Exploring the relationship between large-scale literacy testing programs and classroom-based assessment: A focus on teachers’ accounts

Claire M. Wyatt-Smith

The launch of Australia’s first National Plan for literacy and numeracy education has further intensified interest in and debate about literacy assessment and standards in schooling. In particular, it has brought to prominence on national and state education policy agenda a particular construction of accountability dependent on ‘rigorous’ testing, measurement and reporting of outcomes to systems, parents and the wider community. This paper situates current testing initiatives across the country in relation to a significant tension between different assessment goals. It then examines teachers’ accounts of testing programs, as they and their students participated in them in Year 3 and 5 Queensland classrooms in 1998. The analyses focus specifically on how the teachers talked about the relationship between the testing programs and their own classroom-based assessment. Finally, the paper argues for congruence in assessment and outlines the challenges that must be addressed in order for this to be realised.

Introduction

Literacy education and its assessment are currently major concerns for government, educational professionals and the community, with policy implications that reach beyond educational practice. In countries around the world, many people have come to regard literacy rates as important features of the educational, cultural, and economic status of their communities or nations. In some circles these rates are taken to be indicators of national educational success and of economic potential and well-being; the United Nations Development Program, for example, uses three indices in the composition of its ‘Human Development Index’: life expectancy, literacy and per capita GDP (see UNESCO 1990; Graff 1987). One result of the high profile for literacy education is that improving and assessing literacy have been the topics of many debates, and these debates have been not just about the comparative merits of competing theories of reading and writing, or of different ways of teaching them. They are often about ‘standards’, with such standards often referring almost simultaneously to standards of overall educational, cultural and economic success.
economic activity. So research and policy development in literacy education have come to enjoy high levels of activity, diversity and controversy at state, national and international levels.

Recent literacy education policy initiatives at the national level in Australia have expanded not only the stakes, but also the breadth of the debates, and the kinds of solution and stimulus that enhanced literacy provision is called upon to offer. These debates, and the common practices and aspirations they reflect, make sense only when viewed against the larger changes going on in Australian society. The political, industrial, economic, technological and cultural patterns that make literacy education ‘a problem’, and that give shape to the solutions that are sought, are themselves shifting rapidly. Many of these shifts impact directly on what kinds of literacy activities and competencies people value and claim that they or other groups of individuals need.

Of specific interest in this paper is how the debates also need to be viewed as being centrally concerned with what counts as quality literacy teaching, ways of gauging teaching effectiveness, and the nature of assessment itself—its various purposes, and specifically, the problem of unifying assessment and instruction. More than a decade ago, Cole (1988) outlined the problem as follows:

Assessment is closely associated with two legitimate but different goals—the goal of measurement (and the accountability and policy goals it serves well) and the goal of instruction. The fundamental problem then is the compatibility or incompatibility of these goals (pp. 108-109).

Although Cole (1988) was referring to the American context, her identification of possible tensions surrounding the different goals for assessment applies equally well to other countries. In the United Kingdom, for instance, the measurement goal has gained considerable influence, as reflected in government plans to close so-called failing or ‘sink’ schools, to introduce performance-related pay, and to continue to publish national test results for Science, Maths and English in league tables, including at a web site. Additionally, there are proposals to attach funding to pupils, not schools, in the name of allowing parents greater choice in school selection, and to get private firms to manage schools under contract to the state.

Currently in Australia there are no such published league tables showing school performance in state-based testing programs. However, the country’s first National Plan for literacy and numeracy in schooling, launched in 1998, makes a strong policy commitment to achieving accountability through rigorous testing, measurement and reporting as

1. For interested readers, site details are: http://www.bbc.co.uk/education/schools/. The league table information is one of the options on the home page, the direct link being http://news.bbc.co.uk/hi/english/education/newsid-216000/216975.stm
early as possible on school entry, and then at Years 3, 5, 7, and 9. (Readers interested in a detailed discussion of the range of state-based testing programs operating throughout Australia are advised to see Wyatt-Smith & Ludwig (1998)). Essentially, the Plan, as elaborated in Literacy for All: The Challenge for Australian Schooling (Department of Employment, Education, Training and Youth Affairs 1998), promotes the understanding that the testing-measurement-reporting mix is the vital means for securing improved literacy and numeracy outcomes. The Plan also sets out the policy testing imperative for generating clear, unambiguous information about outcomes, and for providing these to parents and the wider community so that they are informed about school performance, and therefore, school selection. And as the title of the national policy monograph suggests, a distinguishing feature of the Plan is how it reframes the national testing policy initiatives in terms of equity and social justice in schooling, taking up the position that testing is the vital lever for achieving improved outcomes for all.

