Using annotations to inform an understanding of achievement standards

Lenore Adie and Jill Willis

Abstract

The implementation of a new national curriculum and standards-referenced assessment in Australia has been an opportunity and a challenge for teacher assessment practices. In this case study of teachers in two Queensland schools, we explore how annotating student or exemplar assessment tasks could support teacher assessment practice. Three learning conversations between the researchers and the teacher teams are interpreted through the lens of Bernstein's (1999) horizontal and vertical discourses to understand the complexities of coming to know an assessment standard. The study contributes to the literature on the use of annotations by exploring how teachers negotiated the purposes and processes of annotation, how annotating student work or exemplars before teaching commenced supported teachers to experience greater clarity about assessment standards and, finally, some of the tensions experienced by the teachers as they considered this practice within the practicalities of their daily work.

Introduction

Expectations about the quality of knowledge and skills that students should have attained at particular junctures in their education are represented by achievement standards. Public and shared standards are designed to promote transparency of curriculum and assessment

practices. They also work alongside the curriculum to inform and support progressive development of deeper knowledge and understanding, and development of students' higher order thinking skills (Centre for Educational Research and Innovation, 2005; Clarke, Madaus, Horn & Ramos, 2000; Zepke et al., 2005). However, an inherent characteristic of written qualitative standards is that they lack sharp boundaries of differentiation which leaves them open to different interpretations among teachers, parents, and the wider community (Sadler, 2010b). In practice, rather than being transparent, standards without exemplification can be quite opaque. While standards offer opportunities for a range of responses to be recognised, this can result in variable judgements from different assessors. Also, teachers may draw on latent as well as explicit criteria when using standards to make judgements (Wyatt-Smith & Klenowski, 2013). These characteristics can hinder the transparency of the standards. The implementation of standards-referenced assessment has presented both opportunities and challenges for teacher learning.

Understanding how teachers develop a shared and common comprehension of achievement standards is important, both for enhancing the reliability and equity of the assessment system, and for using standards to support student learning. In this article, it is argued that involving teachers in collaboratively annotating responses to an assessment task can support shared understandings of achievement standards, and that this practice is most beneficial when it occurs in the planning stage before teaching commences.

Annotation is defined in this article as "the augmentation of text with additional content ... designed to actively engage with the host text ... and employed by author or reader" (Ball, 2010, p. 138). Annotation is explored in terms of the teacher's use of written annotations on an exemplar or on a previous sample of a student's assessment task to explicitly identify the features that are being valued in the assessment response, and how these features align with different parts of the achievement standard. Negotiation of the annotations with other teachers is an opportunity to develop a shared meaning of what the achievement standard looks like in practice.

The discussion in this article draws on a project involving Year 2 and Year 6 teachers from Queensland, Australia, who were using a model of backward mapping (Wiggins & McTighe, 2005) to develop a unit of work based on a new national curriculum (ACARA, 2012). Backward mapping involved the teachers in a planning sequence of identifying the curriculum, developing the assessment that would provide evidence of learning, then

designing the learning experiences that would support student success. Through this process curriculum, assessment, and pedagogy are aligned. The teachers who were involved in this project collaboratively annotated a work sample during this planning process and before they commenced teaching the work to their students. The researchers acted as facilitators, since using annotations to clearly document valued qualities evidencing a standard was a new practice for the teachers.

The study contributes to the literature regarding the use of annotations in three key ways. First, it identifies how teachers negotiated purposes and processes of annotation to assist in understanding the Australian curriculum achievement standards within their everyday assessment practices. Bernstein's (1999) theory of vertical and horizontal discourse is used to explain the complexity of the practice, viewing the Australian Curriculum as a vertical discourse, and the enacted assessment practices as a horizontal discourse. Secondly, the study demonstrates how annotating student work or exemplars before teaching commences can support teachers to experience greater clarity about key aspects that will be taught. Finally, the study identifies some tensions experienced by the teachers as they considered this practice within the practicalities of their daily work.

Background

A wealth of previous research (Harlen, 2005; Klenowski & Adie, 2009; Maxwell, 2002; Wyatt-Smith, 1999) has highlighted the inconsistency of judgements made by teachers that are based on written standards. Providing too much specificity in standards, in terms of detailed criteria, standard descriptors, and guidelines, has also been deemed problematic (O'Donovan, Price, & Rust 2004; Sadler, 2009; Torrance, 2007). Warnings include the production of unwieldy documentation rather than shared understandings, and criteria compliance rather than deep learning. To overcome these issues, many authors (Freebody, 2005; Harlen, 1994; O'Donovan et al., 2004; Wyatt-Smith, 2000) have suggested the use of exemplars, as well as professional conversations that negotiate understandings of the standards. For example, O'Donovan et al. (2004) identify a number of ways of sharing the tacit knowledge of teachers with higher education students through exemplars, annotated student work samples, marking practice for teachers, self-assessment, and peer discussions, as well as peer-marking exercises. Harlen (1994) has been advocating for many years for "the provision of examples of pupil's work which has been assessed, preferably with a

commentary on particular features used in making the judgement" (p. 12). Similarly, Freebody (2005) advocates for the use of standards, exemplars, and annotations so that over time teachers may develop a shared understanding of a standard. Providing opportunities for the collaborative development of shared understandings of achievement standards, through the use of annotations and dialogue, is a suggested way forward for consistency in judgements within standards-referenced assessment systems.

