

TEACHER JUDGMENT:
BUILDING AN EVIDENTIARY
BASE FOR QUALITY
LITERACY AND NUMERACY
EDUCATION

JJOY CUMMING

CLAIRE WYATT-SMITH

JOHN ELKINS

MARY NEVILLE

Final Report

© **Griffith University, 2006.**

This work has been produced by the Centre for Applied Language, Literacy and Communication Studies and Centre for Learning Research, Griffith University with the assistance of funding provided by the Queensland Studies Authority.

The views and opinions expressed in this document are those of the author/project team and do not necessarily reflect the views of the Queensland Studies Authority.

ISBN 1 920952 71 3 print edition

Published by Griffith University

ABN 78 106 094 461

Contents

Contents	3
Tables	5
Executive Summary	6
Summary of Recommendations	9
Introduction: Project Aims, Research Questions and Policy Connections	11
Review of Literature: Assessment, literacy and numeracy in current policy and school contexts	13
National and International Trends in Educational Accountability and external testing in Australia	13
Effective Assessment and Student Learning	15
Literacy frameworks and instruction in Queensland	17
Numeracy frameworks and instruction in Queensland	18
Schools, teachers, classroom assessment and external test data	21
Professional development and the use of data for learning improvement	25
Conclusion	28
Design of Project	29
Identification of case study schools	29
Identification of teachers and students	30
Ethics	32
Limitations of the study	32
Findings	33
How teachers in Years 3 and 5 currently assess the broad range of literacy and numeracy outcomes identified in Queensland English and Mathematics syllabuses	33
Comparison of teachers' judgment of students' literacy and numeracy achievement with the results obtained from the <i>Year 3 and Year 5 Aspects of Literacy and Numeracy Tests</i> program and similarities and differences	46
How teacher assessments can be used to provide additional information on students' performance in literacy and numeracy.	65
Interpretation and use of the <i>Year 3 and Year 5 Aspects of Literacy and Numeracy Tests</i> data and information by teachers, parents, schools and systems.	68
Issues for students identified as at risk or in danger of being at risk by either teacher judgment or the <i>Year 3 and Year 5 Aspects of Literacy and Numeracy Tests</i> data or both.	79
Standardisation of task and/or moderation processes needed to be developed to assist comparability of teacher judgments	84
Processes or resources to assist teachers, parents, schools, and systems to make more effective use of information provided by the <i>Aspects of Literacy and Numeracy Tests</i> and to assist in teacher judgment of literacy and numeracy achievement	87

Impact of federal tests of literacy and numeracy on discussion and recommendations	96
Conclusion	98
References	100
Appendices	103
Acknowledgements	104

Tables

Table 1: Mathematics Core Learning Outcomes: Level Statements (QSA, 2004)	19
Table 2: Characteristics of Case Study	29
Table 3: Overview of the three strands	47
Table 4: Teacher Descriptions – Continuum of descriptions from high to low achieving performers	55

Executive Summary

This project investigated teachers' assessment practices in literacy and numeracy in Years 3 to 6 in seven case study schools in Queensland. While care is needed in interpreting the outcomes of the project as representative of all teachers' practices in Queensland, the cross sector sample of case study schools did constitute a range of educational contexts of practice with diversity of school size, geographic location and cultural and socio-economic student backgrounds.

The purpose of the project was to examine teachers' assessment practices, compare these with outcomes of the *Year 3 and Year 5 Aspects of Literacy and Numeracy Tests* conducted by the Queensland Studies Authority in accord with mandated external accountability requirements, consider ways in which the outcomes were congruent and differed, and ways in which they could be used as complementary to support both improved learning outcomes for students and systemic data collection. As part of the project, schools' use of the Tests data to plan instruction and improve student learning was investigated.

The project investigated teacher descriptors of achievement in literacy and numeracy and Tests data for focus students identified by teachers as 'above' level, 'at' level, and 'below' level for Years 3 and 5 in 2005. Teacher assessments, and use of the students' Tests results, were investigated in their subsequent Years in 2006. The project did not prespecify either definitions of literacy and numeracy, or definitions of 'above', 'at' and 'below' level. This approach was consistent with the project's intent in examining how teachers arrived at judgments of quality. The major data collection methodology was interviews with teachers and other school personnel and collation of work samples for students. Through the interviews, the frameworks of literacy and numeracy that teachers were using were elicited, as were their conceptualisations of 'levels' of student achievement.

To provide the theoretical and policy context for the project, a brief Review of Literature is presented in the report. This Review considers the national goals for student learning for the 21st century established through the *Hobart and Adelaide Declarations*, including the goals that 'every child shall be literate and numerate'. While these goals establish the basis for the national reportability and accountability focus underpinning the *Year 3 and Year 5 Aspects of Literacy and Numeracy Tests*, we also note in the Review that the original framework of the Declarations was much broader, including goals that students will be technologically-sophisticated, problem-solvers, independent workers and that their education will promote their sense of well-being and self-esteem, especially as learners.

The Review also considers the literacy and numeracy frameworks that are established through policy and curriculum in Queensland. Most notably, these frameworks promote recognition that literacy and numeracy are complex and contextual and that specific curriculum literacies and numeracies are essential for successful outcomes in discipline domains. Recent research on the significant role of assessment for successful student learning is discussed showing the power of effective teacher assessment and feedback for students in directing their learning. The Review also discusses the willingness or reluctance of, and capacity for, schools and teachers to engage with external data to improve planning and instruction. While most of this research is drawn from England and the USA, Australian evidence supporting the findings is present. Effective professional development processes to enhance school and teacher understanding and use of external data are examined. The Review explores the research on the congruence of teacher assessments with external measures of student achievement, noting that such research has often

occurred in educational environments where the external measure is seen as the ‘objective’ and true yardstick against which teacher assessment standards should be measured. The Queensland education context, with its long history of and internationally-renowned respect for teacher assessments even for high-stakes purposes, frames teacher judgment in a different light. Teacher assessment in Queensland is regarded as providing opportunities for deep, meaningful and contextual learning for students. Finally, the Review examines briefly the dimensions teachers bring to their assessments, dimensions that from one perspective may be criticised as outside the official goals of curriculum learning, but on the other hand are seen as supportive of the broader national goals of schooling in Australia and elsewhere.

The project findings led to 36 Recommendations, summarised in the next section. Overall, the outcomes show that:

- Teachers used diverse frameworks in assessing student literacy and numeracy
- These frameworks were aligned with English and mathematics and tended to focus on reading, writing and mathematics content. Oracy and oral reading fluency were noted by some teachers.
- The frameworks did not include curriculum literacies and numeracies or technological applications.
- For literacy, links to the English syllabus were not clear; for numeracy, links to the Mathematics syllabus were stronger, particularly through a commonly-used learning package aligned with the syllabus.
- Frameworks used by teachers did include broader dimensions of student learning including strategic processes and engagement with learning.
- The standards used by teachers to describe student performance drew on the above frameworks that they used and were not clearly aligned to levels in the syllabuses. Much of the commentary provided by teachers reflected point-in-time judgments based on observation, as well as student work samples. There was not a sense that the various sources of information collected by teachers was synthesised in a coherent whole.
- Teachers in many schools reported use of a range of external standardised measures to gauge student achievement in comparison to norms and criteria. While some schools engaged in moderation practices, teachers considered the provision of more exemplars of student achievement at different levels and assistance with moderation across schools desirable.
- Teacher judgments of student levels were broadly consistent with the outcomes for individual students on the *Year 3 and Year 5 Aspects of Literacy and Numeracy Tests*. Teachers also indicated that where there were divergences, the teachers considered their own judgment to have more substantial basis.
- Teachers considered similar dimensions of literacy and numeracy to those measured by the Tests, although it is not possible to determine whether these judgments and the teacher discussions were influenced by the project focus. The dimensions of focus of both the teachers and Tests are narrow in comparison to broader policy and official curriculum frameworks for literacy and numeracy in Queensland. There may be concern that the Tests are already having an impact on narrowing the curriculum and student learning experiences in Queensland classrooms.
- It was noted that the documentation around the Tests is unclear in terms of purpose and what is being assessed. It is recommended that the documentation should be reviewed for clarity and consistency.