Against this backdrop, it is timely to ask: how are we balancing the instructional-measurement goals in practice? How do we account for and understand the playing-out of the goals in classrooms, especially in relation to issues of diversity, equity and inclusive curriculum? Specifically, how do we understand the relationship between state-based standardised literacy testing programs in Years 3, 5 and 7, on the one hand, and on the other, routine classroom assessments, under the control of the teacher? What are the possibilities for congruence between the two? And what is the nature and function of teacher judgement and standards in these different assessment approaches? The remainder of this paper presents the findings of a study that addressed these questions by examining teachers’ classroom-based assessment practices and their spoken accounts, recorded in interviews, of being involved with the state-based literacy testing program in Queensland. The sites and participants involved in the study are outlined briefly below.

Sites and participants
Seven teachers from three schools participated in the study. The three schools selected were all part of Education Queensland’s Special Program Schools Scheme, a Commonwealth-funded, state-administered program designed to measurably improve literacy and numeracy outcomes for educationally disadvantaged students, including students from low socio-economic backgrounds. All three schools catered for students from Years 1 to 7 and had state pre-schools attached. Additionally, all three schools had a high proportion of students whose main language spoken at home was not English. Other noteworthy characteristics of the student population at School 2, identified by the teachers themselves, were the high turnover of students at the school and the high proportion of students with identified learning difficulties. Total enrolments at the
three schools and numbers of students whose main language spoken at home is not English are shown in Table 1.

*Table 1. School composition (data obtained from the on-line Schools Directory database).*

<table>
<thead>
<tr>
<th></th>
<th>School 1</th>
<th>School 2</th>
<th>School 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total enrolment (February 1998 census)</td>
<td>225</td>
<td>570</td>
<td>218</td>
</tr>
<tr>
<td>Number of students whose main language spoken at home is not English (MLOTE Survey, February 1997)</td>
<td>155</td>
<td>139</td>
<td>62</td>
</tr>
</tbody>
</table>

At School 1, 127 of the 155 students identify Vietnamese as the main language spoken at home; 8 students give Hindi as the main language spoken at home; 10 other languages are listed, with small numbers of students for each.

At School 2, 82 of the 139 students identify Samoan as the main language spoken at home; 21 students give Vietnamese; smaller numbers speaking other languages (13) are listed.

At School 3, 20 of the 62 students identify Arabic as the main language spoken at home; small numbers speaking other languages (17) are listed.

Characteristics of the seven teacher participants and their distribution across the three schools are shown in Table 2. All the teachers involved in the study were female.

*Table 2. Characteristics of teachers*

<table>
<thead>
<tr>
<th></th>
<th>Year level taught</th>
<th>Years of teaching experience*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher 1</td>
<td>Year 3</td>
<td>25 years</td>
</tr>
<tr>
<td>Teacher 2</td>
<td>Year 5</td>
<td>30 years</td>
</tr>
<tr>
<td><strong>School 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher 3</td>
<td>Year 3</td>
<td>15 years</td>
</tr>
<tr>
<td>Teacher 4</td>
<td>Year 3</td>
<td>30+ years</td>
</tr>
<tr>
<td>Teacher 5</td>
<td>Year 3</td>
<td>15 years</td>
</tr>
<tr>
<td>Teacher 6</td>
<td>Year 5</td>
<td>29 years</td>
</tr>
<tr>
<td><strong>School 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher 7</td>
<td>Year 3</td>
<td>25 years</td>
</tr>
</tbody>
</table>

*Includes time in administrative/advisory roles as well as classroom teaching.*
Data collection—nature of data
The nature of the data collected at each site varied in accordance with the number of teachers involved and the nature of their involvement. In general, two main types of classroom data were collected—observation and audio recording of classroom talk, and collection of related classroom artefacts. Classroom talk was transcribed for more detailed analysis.

Additionally, teachers were invited to participate in two semi-structured interviews, the first relating to their everyday or routine classroom-based literacy assessment practices and the second, the state-wide literacy testing program in Queensland in the period 1995-1998. The concern was less with the specific design features and tasks that comprise the tests in each of these years than with how the teachers talked about the goals, methods and consequences of the testing program, and its fit with their own assessment practices. Because of timing issues, not all teachers were able to participate in all parts of the study. Table 3 summarises the distribution of data across the participants. As mentioned earlier, only the findings from the interview data are presented in this paper.

In total, approximately 14 hours of observations were made in Year 3 classrooms while 7 hours of Year 5 classroom observation data were collected.

Analyses of teacher interviews
Each of the teacher interviews was audio-recorded and fully transcribed. The analyses of the interviews focused on what they brought to light about how the teachers saw their classroom-based literacy assessment practices.