Research in the use of annotations has focussed on teachers annotating students' work for assessment and feedback purposes (e.g., Ball, 2010; Crisp & Johnson, 2007; Heinrich, 2004; McGuire, 2005), and on the value of students annotating their own work (Johansen, 1998) or annotating text to support comprehension (Porter-O'Donnell, 2004; Zywica & Gomez, 2008). While Heinrich (2004) considers online annotation as a way of providing feedback to current students, and as a resource for future students, McGuire's (2005) focus is on using electronic tools for annotations as a means of promoting dialogue between the teacher and student. Ball (2010), on the other hand, examines the literature on handwritten annotation as critique that provides meaningful feedback to students, and identifies strategies for effective practice. The value of students annotating their own work to identify their thinking behind a response was found to improve the quality of university law students' responses and to support their learning (Johansen, 1998). Other researchers (Porter-O'Donnell, 2004; Zywica & Gomez, 2008) have explored how annotations can support students to learn reading and comprehension skills. Annotations are therefore an established means of promoting dialogue about assessment.

This article explores how primary school teachers use annotations at the beginning of the planning process to explicitly identify the qualities they are valuing as evidence of a standard. The study builds on previous work conducted with ten Year 6 teachers on the use of achievement standards to inform planning of the new Australian curriculum (Authors, 2013). Findings from this first phase of the project provided evidence that the teachers' understandings of the achievement standards were influenced by "their understanding of the achievement standard text, their beliefs about learning and assessment, and the assessment culture of the school" (Authors, 2013). One of the aims of Phase 1 of the project was for the teachers to compile a portfolio of annotated evidence representative of a standard, which they would then share at a moderation meeting involving all of the schools. Annotating student work samples was found to be a difficult process for the teachers. The teachers had a history of working with annotated exemplars provided by the education authority. This history

provided the basis for assuming that the skill of annotating work samples may have been a practice that the teachers could easily transfer to their own work. In reality, this transference proved to be a far more difficult task than anticipated. This finding led to the current project investigating the annotation processes of six teachers across two schools.

In an attempt to elucidate the process of "coming to know" an achievement standard through annotating student work samples and exemplars, Bernstein's theory of vertical and horizontal discourse is drawn upon (Bernstein, 1999). Bernstein (1999, p. 161) distinguishes vertical discourse as "specialised symbolic structures of explicit knowledge" that are hierarchically organised, such that knowledge and understandings progressively become more complex. Horizontal discourse is described as "local, segmentally organised, context specific and dependent" where individuals within a community may draw from a common "reservoir of strategies" but perform and understand them in individual ways (Bernstein, 1999, p. 159). Within this framework, Australian curriculum achievement standards are an attempt at establishing a national vertical assessment discourse, by creating hierarchical progress maps that reflect specialised subject discourses. To enact the achievement standards, teachers need to "recontextualise" the vertical discourses of the Australian curriculum by selectively interpreting and relating the standards to their everyday or horizontal assessment practice (Bernstein, 1999). In recontextualising the vertical discourse, the teachers were "coming to know" the meaning of the achievement standards.

The teachers' conversations and shared understandings are shown to demonstrate characteristics of a local or horizontal discourse that involves initially weak grammars strengthened through negotiation with other teachers working within the same curriculum and school context. Bernstein (2000) uses the concept of strong and weak grammars to distinguish the degree of precision capable of being generated through language structures within horizontal discourses. Bernstein (2000, p. 163) states that "The strong grammar visibly announces what it is". Weak grammars, on the other hand, are indicative of uncertainty, with users unsure whether their interpretations are consistent with others. The Australian curriculum achievement standards, presented as fuzzy and indiscrete, are an example of a weak grammar in which the language is made apparent through the processes of negotiation and annotation. These processes allow the knowledge of the standard to be legitimated and taken as "truth" (Bernstein, 1999, p. 164). The knowledge production of "coming to know" a standard involves negotiation and justification through which invisible or tacit understandings that impact on judgement decisions are made visible to others and are documented within

annotations. This documentation identifies the qualities that are being valued as evidence of the standard. The process of annotating student work samples and exemplars with other teachers in the teaching team provides a context where tacit knowledge of a standard is explicated through the specific evidence identified in the work sample.

Research design

This project responds to current national interest in the introduction of an Australian curriculum, and international interest in standards-referenced assessment practices and the use of assessment to support student learning. While effective assessment is an essential building block for a quality curriculum (Zepke et al., 2005), the Australian curriculum leaves unanswered many questions about assessment, especially the alignment between an end-of-year achievement standard and a requirement to report based on A to E standards (or equivalent). The following research questions are addressed in this article:

1. What are the challenges faced by teachers as they annotate student work samples using A to E standards of evidence based on an end-of-year achievement standard, and how are these challenges addressed?
2. How does the process of annotating student work support teachers in developing an understanding of the qualities that evidence A to E standards of work, given an end-of-year achievement standard?

