First published in the *Australian Journal of Language and Literacy*, volume 26, issue 2 (2003).
Published by the Australian Literacy Educators' Association. Copyright © 2003 ALEA. All rights reserved.
This journal online at: http://www.alea.edu.au/html/publications/16/

The nature of teachers' qualitative judgements: A matter of context and salience

Part One: 'In-context' judgement

■

Claire Wyatt-Smith, Geraldine Castleton, Peter Freebody and Ray Cooksey

This paper and the one that follows report on how two teachers make judgements of primary students' writing. The evidence base used by the teachers in their judgements is analysed using qualitative techniques for mapping the inter-relationships among the various facets of students' performance, behaviours and attitudes. This approach makes it possible to examine and directly gauge the impact of evaluative frameworks that constitute the official, publicly available account of assessment as well as the contribution of otherwise implicit but nonetheless significant determinants of teachers' judgements. Of special interest in the first paper is how the teachers enact judgements as having local, point-in-time relevance for teaching and learning.

WYATT-SMITH ET AL. | • AUSTRALIAN JOURNAL OF LANGUAGE AND LITERACY, Vol. 26, No. 2, 2003, pp. 11–32

Introduction

The intersection of teacher assessment with human judgement is a key concern of this paper and the one that follows. Both papers report on part of a three-year Australian Research Council-funded study that is examining the nature of teacher judgement, using literacy as a case study instance. Of special interest in the papers are the ways teachers arrive at judgements about writing quality, including their use of scoring procedures and any other variables that shape how judgements occur. A distinguishing feature of the large study, and of the pilot data set presented here, is that it starts with the individual teacher and investigates his/her approach to enacting judgements made across a range of situations or stimuli. The latter includes writing of the teacher's own current students as well as writing produced by unknown students at the same year level in other school sites. This approach reflects our interest in studying the teachers' judgement behaviour under the naturally occurring entangled judgement conditions associated with students' written work in the teacher's classroom context, as well as those behaviours that come to the fore when the teacher is unfamiliar with the student and the institutional and pedagogical contexts in which the writing was produced. In this way, the papers take up Lapadat's (2000) interest in the

evaluative context of the institution [school/system] and couples it with pedagogical context to investigate 'the social processes of shared experiences and discussion which amplify and make thoughts, behaviours and events meaningful' (42). This focus provides an opportunity to examine and directly gauge the impact of explicit evaluative frameworks and knowledges as well as the contribution of otherwise implicit but nonetheless significant determinants of teachers' judgements.

As already suggested, what follows represents work-in-progress, designed to generate some data-based insights into the variables that teachers draw on and integrate to judge writing in primary schools. Essentially, an aim of the papers is to build a position from which to better understand individual teacher's implicit judgement policies. A related aim is to reconsider how judgement may be intertwined with and shaped by both officially authorised curricular and assessment practices and by other essentially private, local ways of knowing (Belenky et al. 1986) that can be made available through examining teachers' accounts of how they arrive at judgements. The focus of the discussion therefore is on exploring the actual processes that teachers rely on as they read and appraise writing quality rather than on judgement captured as an end point or final grade.

Background

A starting point for the study was that, broadly speaking, in educational practice the issue of judgement is rarely treated as analytically tractable. Consequently the most interesting aspect of the process of gauging students' performance is glossed either by technical procedures that address 'reliability'[1] after both the items and the training and analytic procedures have been put in place, or by the invocation of professional 'connoisseur-ship' or 'insider knowledge' (Sadler, 1989). The 'guild', or those with insider knowledge and experience, thus names itself as having access to what is at the one time taken to be measurable but resolutely implicit. Thus, crucial aspects of performance appraisal rely partly on unstated judgement practices and procedures. The dilemma for understanding teachers' judgements arises at the intersection of the stated (outcomes, criteria/standards) and unstated procedural rules (Huot, 1990; Sadler, 1986, 1989), with variability from instance to instance and combinations of variables in judgement remaining opaque.

A second starting point was Wyatt-Smith's (1999) data-based research on the role defined assessment criteria and standards play in how teachers read and judged student writing in secondary English classrooms. The research was undertaken in the context of the Queensland model of standards-referenced assessment that depends on the professional judgement of teachers. A key finding from this research was that

1 As used here, the term 'reliability' is taken to refer to both individual rater-consistency over time and inter-rater consistency.

published criteria and standards, in and of themselves, did not wholly regulate how judgement occurred. More specifically, Wyatt-Smith showed that while stated criteria and standards made available an official account of how teachers read and appraised student writing, they did not necessarily explicate the variables that teachers relied on to ascribe meaning and value to writing.

So, as we began the study, we recognised that judgement represented largely uncharted territory in assessment research, this situation explained in part by Phelps (1989) who made the point that the deep structures of teacher judgement do not readily lend themselves to examination, even by teachers themselves. In what follows, our interest is in generating a provisional charting of how two practising Year 5 teachers arrived at judgements of writing quality, the aims being to generate professional conversations about judgement and to inform future analyses in the larger study mentioned in the beginning of this section.

Study design

Participants

The data corpus of the study includes semi-structured interviews with eighteen Year 5 teachers, each interview lasting approximately one hour, and approximately 185 hours of recorded talk, in which 37 teachers presented 'think-aloud' assessments of student writing samples. In reporting work-in-progress, this paper draws on a sub-set of the data, presenting an in-depth analysis of the recorded talk of two of the teachers, known as Val and Sue, as they performed 'think-aloud' assessments of Year 5 writing. These two teachers had already completed the semi-structured interviews and were invited to participate in the pilot study on the basis of their extensive experience and their ability and willingness to probe their own judgement processes. At the time of the study, they were team-teaching 61 Year 5 students (average ten years of age) in a school on the outskirts of an Australian capital city and were invited to share the processes they followed in arriving at successive judgements of students' performance in writing.