- Schools and teachers reported limited use of the Tests data for a number of reasons including usability and access. Systematic planning that did occur related more to preparation of cohorts for future testing in areas identified as demonstrating poor performance. Some planning for individual students did occur. This was mainly through the activities of learning support teachers. One factor that continues to affect usability is the timing of the testing and the provision of data.
- Teachers were able to identify students who were at risk and especially those whose achievement was reported as 'below benchmark'. However teachers considered many students at risk who were reported to have achieved the 'benchmark'. The work samples confirmed teachers' concerns. Teachers considered that the level of satisfactory achievement indicated by the benchmarks were too low and that further, they could distract attention and resources from children at risk.
- The Project Recommendations address the need to:
 - Develop stronger frameworks of literacy and numeracy, statements of standards for levels, and statements of curriculum literacies and numeracies to guide teachers in their assessment practices.
 - Undertake further professional development and research activities that assist teachers in these areas and in developing clearer profiles of student achievement from available evidence.
 - Provide a framework for the assessment evidence and profiles collected by teachers to feedforward into systemic data collection, recognising the breadth of teacher knowledge about students and the capacity of teachers to identify students at risk.
 - Undertake further research to examine the impact of the Tests on the learning experiences of children in Queensland classrooms.
 - Undertake further research into the standard of achievement ascribed by the Tests to the benchmarks and teachers' considerations that such standards are too low to identify all students at risk.
 - Provide professional development to school leaders and teachers in using the external Tests data to support student learning and plan curriculum, and the appointment of a key person in a school or supportive to a school as a 'data guru'.
 - Provide the Tests data in a usable form with assisting software.
 - Consider the implications of the findings for current Queensland Curriculum and Reporting projects.

Summary of Recommendations

Literacy and Numeracy Assessment Frameworks

It is recommended that the *Literacy Framework for Action 2006-2008* (Education Queensland, 2006b) should underpin future projects and professional development in schools on literacy achievement. Similarly, recent Queensland numeracy policy initiatives should underpin projects and development in schools on numeracy achievement.

- ◆ Recommendation 1: pp 44

Literacy and Numeracy Assessment PD

It is recommended that professional development should be undertaken with teachers and schools to provide guidance in literacy and numeracy assessment and documentation, addressing the matters raised in the report.

- ◆ Recommendation 36: p. 95

It is recommended that descriptive elaborations of standards, siting the benchmark performance within such standards, should be developed to guide teacher assessment, with opportunities for teachers to engage through some process of moderation and sharing of exemplars.

- ◆ Recommendations 2: pp. 44-45
- ◆ Recommendations 3-5: p. 45

It is recommended that a website resource providing advice on commercially available standardised tests, and their use, should be developed by the QSA.

- ◆ Recommendations 32-33: p. 92

Data Use PD

It is recommended that considerable statewide professional development on the interpretation and use of Tests data should be undertaken.

- ◆ Recommendations 12-13: p. 78

It is recommended that the Tests data should be provided to schools in electronic forms with accompanying software to assist in data interrogation.

- ◆ Recommendation 14: p. 78

It is recommended that schools should be able to appoint a 'data' guru, or that such a support person should be provided to clusters of schools, to assist in effective interpretation and use of Tests data.

- ◆ Recommendations 15-17: p. 78
- ◆ Recommendation 34: p. 93

As noted, professional development on interpretation and use of Tests data is a recommended priority.

- ◆ Recommendations 23-24: p. 88
- ◆ Recommendations 25-27: p.89
- ◆ Recommendation 28: p. 90
- ◆ Recommendations 30-31: p. 91

- ◆ Recommendation 33: p. 92

Queensland Curriculum, Assessment and Reporting (QCAR) Project

It is recommended that current initiatives such as the statements of Essentials and Standards, and other QCAR projects, be reviewed in light of the findings of this project.

- ◆ Recommendation 18: p. 83
- ◆ Recommendations 19-22: p. 86

Clarification of Documentation

It is recommended that the documentation provided to support the *Aspects of Literacy and Numeracy Tests* should be reviewed and clarified.

- ◆ Recommendations 6-8: p. 52

Systematic Tracking of Student Achievement

It is recommended that all school sectors should explore the potential for schools to systematically track student achievement longitudinally by using a student-specific ID and create a longitudinal database.

It is recommended that the QSA, in conjunction with system authorities, develop an annotated bibliography about external standardised tests, including information on purpose, norm development and appropriateness for Queensland curriculum and students to be hosted on the QSA website.

- ◆ Recommendations 28-29: p. 90

Advice to Parents

It is recommended that feedback to parents should include meaningful descriptions of students' standards and levels in relatively simple English with links to the relevant curriculum.

- ◆ Recommendation 35: p. 93

Further Research

It is recommended that a trial project to assist teachers to construct structured portfolios of student achievement, should be undertaken

Research on Impact of Tests on Classroom Practice

It is recommended that the reported disparity between the identified benchmark standard identified through the Tests and teachers' considerations of the minimal achievements of students at risk should be further investigated. Mechanisms for teacher judgment to provide further systemic advice on students at risk should be investigated.

- ◆ Recommendation 9: p. 63
- ◆ Recommendations 10-11: p.67

Introduction: Project Aims, Research Questions and Policy Connections

This project was funded by the Queensland Studies Authority (QSA) as part of a research program

to provide evidential and analytical support for policy initiation, development and review in relation to QSA's functions ... includ(ing) syllabus accreditation and development, testing, assessment, moderation. (QSA Project Brief, 2005)

The project aims were to explore

the capacity of teachers to make judgments against defined standards such as the national benchmarks, and whether their judgments can provide more valid (comprehensive) and reliable (stable) information on student capabilities

and to address

the question of best use of standardised-test data along with school-based data. (QSA Project Brief, 2005)

Hence, the project provides an investigation of the nature of teacher judgment of student literacy and numeracy performance and consistency between teacher judgment of student literacy and numeracy performance and student achievement on the external *Year 3 and Year 5 Aspects of Literacy and Numeracy Tests* (hereinafter referred to as the Tests). It also explores the degree to which teachers and schools use the external test data to inform individual student instruction. In so doing, the project examines policy implications for effective practices on the use of data for learning improvement, and the potential for teacher judgment to inform systemic data on literacy and numeracy performance.

The research questions identified in the study, and addressed in the following report are:

How do teachers in Years 3 and 5 currently assess the broad range of literacy and numeracy outcomes identified in Queensland English and Mathematics syllabuses?

How do these assessments compare with the results obtained from the *Year 3 and 5 Aspects of Literacy and Numeracy Tests* program? What is the nature of any similarities and differences?

How could such teacher assessments be used to provide additional information on students' performance in literacy and numeracy?

How do teachers, parents and schools interpret and make use of the Year 3 and Year 5 literacy and numeracy test data?

What particular matters arise for students identified as at risk or in danger of being at risk by either teacher judgment or external test data or both?

What standardisation of task and/or moderation processes may need to be developed to assist comparability of teacher judgments?

What processes or resources might assist teachers, parents and schools, if necessary, to make more effective use of such information?

As noted, the project was designed not only to improve practice through enhanced understanding of schools' and teachers' use of assessment data of varying forms, but also to inform policy development by the QSA. Within the above research questions, the project needed to: identify ways in which the QSA would be able to strengthen the capacity of teachers for school-based assessment of students' literacy and numeracy skills; support schools and teachers in making strategic use of test data for teaching and assessing their students; and inform materials development to support appropriate use of test data by schools and teachers (QSA Brief, 2005).

In order to address these research questions, the project team undertook seven case studies over twelve months, collecting information from schools and teachers, and student work, at key times. The design of the study and nature of information collected are described in Chapter 3.