To respond to these questions, the researchers worked alongside six primary school teachers in two schools to collaboratively investigate effective annotation practices using the Australian curriculum achievement standard. Each school was visited several times. In the first visit, the school organisation for planning and assessment and the key school initiatives in this area were discussed with the school principal. At the next visit, the researchers worked alongside the teachers for an initial planning day discussing how planning occurred at the school and in the year level, and the positioning of assessment within these discussions. The researchers acted as facilitators and co-constructors of meaning to (a) share ideas about backward mapping to inform the planning of teaching and learning experiences, and the use of annotations in this planning process, (b) support teachers as they engaged in exploratory conversations with peers about what evidence they valued as representing a standard, and then (c) challenge and prompt teachers' reasoning, such as, "What is the difference between an A-level response and a B-level response?" Articulating how they differentiated between

different standards of performance was difficult for the teachers and required continual questioning, prompting, and facilitation of this process by the researchers.

Data were collected through the observation and audio recording of teachers' conversations as they annotated teacher-made exemplars or samples of student work. The teachers' justification of why the identified section of the student's work was evidence of a particular standard was a focus of the observation. These meetings occurred at the beginning of a teaching semester, and the teachers were then left to implement this unit of work and to trial different ways the annotated work samples could be used to inform their teaching, and student learning. The researchers kept in contact with the teachers through email during the semester to provide follow-up support, and to specifically enquire about the usefulness of the annotated samples to their teaching and student learning. The data set consisted of researcher notes regarding school practices for the development of curriculum and assessment from meetings with the principals and teachers; observation notes and audio recordings of the initial planning days with the teachers; records of emails between the researchers and the teachers during the semester; and copies of annotated work samples provided by the teachers.

The data were analysed to answer the two research questions, first to identify the processes teachers used to negotiate meaning, and then to identify the challenges faced by the teachers as they collectively constructed their annotations. The parts of the conversations that centered on the challenges were analysed as critical incidents (Tripp, 1993) or knotty turning points (Lincoln & Denzin, 2003) that were coded to identify pragmatic and epistemic assumptions and conflicts. These data were then interpreted by the researchers into descriptive case study narratives (Simons, 2009).

The case studies included information regarding the school context (e.g., type of school, year levels, enrolments) and the processes and procedures for planning and assessment that were part of the school culture, as well as initiatives that the schools were working towards. Also included was information regarding the teaching team (e.g., numbers in the team, how long they had worked together, the amount of collaboration that normally occurred). This information was used to understand the conversations between the teaching teams and with the researchers both during the planning day (as the researchers questioned the teachers' decisions and asked them to articulate their reasoning), and during the semester (through the email conversations). The case-study data were interpreted by the researchers against

Bernstein's (1999, 2000) theoretical framework which enabled insights into the relationships between the vertical and horizontal discourses to emerge (Eisenhardt, 2002).

As the study progressed, it was evident that the teachers' thinking about their annotations was a personal and collaborative effort, sharing historic knowledge and reconciling this with the requirements for the new standards-referenced curriculum. To capture this process, three narratives of conversations are detailed here that are indicative of the thinking that occurred as the teachers attempted to articulate their understandings of the valued qualities of a standard. Following the descriptions is an analysis of what was occurring in these conversations interpreted through the lens of Bernsteinian horizontal and vertical discourses.

Conversations with teachers

The first two conversations occurred with two Year 2 teachers from an inner-city state primary school. These teachers were working with new curriculum resources in the form of unit overviews, lesson plans, and associated assessments. This work provoked the teachers to think about new practices, and a need to process meaning together. Their school leaders had introduced to the staff the idea of using assessment exemplars and annotations, with the requirement that, every two months, each teaching team would share their annotated exemplar in a staff meeting. The research team met with the teachers to discuss one of their prepared exemplars. The third conversation occurred with four Year 6 teachers and their curriculum leader, from an independent girls' school. The conversations about annotations occurred over a period of 12 months, and helped the teachers to come to know the potential for annotating to inform their own work, moving beyond performance as systemic compliance.

Conversation One: The dragonfly annotated example

Before our meeting, the Year 2 teachers had collaboratively created an exemplar of a procedural text and written the assessment-task instructions for the students. The task involved students creating a procedural poster for making a model. After writing the task, the teachers had realised that there was a lack of alignment between the national achievement standards and their assessment task requirements. This realisation created a "lot of debate" about the main focus of the task. A consensus was reached that was based on the emphasis in the new curriculum resources, and a decision was made to focus on teaching and assessing

time connectives and command verbs. The Year 2 teachers spoke to the Year 3 teachers to understand the progressive development of the identified skills, and to validate their expectations of the required standard. Finally, to "check it all matched", the teachers cross-checked their task requirements against a checklist provided by the local studies authority that listed "evidence of learning". They added annotations to their exemplar "to show which parts of the Achievement Standard we had covered". The annotations included statements such as "command verb", "time connectives", and "all equipment listed". While one purpose of the annotated exemplar was to show the students a model, it was also to justify to their principal and curriculum supervisors that it matched the required standard. With the introduction of a national curriculum and the associated systemic support materials, the teachers were keen to know if the extra time taken to annotate an exemplar was going to benefit their teaching and student learning, or if this was just another compliance activity.