Table 3. Type and quantity of data obtained at each site

<table>
<thead>
<tr>
<th>School 1</th>
<th>Year level</th>
<th>Interview data</th>
<th>Classroom observation*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher 1</td>
<td>Year 3</td>
<td>Both interviews</td>
<td>3 sessions</td>
</tr>
<tr>
<td>Teacher 2</td>
<td>Year 5</td>
<td>Both interviews</td>
<td>2 sessions</td>
</tr>
<tr>
<td>School 2</td>
<td>Year 3</td>
<td>Both interviews</td>
<td>2 sessions</td>
</tr>
<tr>
<td>Teacher 4</td>
<td>Year 5</td>
<td>First interview only</td>
<td>1 session</td>
</tr>
<tr>
<td>Teacher 5</td>
<td>Year 3</td>
<td>Both interviews</td>
<td>2 sessions</td>
</tr>
<tr>
<td>Teacher 6</td>
<td>Year 5</td>
<td>First interview only</td>
<td>2 sessions</td>
</tr>
<tr>
<td>School 3</td>
<td>Year 3</td>
<td>Both interviews</td>
<td>No classroom observations.</td>
</tr>
</tbody>
</table>

*Although exact start and finish times varied slightly between schools, each school had a morning session from first bell to morning tea (about 2 hours), a middle session from morning tea until lunch (about 1.5 hours) and an afternoon session. No classroom observations were made during the afternoon session.

The Australian Journal of Language and Literacy
practices relative to the statewide literacy testing program. Of special interest was how the teachers constructed versions of themselves as teachers and assessors, and versions of the students as learners and participants in the two assessment approaches.

The analyses started from the position that the teachers' accounts of themselves and their students are part of the world they describe (Garfinkel 1967). As such, readers are asked to consider how the findings presented below relate to other institutional worlds of schooling, literacy education, and its assessment. Further, the analyses drew on the work of Silverman (1993, 1997) who treats interview accounts 'as compelling narratives' (p. 114). Specifically, the transcripts were analysed for the attributes, knowledge and assumptions that the teachers made about themselves, and their students, in the contexts of classroom-based assessment and statewide testing. In this way, the analyses opened up avenues for understanding and practice, in this case with respect to literacy teaching, learning and assessment. As discussed below, four main issues emerged in the analyses, namely: (1) varying definitions of 'essential'; (2) assessment purposes; (3) contexts; and (4) what counts as valued assessment evidence. Each of these issues is addressed separately.

**Findings**

**Policy and teacher accounts of 'essential' aspects of literacy**

In considering the matter of what counts as essential aspects of literacy and numeracy, two points need to be made at the outset. First, the National Plan gives priority to the measurement of students' progress against agreed benchmarks for Years 3, 5, 7 and 9, and progress towards national reporting on student achievement against these benchmarks (Department of Employment, Education, Training and Youth Affairs 1998). Second, the national policy defines the benchmark at each year level as intending:

> To set a minimum acceptable standard: a critical level of literacy and numeracy without which a student will have difficulty in making sufficient progress at school. The benchmarks therefore identify the essential aspects of literacy and numeracy. (DEETYA 1998, p. 23)

According to this statement, the benchmarks capture and make available for scrutiny one account of 'essential aspects': an account that is stable and powerful, at least in policy terms, given the stated expectation that it is to be used to inform national reporting on the literacy and numeracy achievement of Australian school-age students.

Given this, it is surprising—even alarming—that the teachers reported that they had no prior knowledge of the National Plan in general or
the benchmarks in particular. Further, they did not have a clear sense of the purposes of the Year 3 sampling and Year 5 testing programs, but reported that, as far as they were concerned, the programs had little, if any, curriculum relevance. A Year 5 teacher made the point that the reports ‘just told you what the children can do and couldn’t do on the day, on that particular test’. Another teacher commented that the reports ‘just told me what I knew already’. They provided a point-in-time assessment that the teachers saw to be of limited, if any, diagnostic use. They made no mention of using the reported test data to mount intervention programs for individuals or groups of students. A Year 3 teacher talked of her involvement in testing saying:

When we do the test we’re not even going to get the results from the test. So what we’ll do with the test is we’ll look at it before we send it away. Apparently at the end of the year they’re going to send us a follow-up test to see if we’ve improved. I don’t know, but I know that nobody’s going to get reports from this. I think they’re just getting the standard. I’m not sure. We’re just going to do the test, send it away, and we’ll never hear another thing. That’s how I understand it.

Of interest in this extract is the teacher’s use of we and they. She talked of how ‘we do the test’, and ‘we’re just going to do the test’ as though the teachers took themselves to be test takers, along with the students. And in repeating the phrase ‘we send it away’, we hear the teacher emphasising how the test came into the school from an external authority and was returned to that authority, possibly to determine if ‘we’ve improved’. There is a ring here of the teacher and student pitting themselves against the external examiner, reminiscent of the days of the external junior and senior public examination system. There is also the sense that the teacher was largely ‘in the dark’ about testing and reporting purposes, and did not expect to be informed about what was done with the material ‘we’ exported from the school—‘we’ll never hear another thing’. The comment ‘nobody’s going to get reports from this’ shows how, at least from the teacher’s viewpoint, the test data were for ‘in-house’ system purposes, with no direct pedagogical implications for her practice.