Contextualising the project and findings

The following chapter provides a brief review of research and policy in the areas of teacher judgment, standardised tests and literacy and numeracy. It considers the national policies that have informed both the current focus on literacy and numeracy and the external testing requirements and, in particular, the national goals that all children should be literate and numerate. It looks at the role of teacher assessment of student achievement and the contribution of effective assessment to learning; research on the consistency of teacher judgment with external standardised test outcomes; expanded dimensions of teacher assessment; and finally, issues related to effective use of data to improve learning, including the design of appropriate professional development.

Review of Literature: Assessment, literacy and numeracy in current policy and school contexts

Recent Australian and international education reforms have been shaped by the need for governments to continue to deliver quality education in the context of high rates of technological advances, globalisation, and new economic times (MCEETYA, 1999; Queensland Government, 2005; U.S. Department of Labour, 1991). In response to these new complex environments, many countries including Australia have developed futures-oriented policies to deliver educational outcomes that match the capabilities considered necessary for citizens to engage effectively in a world of rapid change. Policy development has made assessment of student performance important to monitor the effectiveness of schooling to promote student learning in the 21st century.

National and International Trends in Educational Accountability and external testing in Australia

Accountability in education is a growing agenda throughout Australia and overseas. This agenda has emerged in part through the perception of many countries that public accountability for education expenditure should be more transparent, and that educational standards can be improved through external monitoring and resultant policy actions. An international trend exists towards the development of national standards upon which to measure and report on student performance to a wide variety of stakeholders including parents, school communities, industries and the public. In countries such as the United States of America (USA) and England the perceived need for educational accountability has witnessed a progressive increase in high stakes and standardised testing of student cohorts. Whetton (2004) suggests such testing in England is mandated for a number of reasons: (a) to increase standards and place more students in university; (b) as a source of reliable public information about academic performance of students and schools; (c) parents have better access to information regarding comparable school achievements; and (d) testing can be used as a favourable instrument for curriculum change.

In Australia over the last two decades, one of the most public educational reform agendas has been the press for improvement in and accountability for Australian students' learning outcomes in literacy and numeracy. National goals and standards were developed in Australia through two significant agreements between the federal Minister and all state and territory Ministers for Education: the *Hobart Declaration on Schooling* (1989) (MCEETYA, 1989) and the *Adelaide Declaration on National Goals for the Twenty-First Century* (MCEETYA, 1999).

The common and agreed goals in the *Hobart Declaration* included:

To develop in students:

the skills of English literacy, including skills in listening, speaking, reading and writing;

skills of numeracy, and other mathematical skills;

and to provide, for the first time, a National Report on Schooling in Australia including information on student achievements (MCEETYA, 1989).

The Hobart Declaration goals were reviewed by MCEETYA in 1997, with particular focus on setting goals and targets, and with Australians invited to comment on a discussion document. The more elaborated draft goals stated that all students should have:

attained the skills of numeracy and English literacy; in particular, every child leaving primary school should be numerate, able to read, write, spell and communicate at an appropriate level. (MCEETYA, 1998)

These specific goals were part of much broader reform goals that

[S]chooling should develop fully the talents and capacities of every student. In particular, when students leave school they should:

- have skills in analysis and problem solving and the ability to become confident and technologically competent members of 21st century society
- have qualities of self-confidence, optimism, high self-esteem, and a commitment to personal excellence as a basis for their potential life roles as family, community and workforce members
- be active and informed citizens with the ability to exercise judgment and responsibility in matters of morality, ethics and social justice; and the capacity to make sense of their world, to think about how things got to be the way they are, to make rational and informed decisions about their own lives and to collaborate with others
- have a foundation for, and positive attitudes towards, vocational education and training, further education, employment and life-long learning. (MCEETYA, 1998)

The 1999 *Adelaide Declaration* endorsed the goals in the Discussion Paper, and the broader context of learners ready for the 21st century and the goal that

every child leaving the primary school should be numerate, and be able to read, write and spell at an appropriate level. (DETYA, 1998, p. 9)

In 1997, MCEETYA had also adopted a new sub-goal—‘that every child commencing school from 1998 will achieve a minimum acceptable literacy and numeracy standard within four years (recognising that a very small percentage of students suffer from severe educational disabilities)’, and had ‘agreed to begin reporting against the sub-goal for each year from 1998’ (MCEETYA, 1997).

To support the achievement of these goals, the Ministers endorsed a *National Literacy and Numeracy Plan* which included

- comprehensive assessment of all students as early as possible, to identify those students at risk of not making adequate progress towards the national numeracy and literacy goals;
- intervening as early as possible to address the needs of students identified as at risk;
- the development of agreed national benchmarks in literacy and numeracy, against which all students’ achievement in these years can be measured;
- assessment of students against the national benchmarks using rigorous state-based assessment procedures;
- progress towards national reporting by systems on student achievement against the benchmarks; and
- professional development for teachers to support the key elements of the National Plan. (DEST, undated)

The development of the national benchmarks for literacy and numeracy was then undertaken by Curriculum Corporation with overview by MCEETYA and public and expert consultation.

This brief overview shows the evolution of the external testing of literacy and numeracy at the state level, with national reporting, from broad national agreement about comprehensive goals to achieve multidimensional learners for the 21st century, a more specific focus on literacy and numeracy. Instigating these moves was the expectation that all students will achieve minimum standards in the early years of schooling, standards described through the National Benchmarks and operationalised through the state literacy and numeracy tests.

In Queensland, the *Years 3, 5 and 7 Aspects of Literacy and Numeracy Tests* are generally undertaken by all schools and all sectors. All sectors are required to demonstrate student achievement for state and national reporting in order to receive funding. However, while one public discourse about the national benchmarks has been about accountability, the major official policy driver has emphasised accountability to improve student learning in the context of the complexity of learning needed for the 21st century.

The Government wants to see improved accountability. Educational accountability should be undertaken co-operatively, not imposed from above, and in ways which collect information of real use to schools, teachers and parents as well as governments. As a nation we need to determine the extent of improvement in broad terms by collecting data about how students are accessing schooling, the ways they are participating in it, where they are going after they leave school and how they are achieving at school, especially in key areas such as literacy and numeracy. (DEETYA, 1998)

Further, for the first time in 2006 schools are required to publish on their school websites, results from the Tests.

Effective Assessment and Student Learning

Teacher judgment of student learning

In Queensland, schools presently rely on a range of assessment processes, both formative and summative, to make judgments of student learning in a variety of assessment conditions and contexts throughout the years of schooling. Most assessment is developed and monitored by teachers, curriculum leaders and administrators at the school level and is supplemented in some years of schooling by statewide school based assessments in the Senior years and standardised testing of student performance in the early and middle years. The Year 2 Diagnostic Net for reading, writing and number and the system of senior secondary assessment are both instances of the value placed on school-based teacher judgment of student performance. Both are validated by other qualified teachers against explicit phase level indicators (Year 2 Net) or specified senior schooling syllabus standards in external moderation processes. The external standardised tests of literacy and numeracy in Years 3, 5 and 7 provide schools with snapshots of student performance against national benchmarks (QSA, 2006b).

Recent research and policy initiatives in Queensland have been exploring new processes for extensions of school based but statewide validated assessment at a number of points of learning between the Year 2 Diagnostic Net and the senior years. These have included 'Rich Tasks' as part of the New Basics research project (see, for example, Education Queensland, 2006a). A current initiative is the Queensland Curriculum, Assessment and Reporting (QCAR) project investigating the development of statewide validated assessment including statements of essential learnings and standards for assessing and reporting student performance. The focus of this aspect of QCAR is on Years 4, 6 and 9 (Queensland Government, 2005).