After the teachers described their process of developing an exemplar, the research team queried how different standards of work would be distinguished. The teachers had not yet considered this. They felt that the annotated sample was an important step in what was a new pedagogic practice of sharing expectations with their students. The resultant discussion and persistent questioning by the researchers enabled the teachers to begin to verbalise their tacit knowledge about the values they would expect in higher and lower quality responses. The teachers drew on their knowledge of their students and typical performances in similar tasks. To inform the A standard, the teachers looked at the achievement standards for the year level above. The teachers recognised that they would expect to see a "variety" of time connectives in an A standard and added new annotations to their exemplar. Through this discussion some of the original annotations were not seen as equally valued, such as "including a final picture".

This collaborative annotating process provided an insight into how the teachers worked across the variety of resources, including their own understandings and existing practices, and tried to align all these information sources. For example, the following extract from the planning day illustrates the process of negotiation as the teachers recontextualised their understanding of the assessment standard into their daily work.

> Researcher: [*Reading one of the teachers' annotations*] "Method is in correct sequence." Is it more [than this], because our other conversation before was about the detail of the method? You wanted the full method for the A standard.

T2: Yes. We wanted where they did the gluing…

T1: It is the main steps for a C [standard]…

T2: Is it [having] the fine detail or parts of it? [for an A standard]
T1: Without them going overboard with the detail. It has to be appropriate… [not] squeeze the [glue] pot and place it back on the table (laughs)

T 2: We've been doing a lot of descriptive writing [in class], haven't we. So I guess if the children can demonstrate interesting sentences within the method rather than just a basic "cut out the template". The children who are an A are able to describe exactly what they are doing, that might be indicating an A student rather the B student.

The teachers' conversation illustrates the messiness of this process of coming to know a standard. The process for the teachers is exploratory, tentative, and individually incomplete as it emerges in the thinking-aloud process. When constructed collaboratively, it becomes a shared understanding.

As a final step, the teachers compared the features that they had valued in their exemplar with the annotated work samples provided as a resource for teachers from the Australian curriculum website. The teachers had not previously accessed these samples and felt that their own expectations were much higher.

T2: We look like we are teaching grade 7 don't we [*laughs*]…

In response, the researchers highlighted an annotation from the Australian curriculum Year 2 exemplars to challenge the teachers to see the detail embedded in these statements.

Researcher: [*reads annotation from Australian curriculum example*] "Writes simple sentences with command verbs, to provide clear instructions and compound sentences to connect ideas"

T1: So they are more focusing on the compound ... We have done compound [sentences] the whole way [year] through. It depends on its purpose.
Researcher: Is it something you have taught them for this purpose? If you haven't explicitly taught it, is it something that you are valuing in this task?

T2: We taught them that in the previous unit. It is not really that necessary in this one.

The teachers rejected the annotation on the exemplar from the national website, as it did not align with their current practices and existing curriculum plans.

This annotating process also revealed that the teachers were able to notice and value the surface features of the writing easily, but struggled to articulate less well-defined qualities. They recognised that they had not necessarily given students the opportunity to demonstrate a

range of abilities. "Thinking about the purpose, it is not just to keep admin happy. It's going to help us be clear about what we are teaching, save us time and help the kids at the same time." They decided that, in the future, their professional discussions about expectations relative to achievement standards needed to occur as part of the planning before teaching the unit, and not merely as a prelude to planning the assessment, which often occurred halfway through the unit of work.

Conversation two: Unpacking a comprehension assessment task

The second conversation with the teachers occurred later that day, and gave the teachers an opportunity to experience backward mapping, by aligning their teaching and learning activities with the provided assessment task. The planned test required students to answer literal and inferential comprehension questions by referencing two texts: one, an information text; the other, a fictional text. The teachers spent some time analysing the test questions and recording their thoughts as annotations on the assessment task sheet. This analysis and annotation involved taking three different but related perspectives of the task: (a) the features that would identify a quality response, (b) the skills and knowledge required of students to demonstrate that answer, and (c) the qualities that would identify different standards of performance. During this process, the teachers questioned some of their own and one another's assumptions. For example, they began to question their practice of using numerical marks in a comprehension task that evaluated students' ability to answer literal and inferential questions.

> Researcher: 4/7 could mean that they get all of the first questions right but don't get the higher order thinking questions right. Does the criteria sheet show you the difference?
> T2: Oh ... no ... They have to think beyond the text.
>
> T1: The student can have a degree of understanding if they can show where they got their answer.
>
> Researcher: See the C statement says 'Locating literal information'. So are you saying that they can show, and also provide key facts? The D then says 'Responds to literal questions'. Is it that the D just locates information?
> T2: So maybe they just write a 1 where they found the information?
> Researcher: C says 'Locate' and D is 'Respond'… not even if the response is correct or not?
>
> T2: So we need to make 'Respond' the D.
>
> T1: Does that mean we are lowering the national standard?

T2: [*reassuring to colleague*] If they can locate the answer, they can extend it. We are getting them to do a bit more in the D, which is locate the correct answer. E is just 'Respond'.