The point here is that while, officially, the benchmarks identify the essential aspects of literacy, and the design of the Year 5 literacy and numeracy tests are informed by the benchmarks, as well as the state’s curriculum, the teachers did not see clear connections among the testing program, official curriculum documents and locally-developed work.

2. The 1998 Year 5 test was a census test, that is, it was sat by the whole cohort of Year 5 students in the state, with some authorised exemptions. Similarly, the Year 6 test in the preceding three years had been a census test. With the introduction of Year 3 testing in 1998, the policy decision at state level was against census testing and in favour of ‘sampling’, that is, selecting a sample of schools to participate in the testing program for that year.
programs, the National Plan which spawned the benchmarks, the benchmarks and the reports returned to the school. Certainly they had no knowledge of the benchmarks themselves or what the benchmarks take to be ‘essential’.

What then did the teachers take to be essential in their teaching and assessment practices? While each of the teachers referred to and drew on relevant syllabus documents to construct their own individual accounts of ‘essential’, broadly speaking the accounts tended to fall into two categories. In the first category was talk about ‘essential’ as inevitably defined in terms of the local school and community context and which was understood to be needs-based. As such, what the teachers counted as the essential aspects of literacy and numeracy could vary even widely from year to year, and from student to student, being determined by their perceptions of individual student’s needs and capabilities—‘where the kids are at and what they can achieve’. In accordance with this view, essential was always being redefined and as such, could not be wholly pre-specified. In the second category, there was talk of essential as a core of learning that was currently ill-defined and in need of stabilising through standards specifications. In this category, the teachers talked of the essential aspects as needing to be fixed in the interests of teachers and students.

Category 1: Essential as locally defined

Year 3 teacher:

No, I don’t think what’s essential stays the same. Depending on what the class is like, the individuals in the class, I think you’re always trying to extend them. So these kids here, I’ll be happy if they can understand what they’re doing. I’ll be happy if they can understand what a maths thing is about and they can understand what they are doing. Like in maths, like Tran, she can do any number fact you give her. You give her $6 \times 9, 12 \times 8$, she can answer it. But you ask it to her in the middle of a problem, she doesn’t realise she’s got to multiply things together. That sort of thing. So with these kids. But I was at, say School B. Those kids are already there. They do that sort of thing. They come from very literate homes. You know they’ve got that coming to school, so then you move on to something else, and you try to give them something else. Different areas, different kids, the teachers will be focusing on different things.

In this extract understandings about ‘essential’ hinge on understandings about local school context in the local community: who the students are, their cultural and linguistic backgrounds, their access to resources, both human and material inside and outside the school; and their fluency with spoken, written and visual English. In effect, what counts as essential is talked about as being inevitably tied to the perceived needs of the cohort, and therefore cannot be wholly pre-specified in authentic or locally relevant ways in syllabus and other policy documents.
Category 2: Essential as core learnings and as standards-related

I think what’s essential, it’s well, more fixed. I think it needs to be, I mean in a way a standard, you know? I think that there are some variations around the fringe and I’ve got some students in my class who are, because they may be very advanced in their learning and their use of language, they might, you know, there might be areas that they, like they might be able to spell very difficult scientific words that other students can’t spell. On the other hand, I’ve got students in my class for whom spelling is very difficult so there may be words that they learn to spell that are like a core for them that, you know, almost like a Year 2 level, so within any year level you would have to make provision for students who are either very much up this end or very much down that end. But I still believe that there is a core of essential learning in language.'

In this extract, the essential is still associated with responsive, needs-based teaching, a characteristic noted above. Additionally, the essential is framed within the teacher’s belief that a core of essential learning in language does exist and that this is tied, albeit in an ill-defined manner, with standards, understood to refer more to content rather than to characteristics or features of achievement levels. Elsewhere in the interview, the teacher reported that in her previous work as an Education Advisor she had encountered repeated calls from teachers asking for greater specificity in terms of pedagogical and assessment content.

Overall, both sets of understandings about ‘essential’ remain fuzzy or undeveloped, with Category 1 tied more tightly to perceptions of students than to understandings about curricular knowledge and skills, and Category 2 presenting a sense of a connection between essential and standards, though the nature of the connection is unclear. When we put the extracts together, what is clear is the need for systems to improve their performance in communicating to teachers information about national and state policy initiatives in literacy testing and how these relate to curriculum and pedagogy. Also clear is the urgent need for teacher involvement in debates about what is essential—essential for what, to whom, and why? This brings us to consideration of assessment purposes and what counts as normal.