At the time of data collection, each of the teachers had approximately 20 years' experience teaching in the state primary school system, with one teacher having an additional three years' experience teaching overseas. The student population at the school is diverse and includes students from a number of European, Middle Eastern and Asian countries as well as Aboriginal and Torres Strait Islander students. The Year 5 class in question reflected this diversity, though there were no students who self-identified as Aboriginal or Torres Strait Islander. The school is situated in a suburb described in Australian Bureau of Statistics on-line data as having a median weekly household income of $698, average household size of 2.5, median age 39, and an unemployment rate of 6.7% (2.1% less than the city's average).

Data collection method

The aim of the 'think-aloud' assessments was to capture *judgement in action*, that is, to capture teachers thinking aloud what was salient to them as they read and appraised student writing. As Val and Sue routinely worked as a team in the classroom, initially they asked to work together for a period of fifteen minutes to establish shared assessment expectations of the writing task they had set for their students. We would like to emphasise that the researchers had asked Val and Sue to bring along samples of their students' writing that they had not judged previously. This request reflected our interest in capturing judgement as it happens, as distinct from how it is recollected and made defensible. Following this short period of collaboration, the teachers worked independently, at different locations, to do the 'think-aloud' assessments that involved three stages.

First, the teachers were asked to verbalise the processes they followed as they read and judged each of 25 samples of written text produced by their own students. We refer to these judgements as 'in-context' to indicate how the teachers were already thoroughly familiar with the institutional, curricular and pedagogical contexts in which the samples had been shaped. Next, the teachers were asked to judge 25 previously unseen samples collected from a number of metropolitan schools. The samples were chosen to represent the range of genres and writing performance typically found in Year 5. These judgements are referred to as 'out-of-context', to capture how the teachers on this occasion had access to a deliberately more limited history of the samples to be judged. The teachers were told the time permitted for students to complete the writing task, whether the piece was a first or final draft, and if the students had been given time to research the topic.

Finally, the teachers made judgements on the same 50 samples, on this occasion judging them against the national writing standard or benchmark for Year 5. For readers unfamiliar with the Australian literacy benchmarks for Years 3, 5, and 7, these benchmarks are used in conjunction with state-based testing programs, to generate data on literacy performance (including reading, writing and spelling) for system recording and funding purposes and in our study, represent a 'system context' for judgement. Readers interested in the National Plan and the literacy benchmarks are advised to see Department of Employment, Education, Training and Youth Affairs (1998).

These three stages represented different assessment contexts and therefore provided an opportunity for the researchers to investigate how judgement processes varied across contexts. In what follows, attention focuses first on 'in-context' judgement processes (this first article) and then on the 'out-of-context' judgement processes (the following article). Analyses of teacher judgement of writing quality using the Australian literacy benchmarks – stage three – attends to the interface of federal

literacy assessment policy and local practice, and will form the basis of a third paper. The following section outlines the analytic procedures used in the part of the study reported in this paper.

Analytic procedures

Drawing on the work of Garfinkel (1967), Baker (1997) and Silverman (1997), the 'think-aloud' data were understood as interactional data that generated accounts of how teachers arrived at judgements. As suggested earlier, the purpose of the 'think-alouds', adapted from Miles and Huberman's (1994) cognitive mapping procedure, was to record judgement in action, that is, as it was being formulated, rather than in talk that sought to recapture judgement as it occurred at some prior point in time. The role of the researcher was primarily to prompt each teacher, where necessary, to make available her thought processes as she made judgements of individual student scripts. The intention was to record and explicate what the teachers, individually and collectively, did as they formulated judgements of quality.

The analytic task for the researchers became one of determining how the participating teachers engaged in acts of judging, of understanding 'what is happening here', by focusing on the teacher's talk around the task of judging. In keeping with the emphasis on understanding how teachers account for their judgements, the study is informed by an ethnomethodological interest in members' knowledge of their ordinary, day-to-day affairs, of their own institutions, where that knowledge serves as part of the same setting to which it brings order (Garfinkel, 1967). According to Garfinkel (1967, p. 1), 'the activities whereby members produce and manage settings of organised everyday affairs are identical with members' procedures for making those settings 'accountable' and therefore available for scrutiny by others'. Ethnomethodologists recognise that this 'reflexivity' is both a phenomenon, and a feature, of all social activity, so that this notion of the reflexive accountability of actions becomes of fundamental interest. Linked to the reflexive quality of social action is the concept of indexicality, that is, the indexical properties of normal language use that are essential features of members' accountability procedures in everyday social activity as they constantly reference their commonsense understandings of social structures and actions. Boden (1994, p. 57) has referred to this process as the 'retrospective-prospective nature of accounts' and argues that talk 'provides the primary medium through which the past is incorporated into present action and each are projected into an evolving, never-to-be arrived at future'.

A basic feature of ethnomethodology is the constant prominence given to the local, moment-by-moment determination of meaning in social contexts. Study of local practices and methods throws light on how people achieve rationality, order and structure to arrive at knowledge

in everyday life. An ultimate outcome of an ethnomethodological stance is the immutable fact that regardless of their insight into the matter, social actors are, through their own actions, unavoidably engaged in producing and reproducing the intelligible characteristics of their own circumstances (Hester & Eglin, 1997). In this study the concern of the researchers was to capture how teachers routinely work as judges, drawing on *ways of knowing* (Belenky et al., 1986) that have, to the present, not been the subject of ongoing, systematic research attention.

With these interests in mind, the corpus of 10 hours of 'think-aloud' talk of the two teachers was scanned to identify and code recurring features of judgement. Statements or sections of the talk that exemplified these features were extracted from this corpus to generate a provisional set of judgement indexes. Then a more detailed examination concentrated on each teacher's 'think-aloud' talk to determine the relevance of the proposed indexes in the teacher's talk.