The system of school based assessment and moderation in the senior years of schooling in Queensland has gained international recognition (Broadfoot, 1999). It has long built on the seminal work of Sadler (1989) linking formative assessment to students working with standards

and the need to make explicit assessment expectations. While the senior system focuses on summative assessment, the underlying emphasis has been the need for students to understand the nature of the work that they are required to complete, to be given guidelines about how the work will be assessed, and to be given feedback in relation to such expectations that is of assistance to improvement. Underlying the system are two assumptions: that teachers as professionals are able to make appropriate judgments about students' work, and moreover, that teachers (and students) are best placed to make judgments about students' work, and to provide full information on student performance in a range of contexts and through a range of assessment opportunities. It is this aspect of assessment, placing high stakes judgment in the hands of teachers, that has gained international recognition for Queensland. Such recognition has increased in more recent times, with the growing international awareness that effective assessment plays a significant role in facilitating student learning, while, conversely, poor or narrowly focused assessment can inhibit effective student learning. Recognition of the power of assessment goes beyond the current accountability focus for public reporting that is occurring worldwide, with its emphasis on external summative testing, and indeed may run counter to the direction the accountability agenda has taken learning and assessment.

A noted comprehensive review of effective formative assessment practices by Black and Wiliam (1998) found that effective formative assessment practices that incorporated explicit feedback to students was decidedly instrumental in improving student learning outcomes—international recognition of the importance of the role of effective assessment practice in the hands of teachers to improve student learning across a range of valued dimensions. Other research being undertaken by members of the project team that focuses on capacity building of teacher assessment capabilities provide further examples of intense interest around formative assessment for learning improvement.

The research evidence on effective assessment practices to improve student learning outcomes provides teachers with promising possibilities to lead quality school based assessment practices (Black & Wiliam, 1998). For example, the Assessment Reform Group (ARG) (1999) provides explicit research based guidelines that they claim are 'deceptively simple' for improving learning through assessment. These include:

- The provision of effective feedback to pupils;
- ◆ The active involvement of pupils in their own learning;
- ◆ Adjusting teaching to take account of the results of assessment;
- ◆ A recognition of the profound influence assessment has on the motivation and self esteem of pupils; both of which are crucial influences on learning;
- ◆ The need for pupils to be able to assess themselves and understand how to improve. (ARG, 1999, pp. 4-5)

Kellis and Silvernail (2002) have also endorsed teacher judgment of student progress for three main reasons: firstly, it provides immediate feedback; secondly teachers are able to use more recent information on student learning to make instructional decisions rather than waiting months for test scores; and, thirdly, owing to the unobtrusive nature of teacher judgment techniques, such as daily observation, teachers have the opportunity to make deeper judgments of student learning that go beyond fact based, short answer responses.

This review focuses on what is known about improving learning through assessment, the roles of external test information both to report on student achievement and for the improvement of student learning, the way teachers make judgments and the way teachers are able to use information from a variety of sources to improve student learning, and the contribution of teacher generated classroom based assessment and the relationship between the two. These are discussed in general but also within the context of literacy and numeracy assessment—the focus of this project.

Literacy frameworks and instruction in Queensland

At the time of this study the *Literacy Framework for Action 2006-2008* (Education Queensland, 2006b) was released. It seeks to focus teacher attention on the relationship of literacy to English and other key learning areas. Policy and strategic initiatives on literacy instruction in Queensland were underpinned by the previous policy *Literate Futures: Report of the Review of Literacy in Queensland State Schools* (Luke & Freebody, 2000). Central to both the Literate Futures initiative and the QSA's Literacy Position Paper (QSA, 2001a) is the analytical framework of literacy needs by Freebody and Luke, known as the Four Resources Model (Freebody & Luke, 1990), a guiding framework or heuristic for identifying and planning for the resources necessary to be an effective reader. This set of practices is situated within the broad definition of literacy found in Education Queensland policy documents and the Years 1-10 English syllabus:

Literacy is the flexible and sustainable mastery of a repertoire of practices with the texts of traditional and new communications technologies via spoken language, print, and multimedia. (Luke & Freebody, 2000, p.9, QSA, 2005b, p. 2)

The current English Years 1-10 syllabus (QSA, 2005b) draws on the work of Green (1998) and is organised in strands that elaborate on the cultural, operational and critical study of language and texts. Contributing to these dimensions is a broad history of approaches to literacy and language instruction including skills based approaches (operational strand), whole language, genre and cultural heritage approaches (cultural) and other critical-cultural approaches (critical dimension). The current Queensland syllabuses are designed around outcome levels, Levels 1 to 6 that depict the expected developmental path of student achievement across Years 1 to 10. Level 2 statements for reading and viewing and writing and shaping provide descriptions of the practices that a Year 3 student at Level 2 in English could be expected to demonstrate:

Reading and Viewing

Students interpret and construct texts with familiar cultural purposes and familiar subject matter, making connections between directly stated ideas and information. They interpret basic generic structure, simple and compound sentences, patterns of short noun and verb groups, dominant images, font choices, gestures and facial expressions to make meaning of texts. They identify, explain and suggest alternative choices in the ways people, places, events and things are represented in texts. (QSA, 2005b, p. 30)

Writing and Shaping

Students interpret and construct texts with familiar cultural purposes and familiar subject matter, making connections between directly stated ideas and information. They interpret basic generic structure, simple and compound sentences, patterns of short noun and verb groups, dominant images, font choices, gestures and facial expressions to make meaning of texts. They identify, explain and suggest alternative choices in the ways people, places, events and things are represented in texts (QSA, 2005b, p. 32)

Similarly, the Level 3 statements below broadly describe what an 'at level' Year 5 student might be expected to be able to use in two of the sub strands of English:

Reading and Viewing

Students interpret and construct texts considering text type and commonly associated purpose of texts with personally significant and some unfamiliar subject matter, making connections between directly stated information and main ideas. They draw on patterns of textual resources, commonly associated with text types that are used to organise and link ideas and information. They identify positive and negative textual representations. (QSA, 2005b, p. 36)

Writing and Shaping

Students interpret and construct texts considering text type and commonly associated purpose of texts with personally significant and some unfamiliar subject matter, making connections between directly stated information and main ideas. They draw on patterns of textual resources, commonly associated with text types that are used to organise and link ideas and information. They identify positive and negative textual representations. (QSA, 2005b, p. 38)

As well as the English KLA syllabus, all other KLA syllabus documents refer to literacy instruction within their specific curriculum area, focused on the development of multimodal curriculum specific skills through reading, writing, shaping, speaking, viewing and listening. Each of the Years 1-10 syllabuses adopts the view that literacy is a social practice, with at least two syllabuses, The Arts and Studies of Society and Environment, explicitly identifying a social-critical literacy role for learning in the subject. Additionally, the Arts KLA syllabus (QSA, 2002), like the English syllabus, frames literacy as a multimodal capability requiring combined use of the design elements in the range of semiotic systems (visual, linguistic, audio, spatial and gestural modes) to demonstrate learning.

The Arts key learning area uses English literacy skills as well as contributing to the development of those skills. In addition, students become literate in the symbols systems used within the arts disciplines to convey meaning using current and emerging technologies. (QSA, 2002, p. 4)

The inclusion of literacy as a priority in all key learning areas is supported by recent research on the curriculum-literacy interface (Cumming, Wyatt-Smith, Ryan & Doig, 1998) research undertaken with students engaging in the senior years of schooling that indicated that each area of curriculum requires students to be taught and to demonstrate literate capabilities that are unique to its domain. The QSA *Literacy Position Paper* (QSA, 2001a) elaborates on the many varied literate practices that student encounter at each level and in each key learning area. Described under the Four Resources framework, typical code breaking, text participant, text user and text analyst practices for each learning area are provided as a heuristic for literacy instruction in the primary and secondary years. Further, the recent *Literacy the Key to Learning Framework for Action—2006-2008* (Education Queensland, 2006b) presents literacy in the curriculum as central to teaching, learning and curriculum leadership actions. The development of literacy standards in the key learning areas is part of the key action area of literacy in the curriculum.

This overview of literacy frameworks and instruction in Queensland shows that literacy is officially seen as a complex area, comprising many modalities, and of significance to all curriculum areas. Consistent across the literature is the observation that literacy is central to learning and therefore to student success across the years of schooling.