Assessment purposes

In the teachers’ talk they consistently defined assessment as being primarily for the purpose of capturing the individual student in action across learning situations and tasks. The point of comparison, they said, is the student with his or her self over time. They reported their primary role as being to monitor individual learning in a range of classroom interactions. A Year 3 teacher commented on her assessment purposes and collection methods as follows:
What I do, I compare the child in relation to where he or she was previously. So to me the only real form of assessing that matters is progress over time. You know, I have collected samples of a child’s writing at the beginning of the year, in the middle and so on, so I’m able to look. We have folders, for every child we have one of these in a filing cabinet … the only assessment that matters is comparison with where they were, so if you look at the beginning of the folder, this is what he was doing in the beginning of Year 1 and if you have a look now it’s only two-and-a-half years later, the middle of Year 3 and look at the writing that he’s doing … you can really see how much he’s learnt, and I think that’s fair, I think to compare them with other children, particularly in a school like this. It’s not equitable, you know, you’ve got so many different … such diversity. So it’s comparing the child with himself over a period of time. I know that that might not please some people, you know, and I know that there would be people in the community in this school, parents who would want something a little more standardised, more formal, more of a benchmark.

In this extract we hear the teacher’s rejection of direct inter-student comparisons in favour of monitoring or tracking individual achievement over time, in fact over years of schooling, including the collection of assessment evidence of a range of types. Assessment, as the teacher talks about it here, is continuous, having a clear curricular and pedagogical relevance, with the insights derived from it used also for reporting purposes. Additionally, there is also a significant hook-up between this type of assessment as active tracking and the teacher’s understandings of equity and diversity. She makes clear her position that students come to school as different individuals and that difference has to be factored into how she goes about doing classroom assessment.

The hook-up between assessment and fairness was a common feature of the teachers’ talk, with one Year 5 teaching saying:

Well I don’t compare them with other year levels, nor do I compare them with other classes. I only very informally and in my own mind compare them to other children in the class. I don’t think that’s fair. It just doesn’t fit in with the way I think about it. I compare them against themselves and that goes on, you can look at their work, the progression of their work over time. Over the year and it’s built up with everything in their work. In their groups if they’re performing they can change groups. It’s very fluid. I compare them against themselves only really.

Across the corpus of teacher talk there was a strongly voiced commitment to formative assessment, understood as including the dual concerns of diagnosis and improvement, as shown in this extract. The talk demonstrated that teachers could recall, sometimes in vivid detail, how individual students responded to particular activities and resources, and how they could see progress over time. In this monitoring role, the Year 3 teachers referred frequently to the Reading and Writing Continua.
(Education Department of Western Australia 1994), indicating its usefulness in informing them about performance characteristics. The Year 5 teachers, however, made no mention of official standards, indicating instead that they relied on their individual histories of evaluative decision-making and internalised or in-the-head standards to determine student progress. This omission reflects how, currently, teachers in Years 4 to 10 do not have available to them authorised, defined standards for use in judging student achievement. One of the implications of this situation seemed to be that while teachers said they were confident about the local appropriateness of their judgements, they reported being uneasy about how these related to teacher judgements of quality in other schools.

The teachers also indicated that they felt uneasy as they anticipated how successive groups of their students would manage test demands. The test situation was far from routine and students had to be readied, even rehearsed, for taking the test. The teachers' talk indicated a sense of vulnerability, even nervousness, as they worried about how their curriculum planning equipped students with the knowledge and skills necessary to satisfy test demands. The test represented the unknown, and as mentioned above, its purpose remained unclear for teacher and student alike. In effect, it was something imposed from above. For some of the teachers, it raised to the fore the issue of what represents normal, saying,

I'm worried that the people who do this test have a different idea of what is a normal Grade 5 than what I do. Well I mean everybody would be worried about that.

The issue of normalcy seemed to have less to do with demonstrated ability level than with cultural capital. Did the students have the cultural knowledge necessary to answer the questions? A Year 3 teacher reflected on this as follows:

When it comes to this [testing] I panicked. I thought, oh no, what if they're going to be asked to do something that is in the syllabus, but we haven't written down for [suburb X] children. So then we went through yesterday and looked at the things that they could be asked to do. And so I'm sort of going over it now with them in case it turns up next week in the test. So they're not going to panic and say 'I've never seen this before. I can't do it'. I mean I don't expect it to be that difficult. I expect that the test will say, you know, give a recount of something. Well they can do recounts with their eyes closed. We do it all the time. But it is a bit of a worry that they'll be asked to do something that [suburb X] children have not been taught how to do just because, maybe it's in the syllabus and we're planning on doing it in maybe Grade 4 or Grade 5 and other schools might have done it in Grade 3.

The contribution of cultural capital to test performance came home.
forcibly to me last week again, when a Year 5 teacher reported that the 1999 literacy test required students to answer questions about a reproduced Internet page. He said that the one and only computer in his classroom was not connected to the Net and most of his students did not have Net access in their homes. On the test day, he reported that many students looked puzzled when they saw the Internet page and asked him to explain it. His reply—'This is a test. Just see what sense you make of it by yourself.' This example calls into question the particular criteria that had been chosen and applied to this year’s literacy test in Queensland: how was it scrutinised for curriculum relevance and equity considerations? It also raises the issue of what is actually being assessed, given the teacher’s comment that the Internet text set for assessing reading was unfamiliar to many of his students. Additionally, while it is widely recognised that technological literacy is a vital concern in literacy education, we need to query whether a paper and pencil test with a static print text is the best means for assessing such literacy. Attention now turns to consider how the teachers talked about ‘context’ in classroom assessment and testing situations.