Emerging from the data was a picture of the complexity of the 'educational ecology' (Eisner, 2000, p. 355), in which teachers make judgements about students' performance. The complexity is such that there is no simple, linear course that teachers follow to arrive at their judgements. On the contrary, what emerges is how dynamically networked indexes come into (and out of) play in the judgement process. What follows in this paper is an unpacking of those indexes active in 'in-context' judgements, with the following paper identifying and discussing those operating in 'out-of-context' judgements.

'In-context' judgements

The teachers' think-aloud judgement sessions, though audio recorded in separate locations, generated remarkably similar accounts of the nature of judgement and its reliance on indexes to inform the logic of judgement in action. Specifically, the recorded talk revealed a set of six main judgement indexes, shown in Figure 1 (see page 27). These include:

1. assumed or actual knowledge of the community context in which the school is located
2. teacher experience
3. moderation practices, both planned and incidental
4. assessment criteria and standards
5. first-hand/in-class observations of students
6. knowledge of pedagogy.

The set of knowledges – conceptualised here as indexes for judging – is a construct on our part and we invite readers to consider how they relate to their experiences of and insights into how judgement occurs in other educational settings. For Val and Sue, the indexes were the 'analytic resources' (Baker, 1997, p.132) that the teachers relied on to formulate judgements of student writing, and more generally, to display their

identities as teacher assessors. In what follows, each index is discussed in turn.

Index 1: Community context

The teachers' talk made clear how, early in a school year, they established a latent, or *in-the-head*, standard (Sadler, 1989) as foundational for acts of judging, with the standard being locally defined rather than drawn from official curriculum documents. The teachers determined their standard by drawing, in part, on their early first-hand observations of students, as well as their knowledge and perceptions of the community surrounding the school. In this way, the process of setting the standard extended to the class members and the wider community, with what was to count as appropriate being firmed up and internalised by the teacher in the early weeks of a school term. Sue spoke of this approach to standard setting as follows, describing it as being reliant on a mix of *knowing the kids, the general area* and *the general feel*:

> There are like base marks that you start from, that you think, okay, after being there I usually give myself to Easter when I've gone to a new school. By Easter you, sort of, know the kids. You know the general area and the general feel and so then you work out okay what sort of a standard you're going to make, then within that standard you've got certain children in your class who still aren't going to fit that standard, for a whole lot of different reasons, so those kids have to, you push them along to aim higher, to reach at least the standard that you've made for the school or that particular class. (Sue)

In this segment, the standard is characterised as a type of baseline that has local (as distinct from system) relevance, though it may not *fit* or be appropriate for all students. Missing from the segment, and in the body of data as a whole, is any direct reference to official curriculum documents including syllabus materials as informing how the baseline is established. This omission is striking, especially given that the teachers consistently claimed that the standard informed how they diagnosed student need – *kids who aren't going to fit that standard* – and put in place interventions designed to achieve improvement – *so these kids have to, you push them along to aim higher*.

The segment also indicates how teacher observations of the community surrounding the school (index 1), especially socio-economic status (SES), play a part in standard setting – *you know the general area and the general feel and so then you work out ... what sort of standard you're going to make*. Such observations typically remained unstated, yet were potent in the initial firming up of the standard, allowing teachers to establish what could reasonably be expected from students in 'this' school (as distinct from the school in the next suburb or elsewhere). One teacher captured the recurring issue of SES as shaping teacher expectations, saying:

> you can't *expect*, or I never *expect*, that a child who's come from a low socio-economic school, area where I've taught will do as well as a child who's come from a higher standard school, and area.
> (Sue) (authors' emphasis)

The tie-up between school and local area is repeated here in a way that discloses the apparently unproblematic, taken-for-granted connection between socio-economic status and teacher *expect*ations. The potency of this link has been reported in previous Australian research, including Freebody, Ludwig and Gunn (1996). Emerging in the present study is the potency of the link between assessment standard and SES, especially when it becomes interwoven with talk about family background (as being good/poor), the un/availability of books in the home, parenting practices, including shared reading time, and demonstrated interest and involvement in school learning. The SES-standard-family background connection is evident in the following extract in which 'a good family background' is characterised as being one in which school learning is actively supported:

> The fact that I've, in schools, depending on where the school is, like low socio-economic schools, the *standard that you expect* from the kids there, for me, are not nearly as high as the standard that you would expect who, you know, comes from a *good family background* where there's *lots of books* in the house and the *parents* have spent, even if they're not spending now, but have spent a *lot of time* with the kids and the parents are at least *concerned* in some way about the kid's work that they're going to produce ...
> (Val) (authors' emphasis)

In this extract the teacher is accounting for how she establishes the *expected* standard as one that has relevance for the school, and by implication the student cohort, depending on perceived socio-economic status. Also emerging is how the teacher sees herself as being licensed to tune into the school and community contexts, even family backgrounds, as the means to establish the locally relevant standard.

Interestingly, both Val and Sue started from the position that an externally defined, stable standard was not an available or relevant point of reference and that their site-specific, locally-relevant standard could reasonably be expected to change from site to site and from year to year. This provides an opening for considering how the teachers carried with them their evaluative experience from site to site and class to class.

Index 2: Teacher Experience

At no point did Val or Sue call into question the need for re-establishing the expected standard with each successive cohort of students, nor did they call into question the appropriateness of relying on a standard that may not have relevance beyond the immediate school and community context in which it was established. The corollary of this was that

variability of standard across sites and even over time within a site seemed to be taken-for-granted, normal, as indicated in the following segment of recorded talk between the two teachers:

Sue: The standard of our class this year is probably lower than the standard for her [Val] class last year.
But that's not to say that next year that in grade five you wouldn't be getting a whole group of students through who are going to be above the standard, so you're sort of adjusting as you go along

Val: So the skills that are lacking are the skills that you concentrate on to bring those kids up to what you think will be the acceptable standard for that year.