From these experiences, the two teachers began to explore the implications for their practice. As part of their coming to know how annotation might enhance their assessment understandings, each teacher chose a different focus. One teacher had used the process of collaborative annotation to identify the valued features that had previously been hidden or assumed, and planned to teach these skills and understandings explicitly to her students. The other teacher explored the process of annotation itself and how it made the valued features visible for her own judgement making. She found it so valuable, she decided to explore if annotating an example with her students would also make these skills and understandings more visible to her students and hence support their learning. As part of the process of coming to know the assessment standards, each teacher explored a practice that seemed to promise the better outcome for her students, and could build on her current assessment understandings. As a result of these differing experiments, both teachers reported that the quality of the student responses had improved significantly. They believed that spending time reflecting on expected qualities within the task had clarified their own expectations about what to teach.

Conversation Three: Annotating an integrated technology task

The team of four Year 6 teachers in school two met regularly to design common learning plans and assessment. Team planning took more time, but led to greater shared understanding and the consistency that the team desired. This collaborative planning occurred during one day every few months when the teachers were released from classes to work together. It was refined in their weekly common "spare", a weekly 1-hour planning meeting after school, and through emailed conversations. The teaching team had changed each year, with one member remaining the same for 5 years. This was identified as challenging, since "we almost have to go back to square one each time to develop a shared understanding". Being Year 6 teachers in a P-12 school meant that the teachers' assessment practices were heavily influenced by the higher school assessment policy regarding the number of assessment items, percentages, and scheduling of end-of-semester tests.

Planning for assessment involved identifying task requirements by referencing the school curriculum documents and reporting spreadsheet to "plan what we want the kids to work towards". The teachers designed the assessment task, and then determined the criteria and

standards descriptors. The teachers' conversation during our planning day centered around an integrated assessment task requiring mathematical knowledge to design and produce a plan for a product and conduct a survey; and language and technology knowledge to write a script and film a commercial for the product. This task had been implemented in the prior teaching term. When marking the resulting student work, the teachers realised that some criteria were not clearly aligned with the task. They identified their greatest challenge to be designing a rubric that maintained meaning for themselves, their students, and parents.

Using this assessment task the teachers began to think aloud about what annotations might look like and the purposes annotations would serve. Time-effective strategies such as arrows, number coding, track changes, or colour coding were discussed. Discussion about efficient ways to annotate led back to a discussion about purpose: "We wouldn't do it for every child in the class ... would we?" The type of annotation therefore was seen to depend on the purpose and audience for the annotated example. Three examples of differing purposes and approaches were identified in the discussion.

1. If it is to justify to parents, then maybe every child's work would be annotated.
2. If it is to provide models to guide students, only one or two examples are needed.
3. If is it to create a portfolio to guide moderation then one or two examples of each standard would be annotated.

This led to an important discussion about the difference between annotations and feedback to students. While both require teachers to document in writing their judgement-making, they differ in their timing, audience and purpose. Annotations are representations of the embodied qualities of a standard rather than the individualised commentary of personal feedback. Also, feedback occurs typically during or at the end of an assessment performance, while annotation can also inform expectations about the standard before the commencement of teaching.

In attempting to annotate scripts, the teachers first identified what they understood by the A descriptor in their rubric: "writing shows maturity and flair". Using different colours, the teachers moved to highlight the student work. It was found to be more effective to look at one valued feature and one colour at a time. The teachers were conscious of noticing positives and what evidence was in the sample, as initially they had noticed the spelling and grammar mistakes and the qualities that were "missing". As the teachers interpreted the work, they were also interpreting how the student may have engaged in the production of the work,

making comments like "I think she got distracted here" and "she is in a hurry to get on camera in this part". There was agreement that in annotating a C standard, they had to notice what was not evident as well as what was. To annotate one student sample took an hour of conversation, as teachers attempted to articulate and reconcile some of the assumptions that informed their judgement-making. At the end of the conversation the teachers returned to questions of finding time-effective ways of including annotations as part of their current practice, such as pencilling in their conversations at the time of moderating, or using sticky notes, or typing their comments to develop a resource bank of expected qualities.

The teachers were challenged to consider whether a time-efficient practice may be to collaboratively annotate an exemplar before they started teaching. Our suggestion was that some of the difficulties they faced when making judgement decisions could come to the surface earlier, and that greater clarity could inform common approaches to teaching. One teacher reflected: "we are often naive about how explicit we need to be ... we walk away assuming we are all going to do the same thing". Another teacher agreed, commenting, "I was thinking before when we were writing down our thoughts, what a waste of time. As if we don't all do this [make judgements in similar ways]. And then I realised actually we are all doing very different things". Others in the teaching team reflected that their differences in interpretation of standards was a natural product of having big teaching teams, and also reflected the very different ways they each processed information. There was agreement that there was room for greater consistency, but also flexibility, and that talking through and taking the time to write down expectations could enhance their own practice and understanding. The teachers resolved to try to include annotations in their future planning meetings.

After the meeting, the teachers continued to experiment with their annotation process, sharing annotations of an A and C standard of work in a biography assessment. Instead of the colour coding, the teachers tried annotations as commentary in the margins that could be scanned and electronically shared. While the teachers were not yet fully confident of the quality of their annotations, they commented on the usefulness of the practice to support their work as a team.