The significance of context(s)
The term ‘context’ frequently recurred in teachers’ talk about classroom assessment practices and testing programs. More specifically, their use of the term context was tied to understandings about teacher-student relationships and the ways in which those relationships are shaped by various conceptualisations of context. Of specific interest are, first, the concept of context as it applies to developing in students’ a sense of the authentic or real-life purpose(s) of activities; second, teachers’ understanding about students’ out-of-school contexts; and third, the context of the statewide test. Each of these will be considered separately.

Teaching context
The teachers talked of devoting considerable time and energy to ‘building pedagogical context’ in order to establish learning as purposeful. They talked of how purposeful learning hinged on students having an authentic cultural and social context for the activities they were required to undertake at school. Essentially, building context was, they said, foundational to good practice. One teacher spoke of this as it applies to teaching writing as follows:

You know, that’s the important thing, that when you ask them to write something you have to first of all create the context, you have to create a purpose for the writing, you have to create a sense of audience, you know, who they’re writing for and so on. If all that’s done well and you’ve done a lot of work in terms of preparing them for that, then they often write quite well. And with regard to spelling, again it depends on what, what the context that we’ve been using, you know, they might be able to spell ‘Nintendo’,
but not spell ‘crown’, you know, for example, because they’re words that they’re using more than others.

In this conversation we hear about the teacher’s understanding of writing pedagogy as being ‘first of all’ to ‘create the context’. We also hear how the teacher deliberately connects context with purpose and audience in her classroom practice in order to prepare students for writing, including spelling. She also indicates that these preliminary work has benefit—‘they often write quite well’.

Students’ life contexts
The teachers’ talk also brought to light their keen interest in connecting teaching, learning and assessment activities with the students’ outside-of-school experiences as a way of achieving relevance. Taking account of context then, was a matter of looking outside the school window and knowing about student characteristics relating to geographical location and cultural and linguistic backgrounds— their life contexts. According to the teachers, the authenticity and usefulness of assessment evidence generated in the classroom depended on the knowledge they developed of students’ life contexts as well as the teachers’ ability and willingness to design activities that were responsive to those contexts. A Year 5 teacher spoke of the need for responsive practice as follows:

Because students learn language in context ... and I believe that’s the way to teach language as well is to create some kind of social context. That social context is determined very much by the cultural context in which the children are, so the children in, let’s say you know, Camooweal or the children in an Aboriginal community in North Queensland, for example, have a very different cultural context from the children who are living in let’s say the Gold Coast or suburban Brisbane, you know, or different again from children who are living in the country, so the problem with the essential elements is how do you define those and at the same time take into account the different cultural social contexts in which children are learning language, and that’s why the new syllabus has not done that. It says that the context is derived from the children’s everyday living and school subject matter, so you can teach information reports in Grade 3, but whether you choose to teach them about the salt water crocodile or birds or spiders is determined by where you’re living and what’s around you.

In this extract we hear the teacher bringing into focus the need, in this case, for language activities to be anchored into cultural and social contexts, as well as the need for the teacher to relate classroom practice to ‘children’s everyday living’. We also hear the teacher’s understanding that the present Queensland English Syllabus for Years 1 to 10 (1994) authorises teachers to take account of outside-of-school contexts or what she refers to as ‘the different cultural social contexts in which children are learning language’.

121 The Australian Journal of Language and Literacy
The required shift in teacher-student relationship

In both of the above notions of context, there is a clear emphasis on the teacher and student in a partnership: the teacher is the master who deliberately and carefully scaffolds learning for the student apprentice, thereby inducting her into cultural knowledges. Further, we hear of how the teaching-learning relationship is foundational to how assessment happens in the classroom. In short, there is no clear demarcation between teaching and learning on the one hand, and on the other, assessment. On the designated test program days, however, at least for a few hours, sitting the test required a radical shift in the teacher-student relationship. For the teacher at least, there was the understanding that on test day, the paper was to be completed as far as possible in silence and as a solo performance. In the test situation, the teacher understood that she was to be detached—to keep an observable distance from student performance.

Teachers were well aware that the students needed to be readied for the test situation and, more importantly, for the inevitable change in teacher-student relationship that this brought with it. In the following extract, a Year 3 teacher talks of her work in preparing students for the test and relationship change.

So we’ll go through all the talk with it and getting them ready and discussing the topic and looking for words and things like that. And then I told them I can’t put it all up on the board for them. I can’t plan it. Because that’s the sort of thing we do, and they write from the board. So, what I’m trying to do is to get them to do that for themselves. So, you know, we put up these boxes and I say now you tell me the words, now you write it down. I’m trying to get them to do it themselves but they’re not used to it.