Numeracy frameworks and instruction in Queensland

In Queensland the key learning area of mathematics is positioned as central to numeracy instruction. In both the Years 1-10 Mathematics Syllabus (QSA, 2004) and the QSA Numeracy Position Paper (QSA, 2001b) the definition of numeracy relates to students being able to demonstrate their numerate dispositions in a range of practices in everyday situations.

Numeracy is the demonstration of practice of dispositions that accurately, efficiently and appropriately meet the demands of typical everyday situations that involve number, patterns and algebra, measurement, chance and data, and space. ... Central to these understandings is knowing how, when and why to engage in a range of practices that include estimating, predicting, justifying and visualising. ... (QSA, 2001b, p. 2; QSA, 2004, p. 8)

Mathematics is divided into five strands: number, patterns and algebra, chance and data, measurement and space. Table 1 displays the level statements for these strands. These statements provide detailed descriptions of behaviours expected to be demonstrated by a Year 3 student who is at (expected) level (Level 2) and a Year 5 student at (expected) level (Level 3) in Mathematics respectively.

Table 1: Mathematics Core Learning Outcomes: Level Statements (QSA, 2004)

Level 2	Level 3
<p>Number Level statement</p> <p>Students demonstrate their developing number sense by comparing, ordering and representing whole numbers to 999 and understanding that the value of a digit in a number determines its place. They understand that a whole can be made up of equal parts and use concrete materials to represent halves and quarters. When using money to purchase goods, they tender different combinations of notes and coins.</p> <p>Students are beginning to recall or work out some addition, subtraction and multiplication number facts. They use a range of computation methods, including mental, written and calculator, to solve problems.</p>	<p>Number Level statement</p> <p>Students compare, order and represent whole numbers to 9 999, common and decimal fractions and recognise the value of each digit. They tender appropriate amounts of money for cash transactions and identify other methods of paying for goods and services.</p> <p>Students recall or work out all addition, subtraction and multiplication number facts and some division facts. They use a range of computation methods, including mental, written and calculator, to solve problems that involve whole numbers and decimal fractions in context.</p>
<p>Patterns and Algebra Level statement</p> <p>Students use rules to create and describe number patterns based on addition and subtraction. They identify number sequences that are not patterns. They complete missing parts of, or continue, a number pattern when given the rule. They know the inverse relationship between addition and subtraction and use this to apply and then reverse simple rules. They display the inputs and outputs of the application of rules in table form.</p> <p>Students represent addition and subtraction situations using equations. They recognise and describe the equivalence or non-equivalence of two sides of an addition or subtraction equation (number sentence) and determine an unknown using a variety of self-generated and learned strategies.</p>	<p>Patterns and Algebra Level statement</p> <p>Students describe relationships between sets of numbers in terms of functions or rules. They draw tables and graphs to display these relationships. They know the inverse relationship between multiplication and division and use this to reverse the effect of a rule or change.</p> <p>Students represent and describe equivalence in everyday situations. They determine the missing part of an equation (number sentence) that requires either multiplication and division or addition and subtraction using a systematic guess and check strategy.</p>
<p>Measurement Strand Level statement</p> <p>Students use non-standard and some standard units to estimate, measure and order length, mass, area and volume. They understand that the larger the unit, the fewer required to measure an object, and that standard units provide consistency when measuring.</p> <p>Students measure and compare durations of events and link these to familiar activities. They read hour and half-hour displays on analogue clocks and any time on digital displays. They use calendars to locate and sequence events that are of significance to them over a year</p>	<p>Measurement Strand Level statement</p> <p>Students use equivalent forms of standard units to compare, order and measure. They select appropriate standard units to estimate and measure length, mass, area and volume. They further develop their estimation skills by identifying and using a set of personal measurement referents.</p> <p>Students interpret and use calendars, simple timetables and diaries to plan and record events in their daily lives. They know and use conventions related to reading and recording time. They calculate the duration of events.</p>

<p>Chance and Data Strand Level statement</p> <p>Students make comparisons and predictions about the likelihood of familiar events. They classify them as likely, unlikely or impossible, though their opinions are often swayed by sentiment. They understand that the outcome of a future event does not depend on the outcome of a previous event.</p> <p>Students collect and organise data, create and interpret a range of data displays and identify significant elements of the displays. They suggest and distinguish between some sources of variation in data and explain the effects of these variations.</p>	<p>Chance and Data Strand Level statement</p> <p>Students describe all possible outcomes from a single situation and order these from most likely to least likely to occur. They identify situations where every outcome has an equal chance of occurring. They estimate the probability of an event occurring by conducting experiments and analysing the results. They distinguish between situations where each outcome may or may not depend on the previous outcome.</p> <p>Students identify issues and topics of particular interest and create, trial and refine questions that allow for appropriate details to be gathered through surveys, interviews and existing sources. They organise data and experiment with a variety of manual or electronic displays, selecting those that represent the data clearly. They make statements regarding the results of their surveys using quantitative and comparative language.</p>
<p>Space Strand Level statement</p> <p>Students sort 3D shapes and objects by identifying common properties, including those that distinguish them from 2D shapes. They identify and match drawings, pictures and other representations of 3D shapes and objects from different viewpoints and orientations. They investigate nets of some shapes by unfolding and refolding packages.</p> <p>Students recognise that maps and plans are representations of environments and use these to follow or give directions related to pathways and the location of objects. They interpret and use alphanumeric grids to describe locations. They create simple sketches that represent parts of familiar environments and recognise the relative size and proximity of objects and locations.</p>	<p>Space Strand Level statement</p> <p>Students identify and visualise the geometric properties that define and distinguish families of prisms, cylinders, spheres, cones and pyramids. They recognise and describe the properties that distinguish trapeziums and rhombuses from other quadrilaterals, as well as the properties that distinguish different groups of triangles. They describe the properties of shapes using terms such as parallel, congruent, symmetrical and other terms related to angles. They draw shapes using various conventions to indicate particular geometric properties.</p> <p>Students describe locations and directions with reference to the four compass points and grid references displayed on simple, authentic maps and plans. They interpret the symbols used on maps by referring to the keys and legends and know that most maps are orientated to the north.</p>

However, similar to literacy, numeracy is also viewed as a cross curricular priority within most curriculum areas. The Years 1-10 SOSE syllabus observes that:

[n]umeracy involves the ability to choose from known number, spatial measurement and data concepts and skills and to apply them to the mathematical demands of schoolwork and everyday life. Numeracy skills are developed as students solve problems by applying numerical and spatial concepts and techniques.

In this key learning area, learners develop and use numeracy skills to solve problems related to their social, built and natural environments. In particular, students are involved in collecting, organising, analysing, critiquing and synthesising data, and using numerical language and reference systems. (QSA, 2000, p. 6)

Beyond the Years 1-10 Mathematics Syllabus, and in each KLA syllabus except English, broad descriptions of the numeracy demands of the learning domains such as the one above are provided as brief, explanatory guides to the numerate practices in which students need proficiency in order to achieve in that learning area. In addition, the QSA *Numeracy Position Paper* (QSA, 2001b) elaborates the different numeracy requirements and demands of the key learning areas, including English, under the headings of *number sense*, *spatial sense* and *measurement and data sense*. The position paper notes that primary teachers have been traditionally well placed to integrate numeracy across curriculum areas but advises that integrated planning for numeracy should not draw attention away from the specific nature of numeracy in particular learning areas.