Present in this talk is the notion of acceptability of the standard for the period of a year, demonstrating what is acceptable can become redefined in response to different student cohorts. Elsewhere in the talk, it became clear that the teachers could determine the locally relevant and 'acceptable' standard by drawing on the evaluative experience that they accrued over their teaching careers. The teachers talked of knowing what students at say, Year 5, could reasonably be expected to demonstrate and know by having worked with successive cohorts of Year 5 in different locations. Sue talked of the connection between how she judged and her teaching and evaluative experience as follows:

> ... you can tell, very easily. I mean, after years of teaching, well I've been teaching 20 years, you can pick up a lot of things that children will try. I mean, there's another aspect of, of our judgement. The longer you've been at the game the more you know what children will either try or cover up or get help from or be sneaky and you know what to expect, um, same with projects. (Sue)

In this segment, Sue can be heard identifying her twenty years of teaching and evaluative experience as a resource for professional judgement. This is of particular interest given the absence of teacher reference to explicitly defined, endorsed standards to inform teachers' judgement work, a point we made earlier. This situation meant that the standard that the teachers firmed up over time at each site remained typically in an unarticulated form and therefore was not readily available for scrutiny or inspection, even by the teachers themselves. The teachers' reliance on their teaching and evaluative experience to inform the standard that they, in turn, knew to be 'shifting' over time and over sites made particular demands on them. The strategies that they put in place to address this situation are discussed next.

Index 3: Moderation practices

As mentioned earlier, Val and Sue worked in a team-teaching situation, sharing the marking for some 61 students. In their talk, both teachers discussed the need for them to be consistent in how they judged student

writing, referring to both rater consistency over time and inter-rater consistency. They also spoke of how they routinely met before marking so that they could jointly establish the features they were expecting to see and value in the writing. Val then described the processes they followed after this preliminary meeting:

> Sue takes her little pile and goes to her house, and I take my little pile and go to my house, and then if there's anybody who we're having a real dilemma about we'll come back and say, 'Have a look at this, you read this and what do you think about this one, you know, is there something there?' Cause reading thirty in a hit is something I don't do. I don't read thirty in a hit. I would never do that. (Val)

In this segment Val indicates how the two teachers not only team-taught, but team-judged on occasions. She makes the point that they were sufficiently confident in their relationship as co-assessors to exchange student papers (and provisional judgements) for the purpose of confirming (or challenging) the suitability of particular judgements. In effect, the teachers initiated their own form of moderation meetings. Sometimes the talk about judgement occurred formally in a scheduled post-marking meeting, and sometimes informally, as one or the other teacher experienced difficulty in arriving at a decision about the value of a paper. This type of mutual support was important to the teachers, given the value that they attributed in their talk to being consistent in their judgements. Val disclosed, for example, that for her a priority was her own consistency over time, and her awareness of how the bases for judgements did not stay stable or fixed throughout the session, especially if she was marking say, 30 pieces of writing. She described this as follows:

> See I find when I go away, if I do them all in one hit I usually end up then going back to the first ones that I've marked because I get tougher as I go on ... if I do them in one hit. Whereas, I find that I'm more even if I don't try to do them altogether. (Val)

Of interest is how both teachers indicated the benefits of opportunities to share judgements in the way described here. They also talked of how such sharing not only provided necessary collegial support (be it in the form of confirmation of or challenge to particular judgement decisions), but also was useful in determining how to engage students in future learning and assessment opportunities. So, for Val and Sue, the judgements had a feedback function to inform students of progress, and importantly, a feedforward role, in the sense that they were used as a resource for reviewing student learning needs and the pedagogy that would be responsive to those needs.

This provides an opening for re-considering the explicit provision of moderation opportunities at system level to support teachers' assessment judgements as they occur in the range of assessment regimes operating nationally. For Val and Sue, as middle school teachers, there was no

such support, and it fell to them to initiate opportunities to share and moderate work samples within the school. Importantly, they put in place opportunities for sharing and moderating, though these did not extend beyond their school. As accountability comes to align with judgement in schooling, it is clear that system support for teacher judgement is vital if public confidence in teacher judgement is to be maintained.

Index 4: Assessment criteria and standards

In the larger study of 37 participating teachers, only the two teachers referred to in this paper reported that they routinely co-developed a statement of assessment criteria for each major writing task. According to these two teachers, the statements did not follow a uniform pattern, with some focusing on textual features exclusively (see Appendix A), while others included planning and editing (see Appendix B). Additionally, the teachers talked of how the development of assessment criteria involved *a valued process of negotiation* (Val). During the think-aloud process, described earlier, the teachers talked of the assessment criteria as being designed to capture the anticipated features of student writing, as well as providing a checklist that, in principle, secured consistency of teacher judgements. Interestingly, they adopted the position that the criteria were the official, public account of judgement, rendering all students on a level playing field. One teacher described how the criteria communicated teacher 'targeting' as follows:

> And the kids are also given the criteria sheet before they do the piece of writing, so the kids know what sort of things we're actually targeting and on each criteria sheet we sort of, target a slightly different thing. (Val)

The status of what was called 'the criteria sheet' as the official or authorised account of how judgement occurred was further strengthened by the practice of sending the sheet home so that parents could also be informed of student achievement. The same teacher spoke of this as follows:

> A criteria sheet, the kids, if they've had a copy they've taken home, they'll be signed by their parents, or parents have seen it, so all year the parents have actually seen how they're going and the marks are the same ... heading as what they get in their final report. (Val)

In this segment the teacher can be heard talking about the criteria sheet as both a record of judgement and a tracking device to assist parents to see progress over time – *how they're going*. It is as though the stated criteria were expected to provide a means for achieving accountability and transparency in judgement. In short, the criteria were the means of making judgements not only available, but also defensible. The following segment shows however, that the criteria alone did not wholly account for how judgement occurred, and that teachers routinely took

account of other factors when judging, including the nature and extent of assistance provided to the student writer:

> He has to be helped to get to the standard we asked. ... And he got to that standard, and he got to that standard but he had heaps of help.... He had teacher's aide help, he had my help, he had not so much help at home, no, but he does spend a lot of time doing the decoration, so, this is another thing that comes into play with assessment, you've got to know the child and what we have to do to build them up. (Sue)

The teacher's talk on this occasion starts with the idea of a fixed standard – *the standard we asked* – and reveals the teacher's assessment decision that the student's work did meet the requirements of the standard – *he got to that standard*. However, the segment also shows that the teacher's act of judging the writing against the standard was not done in isolation from other considerations. More specifically, the teacher could and did readily call on her first-hand knowledge of the assistance that she provided and also of the assistance given to the student writer by the teacher aide and, though to a lesser extent, others at home.