Discussion

In this project, teachers were introduced to the practice of using annotations as a way to develop an early understanding of the valued qualities in student work that evidence a standard. There were two key steps in this process. The first involved the discussions by researchers with the teachers at the planning stage. The research team focussed on the value of backwards mapping (Wiggins & McTighe, 2005) as a practice that would connect the vertical and horizontal discourses of standards-referenced assessment. The second step was the recontextualisation of this discussion by the teachers into their teaching practice.

Negotiating intersections of vertical and horizontal assessment discourses

Annotating student work samples or exemplars involved teachers in articulating the expected qualities of work. As these discussions were occurring before teaching, these qualities are an anticipated standard informed by both the systemic vertical achievement standards and the local horizontal assessment discourse. Historic practices and knowledge clearly influenced the teachers' understandings of what was valued as evidence and what the expected standard would be, regardless of the systemically stated achievement standard. For example, the understanding of some teachers that an A standard must contain elements of higher year levels rather than higher levels of thinking and problem solving was not a belief generated through the current national standards-referenced assessment policy. Also, teachers in both schools were surprised to see that their expectations were above those of the national standard, but resolved the dilemma by prioritising their own expectations and experiences. The teachers were working within their local horizontal discourses, and drew on their knowledge of past student performances within their community, as well as anticipating how the children in their classes would engage with the assessment activities in the future. While the teachers worked between both the horizontal and vertical discourses, it was evident that succinctly articulating these convergences as annotations was a difficult task.

Adding complexity to this process were the historic assessment and reporting routines of the school that had come to be accepted as uncontested practices. The tensions caused by school policies and processes, designed to provide stability and consistency across a teaching team, were evident in the teachers' discussions. Wenger (1998) refers to the creation of authoritative texts or processes such as these as *reification* that become shortcuts to understand complex practice. For example, the criteria sheets produced by the teachers reified the relevant vertical discourses of the Australian curriculum achievement standards and the A

to E assessment descriptors. However, Wenger notes that reification is double-edged. These abstractions in their succinctness can reduce effort, and amplify good ideas, but they can also become a substitute for deep understanding or be "frozen into a text that does not capture the richness of lived experience and that can be appropriated in misleading ways" (p. 61). To mediate, there needs to be a balance between the richness of participatory discussion, and the continuity of meaning that occurs when that discussion is written down. Wenger (p. 65) identifies the principle of a duality of participation and reification, in which each compensates for the others' shortcomings, and meaning can be negotiated in a coordinated and generative way. In this project, annotations were an additional type of assessment reification, made powerful through professional conversations.

In the negotiations of articulating the standards through the annotations, the teachers were challenged to articulate what may previously have been tacit understandings of quality. While Polanyi (1958) proposed that tacit knowledge cannot become fully explicit, it can be represented and learned through practice. In verbalising these qualities to other colleagues when negotiating the annotations, there was opportunity for teachers to challenge their own and one another's assumptions. Representing this tacit knowledge was a lengthy process for the teachers, and there was a need for supported guidance and strategic questioning.

The achievement standard from the vertical discourse of the Australian Curriculum was a reification of quality observed to have little influence over local assessment practices, since it was not yet part of the horizontal discourse. The school-wide assessment practice, such as a spreadsheet used for reporting purposes, had greater influence. It is proposed that collaborative annotation of examples of student work can help create a shared horizontal discourse of quality, and provide some representations to help teachers navigate the intersections of horizontal and vertical discourses.

Annotation before teaching gave greater clarity

The value of annotating student work during the planning process was evident as the teachers started to clarify what evidence would look like at different standards of performance before planning their teaching. The process of annotating enabled teachers to articulate and determine the valued qualities of a standard, and move beyond a checklist developed from the content descriptor. Finding clarity was challenging, and to deeply probe evidence of a standard was difficult work. Understanding quality involved seeing the work holistically rather than as separate, unrelated criteria. Sadler (2010a, p. 544) states that "quality is often

easier to recognise when it presents itself than it is to define in the abstract". Annotating tasks from previous years provided a starting point for the teachers' conversations where they could consider why a response had been awarded a particular standard.

Working towards developing a shared understanding of the qualities that were valued as evidence of a standard assisted the teachers to avoid the pitfalls of criteria compliance, overspecification of criteria, and quantification of criteria. For example, the teachers challenged prior beliefs that quantified a standard, namely that, in an A-standard response five steps are identified, in a B-standard response four steps are identified, and so on. Instead, they started to develop holistic and shared understandings that involved the connections between different elements. The value of conversations is apparent as the teachers negotiated, questioned, and clarified their reasoning on judgement decisions. In one conversation, the teachers considered what elements were missing to award a higher grade, and the perspective and originality of a response. Through that discussion, the teachers delved deeper into the qualities that they were expecting to see in the students' responses. The teachers' conversations moved between identifying specific features to the connections between the identified features that produce the final product.