The QSA *Numeracy Position Paper* clearly suggests effective planning for and instruction of numeracy in all key learning areas through the key actions below:

- ◆ analysis of the ways in which different key learning areas are constructed and the development of explicit descriptions of the numeracy demands of each key learning area;
- ◆ identification of aspects of numeracy that may occur only in one key learning area;
- ◆ identification of numeracy demands for specific community needs but which may not be strongly related to key learning area-specific demands;
- ◆ identification of ways to program for numeracy across all areas of the curriculum with an explicit focus on numeracy learning;
- ◆ devising sequences of lessons in which there is a cumulative development in students' demonstrations of numeracy;
- ◆ devising sequences of lessons in which there is a deliberately planned movement between teacher-led and independent numeracy learning;
- ◆ specific judgments about students' progress in numeracy;
- ◆ evaluation of the success of such programs, redirecting, and refocusing teaching and learning activities where necessary and;
- ◆ selecting and constructing appropriate resources for numeracy learning. (QSA, 2001b, pp. 24-25)

This brief analysis of the Mathematics KLA, the *Numeracy Position Paper* and other KLA syllabuses shows that in Queensland, while there is a strong alignment between a traditionally focused mathematics syllabus and numeracy, numeracy is also seen in a broader context of application, and in particular in terms of curriculum numeracies. At the time of writing a numeracy framework to complement the *Literacy Framework for Action 2006-2008*, mentioned earlier, is being developed with potential to elaborate the central significance of numeracy to all student learning and success.

Schools, teachers, classroom assessment and external test data

Engaging with data to support student learning

Queensland state education policies relevant to school accountability refer to use of external test data to set targets to improve and monitor standards of performance in literacy and numeracy (Education Queensland, 2005). Research by Nisbet (2004) surveyed approximately 125 Queensland teachers on the effectiveness of the use of external test data for monitoring student improved learning. The aim of this survey study was to establish teacher attitudes and actions regarding the external numeracy tests. The research explored teachers' attitudes and beliefs about the tests, how schools and teachers used the test results, the impact of the tests on teachers' practices, and the responses of teachers and students to the tests.

The project findings indicated that the teachers generally had negative responses to the tests. The majority of teachers (60.3%) believed that the tests did not assist students' learning. Only 25.7 per cent of the participating teachers believed that the results gave an accurate indication of students' numeracy ability. Furthermore, 77.7 per cent of teachers believed that the test results arrived too late to be of any use and very few teachers used the results for planning or to improve their teaching. Only 40.5 per cent of teachers reported use of the results to identify students who were having difficulties. Schools tended to use test results more than individual teachers, with the majority of teachers believing schools used the results to establish areas/topics in which particular cohorts were experiencing difficulties. It appears from the research that the majority of teachers taught students how to fill out the test and gave them practice tests before the day, however few teachers reported that they let the tests influence what they taught, how they taught, or how they assessed. Teachers interpreted student responses to the tests as negative experiences.

Only 20.6 per cent of teachers reported that their students coped well with the tests, with the majority of teachers expressing high levels of anxiety in their students before the tests.

A United Kingdom study (Ashby & Sainsbury, 2001), exploring the use primary schools made of national test data, concluded that only 40 per cent of schools had written policies concerning the use of national test data. However, 95 per cent of teachers reported using test results to inform curriculum, and the results for the 229 schools that participated in surveys in 1998 and 2000 showed little change in attitudes. Most schools reported that they had access to assistance from their local education advisor in interpreting the data in different ways. The most common use of the test data, curriculum management, was to check national curriculum attainment and monitor performance. For curriculum planning, data were used most commonly to inform whole school planning. In all three subject areas covered in the national tests (English, maths, science), the test data were reported to be more widely used than any other type of information.

Two studies report conflicting understandings of the purposes of assessment within schools (Brown, 2004; Hayward & Hedge, 2005). Brown (2004) found that teachers were less likely than leaders and managers to identify the purpose of assessment as the improvement of teaching and learning. By contrast, through a series of consultations with educational stakeholders in Scotland, Hayward and Hedge (2005) identified that all stakeholders believed the main purpose of assessment should be to support learning and teaching through informing future learning and providing feedback to other teachers, parents and the students themselves. However, there was less agreement amongst stakeholders when it came to assessment for accountability. Some respondents believed that assessment could be for both learning and accountability, while others believed that if assessment is used for accountability purposes then it would become dominant and drive out assessment for learning. The authors considered the tension between these purposes of assessment ‘very real’ claiming it may have a negative impact on teaching and learning if not reconciled.

Stakeholders in Hayward and Hedge’s (2005) study reported that the primary school based assessment data sent with students to secondary schools had little bearing on secondary teachers’ instruction. Secondary teachers involved in the consultations claimed that reports they received about particular student performance levels from several primary schools were inconsistent with the subsequent standard of student performances demonstrated in secondary school. Hayward and Hedge (2005) reported that this occurrence led secondary teachers to disregard information about student achievement levels passed on by primary teachers in favour of their own assessment. However, the researchers added that while most departments in secondary schools did not use primary school student assessment data, the main use that did occur was in the English department.

These studies report different degrees of acceptance and engagement with external tests and test data. The differences may reflect historical and cultural contexts of testing—in England, external examinations at the end of school years have long been a part of established practice, in the USA, standardised tests are most common, in Australia, and Queensland, the external tests are only beginning to be recognised both as having impact on schools and classrooms for accountability purposes and as potential sources of data for identifying student learning needs and improving learning.

Relationships between classroom assessments and external indicators of achievement

Research from a number of sources (Fuller, 2000; Hodges, 1992; Hoge and Coladarci 1989; Kellis and Silvernail, 2002, Perry & Meisels, 1996) has demonstrated that teacher judgment of student performance is generally consistent with student achievements on standardised tests. Most of this research is undertaken in the USA where, as noted, external standardised tests are frequent. In much of the research, there is an underlying, and perhaps hidden, assumption, that the outcome

that is indicated by the test is the 'true' depicter of student achievement, and the research investigates the degree to which the teachers' own assessments are 'right' against this standard. Despite this, however, it is interesting how frequently teachers and external tests are found to agree about a student's performance. In some instances it is noted that teachers' judgments may be more comprehensive than a test outcome, while in other instances, concerns are raised as to whether teachers protect students or use apparently extraneous factors in making judgments.

In one three year longitudinal study, conducted in an American school district, the researcher followed students from Kindergarten to the end of Year 2, specifically exploring teachers' ranking of their students' literacy achievement obtained through informal assessments and comparing it with the scores students received on standardised tests (Hodges, 1992). It was found that teacher judgments correlated highly with standardised test results and that teachers tended to measure student success on similar factors. In addition the researcher noted that teacher judgments may be a more reliable measure of students' early literacy success because they include knowledge of their students' developmental processes in reading and writing.

Fuller (2000) attempted to gauge how effective teachers in a number of schools were at judging whether their 4th and 6th grade students would pass or fail a statewide test in Ohio. In this study teachers were asked to judge, three months before the state wide tests, whether they thought their students were 'likely to pass', 'uncertain to pass', or 'unlikely to pass'. The findings showed that teachers were more 'accurate' at judging students likely to pass than they were at judging students likely to fail. For example, for judgments made by Grade 4 teachers before the Grade 4 citizenship test, 86 per cent of students who passed were judged as 'likely to pass' and only 4 per cent of student who passed were judged 'unlikely to pass', whereas, only 33 per cent of students who failed were judged 'unlikely to pass'. Fuller concluded that teachers may be unwilling to say a student is going to fail, as this may 'jinx' the student or symbolise that the teacher has given up on them. When a teacher did claim a student was 'unlikely to pass' they were almost always right. The study found no significant difference between the predictions of teacher judgments in high performing schools compared with teacher judgments in low performing schools. Fuller also suggested that one reason for the accuracy of teachers' judgments was that they were informed and aware of the proficiency outcomes by which students were being judged in the tests. Therefore they could discuss or judge how a student would fare on the particular test, knowing the match between their students' strengths and weaknesses and stated expectations of quality.

In a review of 16 studies examining the consistency of teacher judgments with external test scores, Hoge and Coladarci (1989) found that on the whole, the data presented in the studies showed a moderate to strong correlation between teacher judgments and student achievement. A similar review by Perry and Meisel (1996) on the use of teacher judgments to assess student achievements also found that teachers could make informed judgments of student performance, stating that in some cases, teachers were found to make more informed judgments than standardised measures. Perry and Meisel discussed various ways teacher judgments could be characterised, for example, direct—judging students' performance on a particular criterion measure; indirect—making more global judgments such as achievement in a subject area; and specificity—referring to the type of judgment such as ranking, grading and rating. In general, it was found that the more specific and direct the judgment, the greater the consistency. The researchers found that consistency was also found in indirect judgments using a rating scale, as long as teachers had a thorough understanding of the judgments they were making. The research evidence suggested that gender and behavioural issues did not bias teachers' judgments of students. However, like Fuller (2000), this study found that teacher judgment of high achieving students was more consistent with external measures than for low achieving students.