The final sentence ‘I’m trying to get them to do it themselves but they’re not used to it’ suggests the teacher’s sense of how the context of the test fundamentally alters how teacher and student ‘do’ school writing. And in the following extract, we hear a teacher who talks of how the routine teacher-student relationship has to change, given that the testing purpose is point-in-time measurement.

If you’re just getting students, if you’re just setting them a little writing task and saying OK, go and write it, then you’re not going to be doing all that scaffolding. But that’s a control situation, that’s a test situation, isn’t it? I mean, if your aim is to have some kind of national benchmark, and you’ve got to provide the same conditions, then you either, I mean how do you say ‘Teachers you can do this much scaffolding and not this much’, you know? Either they do none, but you know, scaffolding is part of the teaching. This is not a teaching, this is a testing thing, so you live with that.

The teacher’s distinction in the final sentence—‘This is not a teaching, this is a testing thing’—is vital in coming to understand the current
relationship between classroom assessments and the testing program. Essentially, it comes down to the distinction between formative and summative assessment, where teachers saw their assessment practices as primarily formative in nature, being concerned with diagnosis and improvement, rather than with grading or the awarding of summative grades. The teachers talked of themselves as key participants, with students, in their classroom-based assessments, designing and teaching the assessment activities, working with students to enable them to complete the activities successfully.

What counts as valued assessment evidence?
The fact is that the teachers are well placed to make direct first-hand observations of and judgements about student achievement. They are well placed to provide valid assessment evidence in the form of portfolios containing evidence collected over time. However, they report a lack of confidence in the reliability of those judgements. The following extract shows how their uncertainty can be traced directly to at least two main sources.

Do I have a clear picture of student achievement? Probably not, no, I'd have to say I really, I don't feel as if I really have got a really good grasp. I mean I have got gut feelings, because you know, I know my students very well and I believe I know what they can and can't do, but I guess it's a little bit of that feeling of that unless it's objective and unless it's formalised, you know, a real formal collection of data, you don't really trust it. If I go to do something that's an assessment activity, often it will turn into, because I scaffold a lot, because I say well, you know, because I can see that there are students having problems with it, I can't let them just sit there feeling as if they can't do something and failing, so therefore I intervene and they come up to me and they say 'I can't do this' and I say 'All right, who else is having trouble? Come over here and we'll do it together,' and before I know it I've taught it, so I don't have ways ... I mean I have ways, but often the opportunities that I have, the opportunities that I create from collecting hard data on the students' abilities, I lose it because I've scaffolded too much, I've started teaching and before I know it I don't really know whether they can do it or not.

I mean I know, for example, that they don't know this stuff or they can't do it, so it's really I guess my teacher judgement that is probably more accurate in that sense, and I guess that's where sometimes as teachers we don't trust that or we forget to formalise, we forget to write that down, you know.

In this extract, we hear the teacher talking about the competing demands she faces as she juggles different assessment purposes. While she may set out to generate what she refers to as hard data, initially wanting to stay removed from the assessment to see what students can do, alone and unaided, it seems the students actively call her in to participate in the assessment activity. What emerges is a clear picture of the
students working to maintain the teaching-learning-assessment connection in routine classroom practice, even on those occasions when the teacher planned to generate hard data for measurement purposes. Also in this extract, we hear of the teacher’s need for clearly defined assessment standards. It is hardly surprising that teachers report a lack of confidence in their judgements when, currently, they do not have access to explicitly defined year level standards to scrutinise and defend those judgements. Until such standards are developed and implemented, and teacher judgement is targeted as a professional development priority, the danger is that the statewide testing programs and the results that they generate may be viewed as having an authority greater than that attributed to teachers’ classroom assessments, even by the teachers themselves. Already there are observable, mounting fears about the use of test results as reliable evidence of teacher effectiveness. One teacher revealed her fears as follows:

Probably what worries me most is that if the results are low or the benchmarks are not met by a certain school, is it then considered that that school (1) is no good, and (2), the teachers at the school are no good? I think that’s really hard and there’s teachers who work in schools where there are quite a few difficulties for children with reading and writing and numeracy. Well where do we stand, you know? I think you’re going to get other schools where the parents are supporting them—we don’t get any of that here ... You feel like you’re hitting your head against a brick wall. And the other thing here that we find difficult is that we have such a transient population as well, so you think you’re going along nicely and then the next minute the whole class changes, or half your class changes. So you just start all over again and there are your percentages dropping down, and that’s no reflection on the school or the teachers. That’s my big concern, it really is. Testing and benchmarks have got to be handled carefully and this thing that was in the paper about, you know, being divulged to the public. OK, if all the facts are added to it, but you know, why? Because, I mean I could sit there and say oh, wasn’t I the best teacher last year, look at all my Grade 3s. Then the year before I’d be going, I’d be the worst teacher in the school because I had all these low kids. I mean, you can really get yourself your self-esteem knocked at a place like here, I think.

