on which judgements are made" (Sadler 2009a, 809) then the combination of quantitative and qualitative descriptors needs elaboration, for example, if success in one element takes precedence over success in the other. The process highlights the difficulty of working with the multiple dimensions of standards descriptors, in particular those that include quantitative as well as qualitative measures, and the teachers' attempts to negotiate a shared interpretation of a standard.

The examples discussed in this section have illustrated how the variables within the supplied artefacts can influence decision making in the moderation meeting as teachers work with a new model of assessment. Other variables influencing a change in grades were related to whether the teacher was present to justify the reasoning of the judgement. In the following

case, the judgement made on this assessment task at the school level was verified at cluster level moderation when the teacher was there to justify the decision making process but was changed at regional level when the teacher was not present. In the moderation processes being used for the QCATs which were not high-stakes, a change in judgement caused teachers to reconsider the decisions on which their original judgements were made. As this teacher stated;

... the one from our school I will certainly go and discuss it with the teacher; and we have year level meetings which we use to moderate again, and just bring it up and see if we can go, “Oh, actually” you know, see if we can reach the consensus again, and if not justify that within our team, because there’s six in our team anyway, so, you know, it gives a reasonable sort of moderation anyway. So that’s probably what we’ll do. (Regional moderation meeting, Post-moderation interview, 2008)

This brings into question the variability of teacher presence. At this stage we are not arguing for or against involving the teachers who marked the work in the moderation discussions of their work. In the example cited, the changed judgement at regional level resulted in the teachers at the school level reconsidering their judgement making process so that the moderation discussions continued at school level. Sociocultural theories of learning enable us to view this process as a continuation of a community of practice (Lave and Wenger 1991) where meaning is continually negotiated as different groups within the practice meet. At the school level, the teachers do not have to accept the changed judgement suggested at the regional level, but it will cause them to revisit their judgement making and consider a different perspective on this understanding of a standard.

In this section we have considered some of the factors that we observed and that teachers stated influenced the judgement making process as they negotiated the meaning of a standard in the social moderation process. In particular, these factors included the artefacts

provided to support the process and the teachers' prior knowledge which influences the reading of the standards as well as the reading of the artefacts.

Conclusion

The question we sought to answer as a part of this research project asked whether the social practice of moderation involving the application of explicitly defined standards resulted in changed judgements about students' work. The findings of our research reveal that changes in judgement making, and thus the understanding of a standard, do occur as a result of the moderation process, however these changes are not always related to "the application of explicitly defined standards", and are often related to a criterion rather than an overall result. Other factors included the teachers' personal standards and how these related to the official or mandated standards, the knowledge base of the teachers, the teachers' prior experience with the moderation process, the representation of the qualities in the standards (quantitative and qualitative), and the interpretation of the qualities of a standard, all of which could combine to influence the judgement making and negotiation processes.

Sociocultural theories of learning help us to understand the variety of historical, social, cultural, environmental and political factors influencing the development of shared meaning. The influence of many factors results in quite a messy practice but one where we observed evidence of progression of learning through negotiated practice as teachers refined their understandings and developed a shared knowledge of the qualities that denote a standard. We anticipate that as teachers continue to engage in negotiation of the standards, that the reliability and validity of the assessments will increase. Such an assumption would need to be quantitatively verified in a study such as that conducted by Baird, Greatorex and Bell (2004) for this specific set of classroom teachers who are also assessors.

We adopted the metaphors of ‘mind in action’ (Bakhurst 1997; Cobb 1994) and ‘community of judgement’ (Wilson 2004) to support a perspective of moderation as progressive learning. This involved teachers developing a shared language including nuances in meaning of assessment terminology. The examples drawn from in this paper have illustrated how this ‘community of judgment’ develops as new ways of assessing and judging student work are introduced into education systems.

The judgement making process is complex as it involves many elements other than those detailed in official documentation. Sadler (2009a) highlighted the importance of conversations to develop shared meanings of standards and ‘tacit knowing’. The practice of social moderation is important as it makes apparent these processes and causes participants to justify their decisions and to rethink their understanding of a standard which may result in changed judgements of students’ work.

Acknowledgements

The authors wish to acknowledge that the project from which these ideas are derived was funded by the Australian Research Council Linkage Program and we wish to further acknowledge the involvement and support provided by our Partners, the Queensland Studies Authority (QSA), the National Council for Curriculum and Assessment of the Republic of Ireland and Queen’s University, Belfast, Northern Ireland. Peta Colbert is also acknowledged for her work in the project, including the initial work with Lenore Adie in establishing the IT-mediated moderation meetings.

References

- Baird, J., Greateorex, J. and Bell, J.F. 2004. *What makes marking reliable? Experiments with UK examinations*. *Assessment in Education* 11, no.3:331-348.
- Bakhurst, D. 1997. Meaning, normativity and the life of the mind. *Language & Communication* 17, no.1: 33-51.
- Bredo, E. 1994. Reconstructing educational psychology: Situated cognition and Deweyian pragmatism. *Educational Psychologist* 29, no.1: 23-35.
- Bruner, J. 1990. *Acts of meaning: Four lectures on mind and culture*. Cambridge, MA: Harvard University Press.
- Clarke, E. 1987. *Assessment in Queensland secondary schools: Two decades of change 1964-1983*. North Quay, Australia: Department of Education, Queensland.
- Cobb, P. 1994. Where is mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher* 23, no.7: 13-20.
- Darling-Hammond, L. 2010. *The flat world and education: How America's commitment to equity will determine our future*. New York: Teachers' College Record.
- Elwood, J. 2006. Gender issues in testing and assessment, in ed. C. Skelton, B. Francis and L. Smulyan, *The Sage handbook of gender and education*. (pp. 262-278). Thousand Oaks, CA: Sage.
- Harlen, W. 1994. Concepts of quality in student assessment. Paper presented at the annual meeting of the American Educational Research Association, April, in New Orleans, LA.
- James, M., and Conner, C. 1992. *Moderation at Key Stage One across four LEAs 1992*. Cambridge: Cambridge Institute of Education.