Interestingly, Black and Wiliam (1998) in an extensive review of assessment research studies considered that the reason teachers' prediction of the likelihood of students passing external tests is often high is because their own tests emulate the content of such tests. Black and Wiliam (1998) stressed, therefore, that such predictions did not necessarily mean teachers had depth of

understanding of the interaction between assessment and learning in their routine classroom assessments.

Dimensions of teacher judgments of student achievement

A recent Australian school based research project (Scarino, 2005) focusing on the complex processes associated with specifying and applying criteria for judging student performance problematised the comprehensiveness of teacher judgment to consistently determine student achievement. Scarino considers that assessment tends to be seen as a positivist process, rather than a process that is informed by cultural and social processes. However, Scarino (2005) claimed that findings from moderation style exchanges between teachers showed that teachers 'for a range of reasons related to their students, their experiences, their expectations and values, their social relationships and their personal and professional identities...' (p. 16) moved beyond the official criteria in making judgments about their students' performances.

Understanding one's own constructs and values, their formation, and power in use, and for the researcher, understanding the teacher's constructs and values and how they are brought into play, is (sic) the starting point for gaining deeper understanding of the processes of judging as central to validation. (Scarino, 2005, p. 16)

Other research by members of the project team has focused on Queensland middle years' teachers' work with the alignment of assessment, curriculum and pedagogy. In this work teachers reported that they had little prior experience in using stated assessment criteria and standards and making standards based judgments. Teachers claimed this gap in their knowledge of the specification of criteria and standards often led to judgments based on other social and cultural factors about the child and not the assessment task. Teachers reported challenges in developing statements of criteria and standards that took account of the demands of both curriculum knowledge and curriculum literacies and numeracies. Teachers also talked about the need to include in their statements information on student performance characteristics such as independence in completing work.

Kenny and Chekaluk (1993), comparing teacher based assessments of students' reading performance with student achievement on various language and reading standardised tests, found that while teacher assessment had many benefits, factors such as classroom skills and personal-social characteristics unrelated to students' 'ability' could predispose teacher judgment of student performance. In this Australian study, teachers of students between Kindergarten and Year 2 filled out a questionnaire about each student's reading, behavioural and academic performance. In addition teachers were also asked to judge students as low achievers (at risk), average achievers (not at risk), or high achievers. While the findings indicated that teacher assessment was cost effective, less time consuming, and had the advantage of being performed by people in daily contact with the students the researchers reported that research on teacher assessments for referral to special education found that such teacher assessment could be restrained by bias or perceptions of the student, such as behaviour, social, or cultural disadvantage. However, the results of the questionnaires completed by teachers were found statistically to be the best predictor of reading achievement for students in Years 1 and 2. For Kindergarten there was more variance, the researchers suggesting this may be because of the rapidly changing cognitive and behavioural characteristics of students in that stage of their schooling. It was also found that teachers in this study, and perhaps these early years, were more likely to make false positive errors, assigning students to the 'at risk' category when the tests placed them at 'no risk'.

The question that arises when considering the extra dimensions that teachers bring to their assessments of students' achievements is whether these represent bias, or additional information identified by teachers as important. This issue was examined in a study by Wyatt-Smith et al (2003a, 2003b), where such dimensions were focal in teacher judgments in the absence of other standards or criteria. While Kenny and Chekaluk (1993) considered that teacher assessments

could be biased, the results of their study indicated consistency with external results. External tests of literacy and numeracy in Queensland, and Australia, were introduced within a reform agenda intended to broaden the learning goals and outcomes for students to those needed for the 21st century, and a complex, technologically based global economy. As discussed earlier, the goals were far broader than literacy and numeracy and included lifelong learning skills, 'self confidence, optimism, high self-esteem, and a commitment to personal excellence', 'the capacity to make sense of their world, to think about how things got to be the way they are, to make rational and informed decisions about their own lives and to collaborate with others', and '(to) have a foundation for, and positive attitudes towards ... life-long learning' (MCEEYTA, 1998). These dimensions go beyond the curriculum content areas that may be assessed through external standardised tests but are intended to be the focus of teaching and learning in the classroom. Similar goals are stated in all Queensland syllabuses as intended learning outcomes. The question then arises as to whether teachers are expected to focus on these outcomes in a generic sense, independent of any curriculum context, or as part of the learning framework of students' learning experiences. It has been argued that these goals are extremely important for students' future success and may be stronger determinants of such future success than more traditionally valued learning outcomes (Cumming, 2001, 2002). It may be timely to consider the way that teachers value these dimensions of learning and whether they form an important addition to, rather than a distraction from, teacher judgment of student learning. It is also important to remember that in the examination of limited external information on student literacy and numeracy performance, such testing was intended to inform a much broader educational reform agenda.

Professional development and the use of data for learning improvement

It was noted earlier that teachers may be reluctant to engage with external data to improve student learning. While professional development is clearly necessary to resolve this reluctance and assist school leaders and teachers in using all available resources and information about students as effectively as possible, it is essential also that such professional development needs to be effective. Lesley and Nelson (2004) conducted an ethnographic study of two American secondary English teachers' classroom practices within the context of district mandated accountability aimed at increasing students external test scores. At the time of the study the district had mandated English teachers to regularly prepare students with test like assessments and formats that combined the intended English curriculum with anticipated external test questions. The aim of the study was to explore how two committed teachers described their teaching practices and deployed them in a climate of test score accountability. The results demonstrated that even when the teachers had positively acknowledged test preparation professional development, externally imposed accountability measures, school based professional support and school leadership directives, it was their own sense of who they were as professionals and not external mandates and school policies, that governed their pedagogical and assessment practices. As a consequence of such secondary positioning of test preparation, the research recommendations advocated for a 'new rubric of professionalism' that supports the time and space for teacher inquiry and reflection in order to interpret the assumptions that predominantly guide their classroom practices.

Related challenges regarding teacher efficacy for use of test data were found in the studies by Kirkup et al. (2005) and Ingram et al. (2004). The main challenges inhibiting teachers' use of data to promote learning according to Kirkup et al. (2005) were: lack of time, limitations of data, trouble applying data to everyday classroom situations, and ICT related issues. Ingram et al. (2004) similarly noted the issue of time but also reported issues of teacher efficacy in use of data related to cultural assumptions about teacher roles. The researchers suggest because traditionally teachers may view their role as teaching the curriculum, not as responsibility for student

outcomes, the use of data for improvement may be perceived a relatively new and divisive concept in schools.

International research that has shown effective use of data to promote learning is often related to how much specific support schools provide teachers to work with and use the data. A number of studies provide explicit guidelines about the type of professional development and leadership that assist schools and teachers in using and interpreting data, particularly to improve planning and student learning.

In their English study, Kirkup et al. (2005) had undertaken a quantitative analysis of 529 questionnaire responses and qualitative analysis of six focus group interviews to investigate teacher and school use of data. They found, among other things, that effective use of data came from meaningful discussions between staff. This effective practice tended to occur most when there was one staff member 'in charge' of ensuring the data were being used meaningfully and supporting colleagues in this area. The most common uses for the data in the responding schools were: tracking student progress, identifying underachieving students, setting targets, and informing planning. At a classroom level, teachers also used data to: identify individual student and whole class weaknesses, develop accurate targets for students and provide evidence to support curricular decisions. Additionally, the focus groups reported that lack of training or support in using data to promote learning could lead to teachers seeing data as a threat. The recommendations made by the report are, for the most part, specific to English data and their collection and storage tools. However, more generic recommendations such as ensuring teachers have the time and training to interpret the data, and making teachers aware of the availability and potential capabilities of the data, have international applicability.