This extract shows the teacher’s attempt at exoneration—the school is exonerated in the talk from major responsibility for low literacy achievement—and the allocation of blame. We hear the blame focused on student deficits, parental involvement and transfers—all factors over which teachers have no control and yet they remain, according to the teachers, powerful influences on learning outcomes. There is also a clear voicing of vulnerability—‘you can really get yourself your self-esteem knocked at a place like here’—accompanied by a fear of being classified publicly as a failing school/failing teacher.
Outlining the challenges in the search for congruence in assessment

Implicit in much of the preceding analysis is the need to improve ways of assessing literacy, both in classroom-based assessments and in large-scale, standardised testing programs. This section outlines five main challenges that we need to address if we are to achieve congruence in our ways of assessing literacy in schooling.

First and foremost, the education community needs to be clear about the purposes of various assessment programs, and how the various programs relate in terms of purpose, one to the other. If statewide testing programs are to have a genuine purpose of improving outcomes, as distinct from reporting outcomes, then we need to reach agreement that the teacher, not the test, is the primary change agent. If we agree on this, then we must bring teacher judgement to centre stage. The point is that teacher judgement is central to a much-needed review and discussion of all performance evidence, including that generated in standardised testing and in classroom-based programs. The challenge is to confirm the consonance of the evidence or to identify outlying aspects of the students' performances. In short, teacher judgement can be used effectively to interrogate the links between the school assessment program and the evidence it generates and the evidence generated in the test program.

Essentially, we need to map the nature and scope of the evidence generated in the standardised testing program, and the nature and scope of the evidence generated in classroom-based assessment programs. We need to know how the two interface, and what they tell us, separately and together, so that we know the whole assessment story. What is the nature of the knowledge and skills assessed in the two programs? And what definitions of literacy and numeracy are informing the programs? If we keep the programs separate at policy level, we run the risk of going down the pathway leading to test results being used to announce so-called beacon and failing schools.

Second, we need to re-value teacher judgement, on the understanding that it lies ‘at the heart of good teaching and good assessment’ (Sadler 1986, p. 6). Currently however, there is an urgent need to invest in teacher judgement, training it up through professional development programs focussing sharply on assessment, and through system support mechanisms including those provided through internal and external moderation networks. The extracts considered earlier show how teachers face competing demands as they struggle to distinguish formative and summative assessment purposes. Teacher professional development programs are needed to assist teachers to distinguish between assessment with teaching-learning significance and assessment with measurement significance, showing how they are best placed to do both in their classroom practice.

Third, there is an urgent need to make explicit performance expecta-
tions for literacy education: what is it that we expect students to be able to know and do? There is an urgent need for anchoring teacher judgement into standards, written as verbal descriptors of outcomes with accompanying exemplars that make clear how each exemplar matches the characteristics of the stated standard. Currently, while teachers in the early years can look to the Continua for some standards advice, teachers in Years 5 to 10 make judgements about student literacy and numeracy in the absence of a clearly defined standards framework.

Fourth, we need to use the testing programs themselves as a professional development opportunity. This can be achieved by feeding back to schools not only quantitative reports, but also reports about the features of good assessment task design at various year levels. Also, we need to make available for teachers information about the scoring guides applied to student scripts in the testing programs and the training given to the assessors.

Finally, we need to be mindful that all assessment activities are contextualised and value-laden. There is no such thing as value-free assessment! The setting of cut scores in national equating exercises is included here. Dwyer (1998) made this point, writing that ‘any use of a cut point, no matter how sophisticated or elaborate its technical apparatus, is at heart a values decision. The underlying question in setting any cut score can be phrased quite simply: “How much is enough?” There is, of course, no technical answer to that question; there is always a value answer to it’ (p. 18). So, in informing the much-needed debate about how assessment can be re-theorised to take account of diversity, we should know more about how test scores are actually treated with respect to cut scores and who takes the responsibility for and acts on these decisions in institutional uses.

In conclusion, I am mindful of the divergent priorities and goals of key education stakeholders in Australia, and aware of the pressure on some to follow short-term political imperatives of appearing to be delivering improved results. The challenge for the educational community is to ward off this pressure, focussing instead on providing support for the long-term professional development change necessary to effect actual pedagogical change and improved outcomes. If the aim of standardised testing, measurement and reporting, as proposed in the National Plan, is to secure literacy for all Australians, then teachers must be key players in instigating, developing, implementing and reviewing systems of assessment reform. As all teachers know only too well, assessment procedures, of themselves, do not necessarily lead to improvement.

Acknowledgments
I wish to acknowledge the helpful feedback and assistance provided by Ms Jill Ryan and Ms Stephanie Gunn in the preparation of this paper.
References
Department of Education, Qld, 1994, English Syllabus for Years 1 to 10. Brisbane.
Education Department of Western Australia 1994, First Steps: Reading: Developmental Continuum, Longman, Melbourne.