What is made available in the above segment is that the teacher did not read the writing as being a-historical and de-contextualised. Instead, as she read, she brought together her knowledge of the expected standard and what she had observed about the history of the piece of writing – its production history, and how various parties contributed to it. Also emerging in the segment is the emphasis once again on *knowing* and *building* the child, with assessment construed as actively constructing the identity of the student writer.

Index 5: Observations of the student

Val and Sue repeatedly drew on the notion of 'knowing' and 'coming to know' students individually and collectively (as a class) by tracking their development over time, or to use Val's words on one occasion, *See...we know them so well, we can tell what they do from project to project, or piece of writing to piece of writing'*. Close analysis of the talk showed how claims to 'know' the student were tied up with how the teacher recollected prior observations of ability, motivation, personality, and how the student applied him/herself to school learning. This included on-task behaviours both in the classroom and out of school, the latter relating to if and how homework was completed. Of interest was how these perceptions typically remained latent or in unarticulated form, and yet exerted a powerful influence on how teacher judgements actually occurred. In the segment below, for example, Thomas is talked about as a student with limited writing ability who needs encouragement:

> ... so then, um, young Thomas here ... he's tried to use an old worldly sort of a font you know, he, he needs, and too, with Thomas we know that this is a kid that needs encouragement so *(laugh)*. So that influences what you're going to do and what you'd write on here. (Val)

Once again referring to the same student, the teacher drew on a second piece of writing, this time factoring in other perceptions of personality and on-task behaviours:

> Well the thing was with poor old Thomas ... he's one of these kids, he threw his whole heart and soul into this explorer project, his absolute, he was fast, he was active constantly, he did huge amounts of researching, he read lots of stuff, he took mountains of notes and, in the end probably not much saved him, unfortunately for Thomas ... you see the explorer project was actually a spoken presentation with the chart as a support. So his spoken presentation was very good, very good but his, his written presentation is not, I mean he's got lots of things but his mark is probably indicating how much I knew he put into this. You know, he finds it difficult to do anything like drawing a straight line or being able to use a ruler. His fine motor skills, I mean, as you can see by the handwriting are exceedingly low um but for him to produce this map was just you know, really good, really good. (Val)

On revisiting this judgement later in the meeting, the teacher went on to say:

> I guess in a lot of ways I was more lenient than I, if I, he had a just given me this and I had never seen the amount of effort that went into it, if I'd never seen how much of himself he'd poured into this, his mark would probably have been lower. (Val)

Emerging in these extracts is how, as the teacher looked at Thomas's writing, she could readily recollect his spoken presentation, describing the latter as *'very good, very good'*. Also evident is how the teacher recollected the student's enthusiastic engagement with the task, and the emotive language she chose to use to characterise this – *he threw his whole heart and soul into this explorer project, his absolute*. Essentially, it was this act of contextualising the project in relation to what the teacher knew of the student – she had observed (and valued) his effort on this task and knew his writing ability to be limited – that combined to arrive at the assessment judgement of Thomas's map as *'really good, really good'* [for him].

The key point here is that the teacher's judgement is not informed by a fixed standard of textual quality in some absolute sense, but a standard of quality that relates directly to the student and how the teacher is interested in tracking his achievement over time and across tasks. So, in effect, the teacher's repeated words about quality – the description of the explorer project as 'really good, really good' pertain less to demonstrated knowledge and linguistic features of the map than to what she 'knew' of the student.

Throughout the transcripts of the 'think-aloud' talk there were numerous instances of talk showing how the teacher's perceptions of student ability, personality, and effort were foundational to how judgements occurred. It was as though the teacher actively sought to align the

writing to be judged with what they had directly observed of the student/child during class. Also at play was teachers' sensitivity to the ways in which poor grades could impact on student motivation, with judgement being shaped as much by these considerations as by the quality of the writing itself. In short, in the recorded talk the issue of writing quality was represented as inseparable from what the teacher could claim to know of the student/child writer, and especially the student's need for encouragement. In the extract below Sue refers to another male student and similarly shows how perceived effort and need for encouragement become key determinants of judgement:

> We know you've got to also build, because knowing the children enables us to build their confidence, right, so he's weak at language, we know that, so therefore, I know how much effort he put in this and he wrote it out, his draft out about three times so when you see a final product you know, being a classroom teacher, we see what happens in between and what he's built up to and therefore, he got a satisfactory for content, cause he's actually got the content, but overall we know what he's like. (Sue)

There is a clear connection here between the teacher's knowledge of the student as *weak at language* and her observation of his effort – *I know how much effort he put in this*. Also evident is the teacher's interest in protecting the student from the negative effects of grading decisions, especially as these could impact on self-esteem, motivation, and the student's relationship with the teacher. Interestingly, however, the need for teachers to be lenient in judging – *to build their confidence* – was restricted to those students perceived to be underachieving or needing encouragement. For students perceived to be more able or more confident, the role of the teacher-assessor to build and protect student self-esteem did not come into play in arriving at judgement, as shown in the extract below in which the two teachers discussed a piece of writing produced by a 'good' writer:

Val: *(reads student script aloud)*
Departed, not 'departed from'
Sue: Now you're being picky
Val: With the kids that
Sue: 'Cause he's so good
Val: Are of a higher standard ... you tend to
Sue: Be more picky
Val: Look for things that they could, you know, improve on.