In discussing and writing the annotations, the teachers were also making links to their teaching practice, to writing the criteria sheet, to student feedback, and to parent reporting. One experienced teacher commented that where she would have marked a task as 4/10, she could now see the value of providing explicit feedback related to the valued qualities of a task. This teacher revealed that, for the first time in her teaching career, she had really unpacked and clarified the skills she was focusing on in her teaching. Another teacher in the project communicated the value of the annotation discussion and writing in her school as supportive for new year level teachers to "see the end product before we start". The teacher also noted that showing these samples to students had "improved the quality of the work they [the students] complete as they can see that either they could do better, or that they can see where to aim for". A novice teacher indicated that, while many of these strategies had been discussed in her university lectures she had not understood their value until she was involved in the negotiated practice of writing annotations that evidenced a standard.

Tensions

As teachers contemplated how annotating student work might become part of their professional routine, they faced both practical and epistemic difficulties. Practical difficulties

included finding time-effective ways to include annotation in their planning routines. The teachers also wanted clear differentiation between their current practice of giving feedback to students and the new practice of annotating to establish shared understandings of achievement standards. Identifying the purpose of the rubric or criteria sheet in the annotating practice was part of the coordination needed between the vertical discourse of the Australian Curriculum Achievement Standard and the horizontal discourses of local assessment practices. The study demonstrated the importance of considering these tensions when designing systemic support for teachers working in standards-referenced assessment systems, and in developing annotation practices.

With the hectic workload of a teacher new practices need to have perceived value to student learning. It took time for the teachers to reflect deeply and to inquire about the standards, and then to record this thinking as annotations in the backwards-mapping process. Teachers were ready to rush through the backwards mapping by identifying surface features, and found it challenging to check for a shared understanding of complex thinking processes. While the teachers noted the significant time taken to reach this shared understanding of expectations, they also commented that the clarity gained through these conversations enabled them to prioritise their teaching time, and effectively saved time later as it focussed their teaching priorities. The teachers also commented that conducting the annotation conversation during the planning phase had been more beneficial than annotating student work samples in preparation for a moderation meeting at the completion of a unit of work.

For the teachers, epistemic difficulties related to the shift in the timing of planning for assessment and the use of annotations in this process. Instead of assessment being a short-term event occurring at the end of an extended teaching sequence, annotations were extending the reach of assessment tasks to inform the planning before teaching. It meant that planning was moving from being curriculum centred (that is, what will we be doing?) to assessment driven (that is, what qualities are we valuing?) This movement involved thinking about assessment as contributing to learning rather than a point in time response to one semester of teaching and learning activities.

A further epistemic shift occurred in the teachers' identities as assessors during the process of collaborative annotation. Instead of a focus on developing individual assessment skills as teachers, the annotation conversations were shifting the focus to the development of a team's shared assessment and teaching capacity. Annotating assessment judgments also meant that

teachers shifted from being an assessment reader to becoming a judgement writer. Being an assessment reader had involved teachers in tasks such as reading many examples of student work and comparing each example to a criteria sheet before reaching an overall judgement. As a judgement writer, the teachers were required to search for the words that captured holistic understandings, and to commit to writing this commentary on examples of these qualities as they appeared in student work. While this writing involved similar skills to marking or giving student feedback, the audience was significantly different. Annotations are written for an audience of teacher peers, and the quality of a teacher's judgment making is being made visible to their peers. These tensions were complex and involved a range of connected elements, and provide one explanation as to why the process takes time to master, even within a supportive peer environment.

Conclusion

Recording annotations on students' work is not a new process. Since Plato lamented that the written word limited the opportunity of others to question, clarify, or debate the meanings of text, ways to clarify the ambiguity of words have been developed. However, in societies that value visibility and accountability, annotations can be revisited as a way to engage teachers in deep discussions about the qualities they are valuing in student work within a framework of achievement standards.

In a recent review of the literature, Opfer and Pedder (2011) identified that certain preconditions are necessary for new patterns of learning to emerge in teacher practice. Teachers need the opportunity to engage in learning that is recursive and includes the opportunity to challenge their beliefs, and to learn new practices in personal, practical, and formal ways. Some disequilibrium is important to provoke new learning and change, but too much leads to rejection of the change. Annotating conversations provided an opportunity for some disequilibrium, but it was evident in our study that this needed to be a facilitated process that probed yet scaffolded the teachers' thinking. It is proposed that developing a culture of deep discussions of standards will require leadership, and time for sharing and dialogue across year levels.

It was apparent throughout this project that annotation in itself is not necessarily going to lead to improved learning outcomes for students. Before our arrival, the teachers had already annotated their exemplars. While it was helpful as a management tool for their teaching team

to coordinate some common practice, the annotations often focused on surface features of the task. It was only as the teachers were prompted through a process of critical inquiry about the differentiation between achievement standards that they began to develop a more sophisticated understanding of assessment standards and the qualities that would evidence the standard for that particular year level and task.

When teachers experienced the value of engaging in substantive conversations regarding expected standards of performance and articulating this as annotations, they were eager to continue refining their practice. Yet it was evident in the study that annotations will not resolve all issues in developing a shared understanding of standards. By developing and annotating an example of student work, teachers may instead lean towards criteria compliance. Strategic and challenging questioning was required to progress teachers from a compliance focus to consideration of what different levels of performance will look like. This took time, both for the conversation and also to give teachers the opportunity to recontextualise their emerging understandings. The outcome of greater focus in their teaching, explicit teaching of skills, and understanding the expected trajectory of learning, provided the impetus for sustained engagement in the professional conversations required to articulate the valued qualities of the standards and record these as annotations of shared understandings.