Johnson, R.L.; Penny, J. and Gordon, B. 2001. Score resolution and the interrater reliability of holistic scores in rating essays. *Written Communication* 2001, no. 18: 229-249.

Klenowski, V. 2011. Assessment for Learning in the accountability era: Queensland Australia, *Studies in Educational Evaluation* 37, 78-83.

Lave, J. and Wenger, E. 1991. *Situated learning: legitimate peripheral participation*. Cambridge: Cambridge University Press.

Matters, G. 2006. Statistical moderation and social moderation around Australia. Paper presented at the International Association for Educational Assessment Conference, May, in Singapore.

Maughan, S. 2009. Improving the acceptability of teacher assessment for accountability purposes - Some proposals within an English system. *Cadmo* XVII, no.2: 39-54.

Maxwell, G. 2001. *Moderation of assessments in vocational education and training*. Brisbane, Australia: Queensland Government Department of Employment and Training.

Maxwell, G. 2004. Progressive assessment for learning and certification: Some lessons from school-based assessment in Queensland. Paper presented at the Third Conference of the Association of Commonwealth Examination and Assessment Boards, Redefining the Roles of Educational Assessment, March, in Nadi, Fiji.

Maxwell, G. 2006. Quality management of school-based assessments: Moderation of teacher judgements. Paper presented at the 32nd International Association for Educational Assessment Conference, May, in Singapore.

Queensland Department of Education and Training. 1998. Year 2 Diagnostic Net: Information booklet for state schools: Moderation.

<http://education.qld.gov.au/students/policy/assessment/y2dn/html/ibss3.html>

Queensland Department of Education and Training. 2004. The New Basics project.

<http://education.qld.gov.au/corporate/newbasics/html/lce.html>

Queensland Department of Education and Training. 2008a. Queensland Curriculum, Assessment and Reporting (QCAR) Framework.

<http://education.qld.gov.au/qcar/index.html>

Queensland Studies Authority. 2007. *September 2007 Trial common assessment task: Teacher guidelines*. Brisbane, Australia: Author.

Queensland Studies Authority. 2010. *School-based assessment: The Queensland system*. Brisbane, Australia: Author.

Queensland Studies Authority. 2011. Moderation and quality assurance.

<http://www.qsa.qld.edu.au/586.html>

Rogoff, B. 1990. *Apprenticeship in Thinking: Development in social context*. New York: Oxford University Press.

Rogoff, B. 1995. Observing sociocultural activity on three planes: Participatory appropriation, guided participation, and apprenticeship. In *Sociocultural studies of mind*, ed. J.V. Wertsch, P. del Rio and A. Alvarez, 139-164. Cambridge: Cambridge University Press.

Rolph, L. and Jordan, P. 2010. Externally moderated school-based assessment in Queensland: Lessons for implementing an Australian curriculum. Paper presented at the IAEA conference, Bangkok, Thailand. www.iaea2010.com/fullpaper/511.pdf

Sadler, D.R. 1986. *Subjectivity, objectivity, and teachers' qualitative judgements*.

Queensland, Australia: Board of Secondary School Studies.

Sadler, D.R. 2008. Indeterminacy in the use of preset criteria for assessment and grading.

Assessment and Evaluation in Higher Education.

<http://dx.doi.org/10.1080/02602930801956059>.

Sadler, D.R. 2009a. Grade integrity and the representation of academic achievement. *Studies in Higher Education* 34, no.7: 807-826.

Sadler, D.R. 2009b. Transforming holistic assessment and grading into a vehicle for complex learning. In *Assessment, Learning and Judgement in Higher Education*, ed. G. Joughin, 45-63. Dordrecht, Netherlands: Springer.

Stobart, G. 2008. *Testing times: The uses and abuses of assessment*. London: Routledge.

Suto, W.M. I. and Greatorex, J. 2008. A quantitative analysis of cognitive strategy usage in the marking of two GCSE examinations. *Assessment in Education: Principles, Policy & Practice* 15, no. 1: 73-89.

Vygotsky, L.S. 1978. *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.

Vygotsky, L.S. 1997. The instrumental method in psychology. In *The collected works of L.S. Vygotsky Vol. 3: Problems of the theory and history of psychology*, ed. R. Rieber and J. Wollock, 85-90. New York: Plenum.

Wilson, M. 2004. Assessment, accountability and the classroom: A community of judgement. In *Towards coherence between classroom assessment and accountability*, ed. M. Wilson, 1-19. Chicago: University of Chicago Press.

Yung, B.H.W. 2010. Examiner, policeman or students' companion: Teachers' perceptions of their role in an assessment reform. *Educational Review* 53, no.3: 251-260.