While this segment shows Val's interest in reading to inform the student writer of ways to improve, Sue is heard making the point that her colleague is bringing a more critical eye to the writing because she knows the student to be *of a higher standard*. This opens the opportunity for considering how teachers' expectations of individual pieces of student writing could become ramped up (or down) dependent on what they

knew of the student. Of more significance perhaps is the insight that the student became categorised as being of a higher (or lower) standard, with the student being the point of reference for determining distance travelled over time.

So, while much talk focused on effort and knowing the student, only limited references were made in the in-context judgement talk to students' gender. It is worth mentioning that this contrasts with the out-of context judgement talk, and this is discussed in the following paper. Further, the only reference to ethnicity in the in-context talk was made in relation to i) parental pressure that a student was under to produce 'good grades', and ii) how individual student's cultural background could shape their analytic and imaginative capabilities. Consider the following:

> Charlie okay, this child okay, he's um, Vietnamese background. He's very, very clever sort of, typically Vietnamese in terms of mathematics. He's very um you know, analytical. He can think, he thinks about sort of little components and how to do it. He's got a very, so everything with Charlie is going to be exceedingly organised but not overly imaginative. (Val)

On this occasion, the teacher 'read' the student writer as 'typically Vietnamese in terms of mathematics', 'analytical' and 'exceedingly organised', with these features being contrasted with his limited ability to display imagination in his 'Traditional Story' (see Appendix B), even though imagination is not listed as an assessment criterion. For this teacher, the student's cultural background allows her to attribute to the student certain positive features (organisation and analyses), as well as limitations (imagination). Further, it was with this set of expectations that *Charlie is going to be exceedingly organised but not overly imaginative* that the teacher began to appraise the writing. What emerged on this occasion was a picture of judgement as involving trade-offs or compensations. On this occasion, the limitations in imagination were compensated by strengths in organisation. There were several other occasions when the teachers looked at the pattern of performance across stated and un-stated criteria (consider the case of Thomas earlier) and decided to trade-off some features against others to arrive at a judgement of quality. This was done without recourse to any published formula for combining the criteria. Instead, the teachers' practices of trading-off appeared to be triggered by what they saw in the script at hand. Further, these acts occurred without leaving a trace (except in the talk). This makes an opening for considering that much needs to be learned about how teachers undertake such trade-offs as they formulate judgement.

Index 6: Knowledge of pedagogy

Just as the teachers drew on various observations and perceptions of the child/student writer, some of them quite detailed, so too they drew on

detailed recollections of their own pedagogy, including classroom talk and other interactions, as they read and judged student writing. Repeatedly, the talk made clear how the teachers not only related the piece of writing to the child writer, but also they tied the writing and their valuation of it to the teaching and learning contexts and practices in which the writing had its origins and purpose. Val captures these connections in the segment below in which she discloses how, as she judges the writing, she can recollect episodes of 'kidwatching' (see Gollasch 1982):

> You can see it, you know, but the feeling that you get about the kid, that's influencing what goes into this is all of the other things that you see every day, you know, when you're sitting there watching that kid or when that kid's coming to your table and he's asking, you know, Does this sentence make sense? Is this sentence right? Then that kid will change that sentence because of some talk that you've had, not because, you know, he necessarily knows that (pause) ... you know, you know, like I know that he's used the word rotated in that story because we talked a lot about what rotated was (pause), and I can just about guarantee that every single kid in there has used the word 'rotated'. (Val)

Even though Val says that *you can see it*, her talk reveals that she is the agent who does the seeing, as teacher, and more specifically, as authoritative teacher-of-writing in the classroom. The disclosure on offer is that Val not only sees the child/student *in situ*, but sees the writing in front of her *in situ*. She can recall, for example, how the student asks questions and how she makes suggestions to improve drafts. Also disclosed is Val's awareness of her authority not only in the classroom, but also in shaping how the writing comes to be. Consider for example her prediction that she could *just about guarantee that every single kid in there has used the word 'rotated'*.

The pedagogy index was particularly salient in that it was regularly updated by what Val referred to above as *All of the other things that you see very day*. It was all of these other things that enabled the teachers to locate a single piece of writing in relation to the larger body of writing produced by the student, as well as in relation to recollected talk about writing during its development. Once again, Val described how judgement was tied to such recollections as follows:

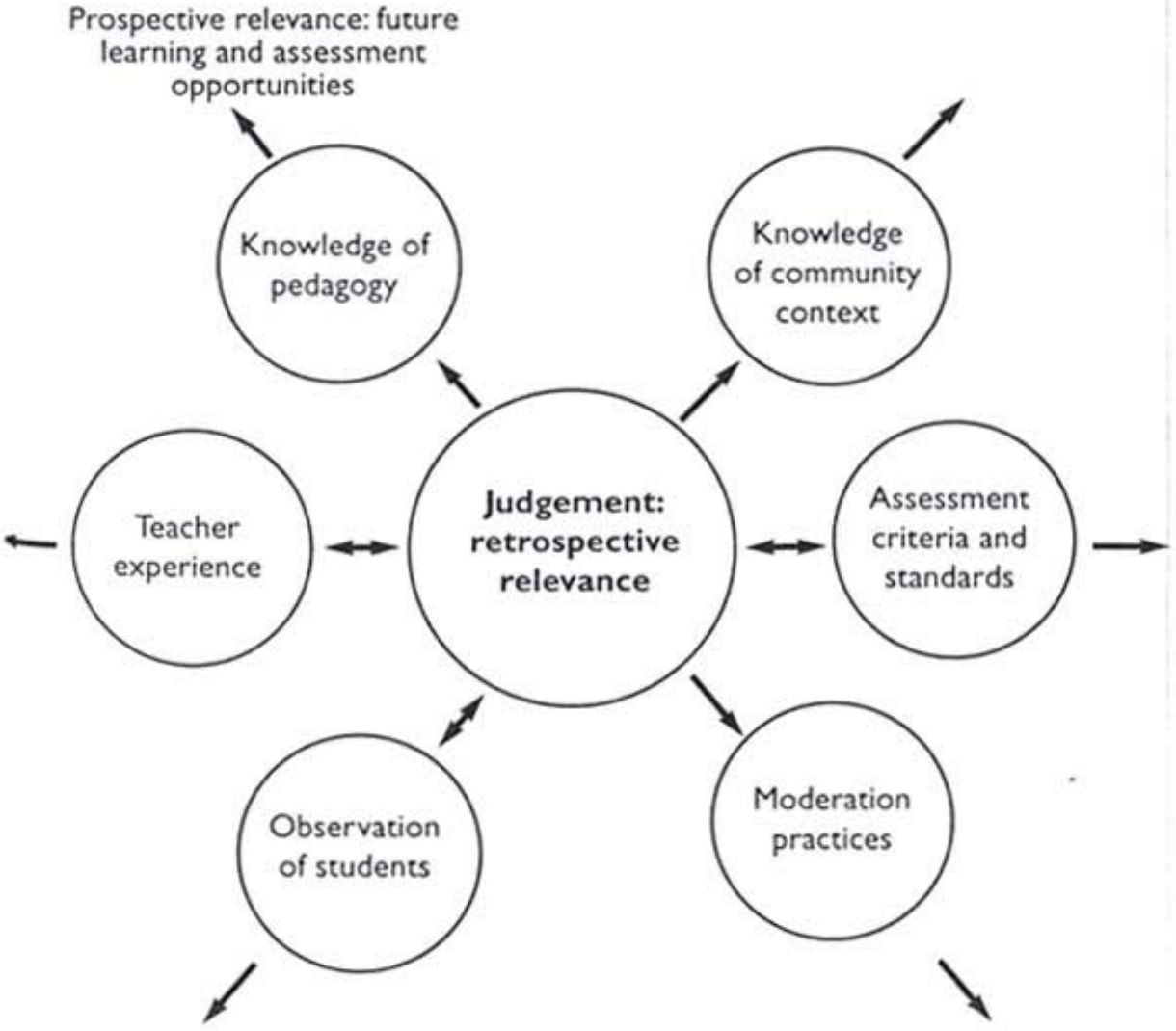
> We know our group of kids because we've read so much of their work and we've talked to them a lot you see, it's the talking bit that, little individual chats that you have with kids about certain parts of their story that make all the difference, you know, Has he taken on board what I said about such and such a thing? Have they actually, or are they still writing in the same way without having tried to improve their sentence starters. (Val)

On this occasion, Val highlights how, as she appraises student writing, she recollects *little individual chats*, as well as observations about how

students have reacted to her suggestions about how to improve the writing. The talk also shows that the teachers deliberately look to see if the students have initiated their own improvements. The teachers' willingness to tune into signs in the writing of a student's effort to improve was also evident when they talked about rewarding those students who used resources and ideas in addition to those provided in class. Essentially, what was at play was an un-stated set of assessment criteria relating to recollected classroom interactions and student initiative, the key point being that these were not made public or available to students or parents as factors contributing to judgement.

In the remainder of this paper we ask readers to revisit the notion of indexicality as we apply it to judgement. To this end, we have developed a diagrammatic representation of the indexes (see Figure 1 below), suggesting how they are interrelated, shape judgement, and have both retrospective and prospective relevance to how teachers work as judges of student writing. These matters are taken up next.

Figure 1. Representing in-context judgement



How the indexes worked

Essentially, the indexes were the means that the teachers used to capture both the logic of judgements and the social relationships that they (as teacher assessors) shared with their students (as recipients of teacher judgements). As such, the indexes allowed the teachers to account for how judgements are 'made to happen' and in this way, they were relied on to impose order on judgement possibilities at a particular point in time. In effect, the indexes were called into play or activated in such a way that they had a point-in-time relevance. That is, the set of indexes worked to constitute the judgements as rational and defensible.

So, while the set of indexes remained stable across the body of data, the emphasis or priority given to particular indexes and the ways in which they were combined were shown to vary not only from teacher to teacher, but also from judgement to judgement, as the teachers moved from one piece of writing to the next and oriented to different aspects of the text. The set of indexes can be understood therefore as comprising a dynamic network of available knowledges that can be combined variously to account for how judgement occurs on a particular occasion.

The third point relates to the *retrospective* and *prospective* relevance of the indexes. This paper has suggested how the indexes had a retrospective relevance in the sense that they worked as the medium through which the teacher linked prior knowledges to present judgement action. Further, the indexes carried a prospective relevance, having potential to carry over to future teaching and learning interactions in the classroom, and the individual teacher-student relationship more specifically.

Conclusion

This paper draws on a limited data set that will be supplemented by data from the larger study involving a total of 20 (of 37) teachers for whom there were verifiable and complete data sets. Already, however, the analyses show that the texture of teacher judgement is complex, having sensitivity to localised contexts and dynamics tied to specific pieces of writing. This of course, raises significant issues about judgement reliability both across teachers in one site and across sites. Further, the paper raises the worrisome issue that the processes of arriving at site valid judgements may be fundamentally different from those adopted to achieve system validity. If a high level of congruence/agreement in judgements about student work is a desirable characteristic (as traditional psychometric models would assume), the pilot reported in this paper suggests that much needs to be learned about teachers' judgement processes, and far more system support is required for teachers to engage in rigorous moderation processes using defined standards.