References

ACARA. (2012). *The Australian Curriculum*. Retrieved from
http://www.australiancurriculum.edu.au/

Ball, E. C. (2010). Annotation an effective device for student feedback: A critical review of the literature. *Nurse Education in Practice, 10*, 138–143.

Bernstein, B. (1999). Vertical and horizontal discourse: An essay. *British Journal of Sociology of Education, 20*(2), 157–173.

Bernstein, B. (2000). *Pedagogy, symbolic control, and identity: Theory, research, critique*. Maryland, USA: Rowman & Littlefield.

Centre for Educational Research and Innovation. (2005). *Formative assessment: Improving learning in secondary classrooms.* Paris: Organisation for Economic Co-operation and Development.

Clarke, M., Madaus, G. F., Horn, C. L., and Ramos, M. A. (2000). Retrospective on educational testing and assessment in the 20th century. *Journal of Curriculum Studies, 32*(2), 159–181.

Crisp, V., and Johnson, M. (2007). The use of annotations in examination marking: Opening a window into workers' minds. *British Educational Research Journal, 33*(6), 943–961.

Eisenhardt, K. (2002). Building theories from case study research. In A. M. Huberman & M. B. Miles (Eds.), *The qualitative researcher's companion* (pp. 5–35). Thousand Oaks: Sage Publications.

Freebody, P. (2005). *Background, rationale and specification: Queensland Curriculum, Assessment and Reporting Framework.* Queensland, Australia: Queensland Government, Department of Education and the Arts.

Harlen, W. (1994). *Concepts of quality in student assessment.* Paper presented at the Annual meeting of the American Educational Research Association, New Orleans, LA.

Harlen, W. (2005). Trusting teachers' judgement: Research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education, 20*(3), 245–270.

Heinrich, E. (2004). Electronic repositories of marked student work and their contributions to formative evaluation. *Journal of Educational Technology and Society, 7*(3), 82–96.

Johansen, S. J. (1998). "What were you thinking?": Using annotated portfolios to improve student assessment. *The Journal of the Legal Writing Institute, 4*(1), 123–147.

Klenowski, V., and Adie, L. (2009). Moderation as judgement practice: Reconciling system level accountability and local level practice. *Curriculum Perspectives, 29*(1), 10–28.

Lincoln, Y. S., & Denzin, N. K. (Eds.). (2003). *Turning points in qualitative research: Tying knots in a handkerchief.* Walnut Creek, CA: Rowman Altamira.

Maxwell, G. (2002). *Moderation of teacher judgements in student assessment: Discussion paper on assessment and reporting.* Brisbane, Australia: Queensland School Curriculum Council.

McGuire, L. (2005). Assessment using new technology. *Innovations in Education and Teaching International, 42*(3), 265–276.

O'Donovan, B., Price, M., and Rust, C. (2004). Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teaching in Higher Education, 9*(3), 325–335.

Opfer, V. D., and Pedder, D. (2011). Conceptualizing Teacher Professional Learning. *Review of Educational Research, 81*(3), 376–407.

Porter-O'Donnell, C. (2004). Beyond the yellow highlighter: Teaching annotation skills to improve reading comprehension. *English Journal, 93*(5), 82–89.

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading in higher education. *Assessment and Evaluation in Higher Education, 34*(2), 159–179.

Sadler, D. R. (2010a). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education, 35*(5), 535–550.

Sadler, D. R. (2010b). Fidelity as a precondition for integrity in grading academic achievement. *Assessment and Evaluation in Higher Education, 35*(6), 727–743.

Simons, H. (2009). *Case study research in practice*. London: Sage.

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education, 14*(3), 281–294.

Tripp, D. (1993). *Critical incidents in teaching: Developing professional judgement*. London: Routledge.

Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. Cambridge, UK: Cambridge University Press.

Wiggins, G. P., and McTighe, J. (2005). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.

Author reference to be inserted here.

Wyatt-Smith, C. M. (1999). Reading for assessment: How teachers ascribe meaning and value to student writing. *Assessment in Education, 6*(2), 195–223.

Wyatt-Smith, C. M. (2000). Exploring the relationship between large-scale literacy testing programs and classroom-based assessment: A focus on teachers' accounts. *The Australian Journal of Language and Literacy, 23*(2), 109–127.

Wyatt-Smith, C. M. & Klenowski, V. (2013) Explicit, latent and meta-criteria: Types of criteria at play in professional judgement practice. *Assessment in Education: Principles, Policy and Practice*, *20*(1), 35–52.

Zepke, N., Leach, L., Brandon, J., Chapman, J., Neutze, G., Rawlins, P., et al. (2005). *Standards-based assessment in the senior secondary school: A research synthesis. Wellington, New Zealand*: New Zealand Qualifications Authority.

Zywica, J., and Gomez, K. (2008). Annotating to support learning in the content areas: Teaching and learning science. *Journal of Adolescent and Adult Literacy, 52*(2), 155–165.