The larger study referred to earlier in the paper is building on the pilot study reported here and will engage with debates about system and site validity, examining in more detail the application of teacher's

individual judgement policies. It will bring together judgement analysis methodology (Cooksey 1996) and verbal think-aloud protocol methodology. As Einhorn and Koelb (1982) indicated, these methodologies target the understanding of judgement at different levels of generality. Of special interest will be the complementarity of the judgement modelling technologies and how they target the understanding and enacting of system and site valid judgements. The challenge is to capture the interior dynamics of individual teacher's judgement processes that, as mentioned earlier, represent largely unchartered terrain. For us, the key question that remains unresolved is: Which type of validity, system or site, should dominate in any public statements about the written literacy level of students? Already from the pilot it is clear that site-validity and contextual dynamics in teacher judgement processes cannot be ignored or assumed to be arbitrary or inherently flawed.

At this point, we turn to Part Two. This paper considers the indexes that the teachers in the pilot called on when asked to judge student writing in a judgement context where they did not know the student and had limited information about the institutional and pedagogical settings that generated the writing.

References

Baker, C. (1997) Membership Categorization and Interview Accounts. In D. Silverman (ed.), *Qualitative Research: Theory, Method and Practice*. London, Sage Publications.

Belenky, M.F., Clinchy, B.M., Goldberger, N.R., & Tarule, J.M. (1986) *Women's ways of knowing: The development of self, voice, and mind*. New York: Basic Books.

Boden, D. (1994) *The Business of Talk: Organisations in Action*. Cambridge, Polity Press.

Cooksey, R. (1996) *Judgement Analysis: Theory, Methods and Applications*. San Diego, Academic Press.

Department of Employment, Education, Training and Youth Affairs. (1998) *Literacy for All: the Challenge for Australian Schools. Australian Schooling Monograph Series No.1*, Canberra, Australian Government Publishing Service.

Einhorn, H. & Koelb, C. (1982) A psychometric study of literary-critical judgement. *Modern Language Studies*, 12, pp. 59–82.

Eisner, E. (2000) 'Those who ignore the past ...: 12 'easy' lessons for the next millennium', *Journal of Curriculum Studies*, 32(2), pp. 343-357.

Garfinkel, H. (1967) *Studies in Ethnomethodology*. Englewood Cliffs, New Jersey, Prentice-Hall.

Freebody, P., Ludwig, C., & Gunn, S. (1995) *Everyday literacy practices in and out of schools in low socio-economic urban communities*. Commonwealth Department of Employment, Education and Training Griffith University, Brisbane, Centre for Literacy Education Research.

Gollasch, K.K. (ed.) (1982) *Language and Literacy: the Selected Writings of Kenneth S. Goodman*. London, Routledge & Kegan Paul.

Hester, S. & Eglin, P. (1997) *Studies in Ethnomethodology and Conversation Analysis*. Lanham, Maryland, International Institute for Ethnomethodology and Conversation Analysis & University Press of America Inc.

Huot, B. (1990) Reliability, validity, and holistic scoring: What we know and what we need to know. *College Composition and Communication*, 41, pp. 201–213.

Lapadat, J. (2000) Evaluative Discourse and Achievement Motivation: Students' Perceptions and Theories. *Language and Education*, 14(1), pp. 37–61.

Miles, M. & Huberman, A. (1994) *Qualitative Data Analysis: An Expanded Sourcebook* (2nd edn). Thousand Oaks, CA, Sage Publishing.

Phelps, L. (1989) Images of student writing: the deep structure of teacher response. In C.M. Anson (ed.), *Writing and Response: theory, practice, and research*. Urbana, Illinois, National Council of Teachers of English.

Sadler, D. (1986) *Subjectivity, objectivity, and teachers' qualitative judgements*. Assessment Unit Discussion paper. Brisbane, Board of Secondary School Studies.

Sadler, D. (1989) Formative assessment and the design of instructional systems. *Instructional Science*, 18, pp. 119–144.

Silverman, D. (1997) Introducing Qualitative Research. In D. Silverman (ed.), *Qualitative Research: Theory, Method and Practice*. London, Sage Publications.

Wyatt-Smith, C. (1999) Reading for Assessment: how teachers ascribe meaning and value to student writing. *Assessment in Education: principles, policy and practice*, 6(2), pp. 195–223.

LITERACY ASSESSMENT

A Collection of Articles From the Australian Literacy Educators' Association

Heather Fehring, Editor

Stressing the importance of teacher judgment in literacy assessment and reporting, this collection of 19 previously published articles from the Australian Literacy Educators' Association will help you make informed decisions about your K–12 students' literacy achievement. The articles explore literacy assessment from a theoretical perspective, report on applicable research studies, discuss the implications of standards and benchmarking, and demonstrate practical classroom implementation of assessment and reporting techniques.

Available from
ALEA Bookshop
PO Box 3203, Norwood SA 5067 tel (08) 8332 2845
Price $39.00 plus $4.00 postage



INTERNATIONAL
Reading Association
800 Barksdale Road, PO Box 8139
Newark, DE 19714-8139, USA (302) 731-1600
www.reading.org

APPENDIX A

School Camp – Newspaper Report

Overall Assessment _____

NAME: _____ TEACHER: _____

CRITERIA COMMENTS	1	2	3	4	5
Presentation/Format					
Heading Picture/ad					
Punctuation					
Paragraphing					
Sentence Structure					
Tense Verbs					
Sequencing					
Intro & Conclusion					
Spelling					
Vocabulary					

General comment:

APPENDIX B

Traditional Story

Overall Assessment _____

NAME: _____ TEACHER: _____

CRITERIA COMMENTS	1	2	3	4	5
Planning					
Spelling					
Punctuation					
Grammar-Sentence Structure					
Vocabulary					
Sequencing					
Paragraphing					
Editing					
Presentation					

General comment:

First published in the *Australian Journal of Language and Literacy*, volume 26, issue 2 (2003).
Published by the Australian Literacy Educators' Association. Copyright © 2003 ALEA. All rights reserved.
This journal online at: http://www.alea.edu.au/html/publications/16/