These common themes run through other research literature identifying successful evidence based practices for schools' use of data as a way to improve student learning (Boudett, City & Murnane, 2006; Cromey & Hanson, 2000; Matters, 2006). Cromey and Hanson's (2000) case study investigation of nine American schools' assessment systems found that the four schools who demonstrated successful use of school based assessment data for informing teaching and learning exhibited a set of enabling practices and conditions, including the existence of a school based 'data guru' and a 'number cruncher' who were instrumental in maximising improvement. The researchers claim, to minimise the demands placed on schools to merge state based assessments with their own school based performance measurements:

... schools must implement assessment systems that are not only aligned and integrated with local curricular, instructional practices and professional development strategies but also contribute to the goal of increasing student achievement based on rigorous content standards. This is complex, demanding work that can take several years. Similarly managing, synthesising, interpreting, and using student assessment data obtained from a multifaceted assessment system can be a daunting task for educators particularly when (a) the assessment system lacks coherence; (b) school staff do not have training or experience in student assessment; and (c) the time, attention and energies teachers and administrators are stretched to personal limits. (Cromey & Hanson, 2000, p. 2)

In essence, the investigation revealed that schools with well developed assessment systems:

1. ...align local curriculum, standards, and assessments to state content standards;
2. ...analyse assessment results to monitor student progress;
3.use state assessment results to check the validity of local assessment systems;
4. ...use assessment results to evaluate the efficacy of local curriculum and instructional practices;
5.limit the number of student assessments used to those that are purposeful and can be aligned with local curriculum and state standards;

6. ...allocate time for teachers to collaborate, reflect, and make data based decisions-individually or in teams- based upon student assessment data and the instructional implications. (Cromey & Hanson, 2000, p. 4)

Cromey and Hanson's findings are similar to the school improvement process *Data Wise* proposed by Boudett, City and Murnane (2006). The 'Data Wise Improvement Process' was born from acknowledgement that processes of using data involving reading, analysing and interpreting were such an overwhelming prospect for many school leaders and practitioners that they often did not know where to start. This phenomenon led a group of Harvard Graduate School of Education researchers and Boston public school leaders to develop solutions for addressing the existing turmoil surrounding data use in local schools and turning them into a more manageable process. The end result was a set of eight scaffolded steps in three distinct phases, preparation, inquiry and action, for schools and educational districts to use in order to improve teaching and learning.

The first phase (Boudett et al., 2006, p. 1) of the *Data Wise* process includes two steps:

Prepare

1. Organize for collaborative work
2. Build assessment literacy.

This is followed by two further phases of inquiry and action, each with three steps:

Inquire

3. Create Data Overview
4. Dig into student data
5. examine instruction

Action

6. Develop action plan
7. Plan to assess progress
8. Act and assess.

This step by step approach is cyclic which, if repeated sufficiently, will not only support school leaders to become more 'data wise' themselves but should provide leaders with more credibility to gain greater school wide use of data for improvement.

A similar process was proposed by Protheroe (2001), who reported evidence that use of 'high-quality, targeted assessment data, in the hands of school staff trained to use it effectively, can improve instruction'. Among the steps identified by Protheroe, on the basis of research evidence, was the need to have data disaggregated by school, classroom and specific groups of students, and the need to undertake detailed analyses of results by learning area or goal. Most importantly, identifying learning needs is not effective unless an instructional response is also made. Recommended staff development activities included training to 'read' and analyse results, principal and individual teacher meetings, training of 'an in-school data expert (teacher)', and the necessary provision of time for analysis and reflection. Protheroe notes that such activities should be undertaken in conjunction with effective classroom based assessments to inform teaching and learning most effectively.

Professional development in using data, and teachers' attitudes and expertise, were reviewed by Matters (2006). Matters noted that data analysis and interpretation are specific skills requiring explicit professional development. The 'Getting it Right Literacy and Numeracy' research project in Western Australia (Meiers, 2005) has indicated growth in teacher reflection on the usability of data through professional development. Support for teachers' professional development in using data to improve learning exists in various forms and degrees around Australia. For example, the

Brisbane Catholic Education Office (2005) produced a special publication themed 'Using data'. The issue covered various school initiatives while providing explicit guidelines for using datasets from the Queensland Tests. Effective uses of the data included

1. Build a professional culture
2. Create collaborative structures to allow the time to 'dig into data'
3. Engage in data driven dialogue and collaborative enquiry
4. Learn what you can from tests
5. Use multiple measures (BCEO, 2005, p. 15)

It is clear from the research that considerable and effective professional development in understanding and using external test data is important if such data are to be valued by teachers as part of a set of indicators for reporting student achievement and for integration with classroom based assessments to improve teaching and learning (Cromey & Hanson, 2000). Two recent studies (Ingram, Seashore Louis & Schroeder, 2004; Lesley & Nelson, 2004) highlight that even in the context of high stakes accountability environments, professional development design and resource allocation need to be addressed if schools, districts and systems are to eliminate barriers to the effective use of external test data.

Conclusion

This review has encompassed the many dimensions that inform this project. Literacy and numeracy are clearly complex curriculum and policy areas in Queensland and elsewhere. Teacher judgment of student literacy and numeracy performance can be a significant and effective tool in improving student learning and is also, if held against an external yardstick, shown to be reliable. The concerns may be that such reliability comes at the expense of breadth of assessment focus, perhaps emulating external measures.

The next chapter provides information on the design and conduct of the project, while the following chapters provide the project findings with respect to each of the research questions posed in Chapter One.

Design of Project

The project was conducted over a 12 month period between mid 2005 and mid 2006. A dual sampling procedure was used. First, case study schools were identified who were willing and interested participants in the project. Second, within these schools, teachers were asked to identify students with selected achievement characteristics as focus students for information gathering. In 2005, Year 3 and Year 5 teachers were involved in discussing student achievement for focus children. In 2006, the Year 4 and Year 6 teachers who taught the children identified in 2005 were interviewed, in conjunction with principals and other support staff.

Identification of case study schools

The seven case study schools were identified through a questionnaire (Appendix One) sent to 200 randomly selected schools across all sectors and geographical regions in the state. The questionnaire focused on assessment of student literacy and numeracy achievements and schools' reporting strategies with, most importantly for the study, the opportunity for schools to identify if they were willing to participate in the case study project. The questionnaires were intentionally designed to offer schools space to make comments about their practices in order to limit the influence of the project team of schools' discussions.ⁱⁱ

While questionnaire responses were limited in number (n=14), (Appendix Two), several schools identified that further engagement would be of value to the schools and seven case study schools were selected for the investigation. While the number of case study sites is necessarily small, the initial call for expressions of interest through the survey provided particularly comprehensive information on school practices and identified areas of need for the use of external literacy and numeracy Test data. The sample of schools is not intended to provide a representative sample of school sectors and types in Queensland. However, in selecting the schools the project team were mindful of the need to ensure that any findings were not a result of biased selection. The seven case study schools comprised four state and two independent schools and one Catholic school and were representative of a diverse range of socio economic and cultural student communities. Importantly, the schools had geographic diversity, with locations across the state of Queensland. Table 2 outlines the characteristics of case study schools.

Table 2: Characteristics of Case Study

Sector	Characteristics
EQ	Lower SES, Co Ed, Outer Metropolitan
EQ	Lower middle SES, Co Ed, Rural, 2 teacher school
EQ	Indigenous, Lower SES, Co Ed, Rural
ISQ	High SES, Co Ed, Christian, Metropolitan
ISQ	High SES, Co Ed Year 3, All girls Year 5, Christian, Metropolitan
Catholic Education	Middle SES, Co Ed, Catholic, Metropolitan

Identification of teachers and students

School principals invited Years 3, 4, 5 and 6 teachers to attend meetings with the project teams. The number of teachers participating in each school varied according to the size of the school and availability of teachers for meetings. In all, 56 teachers across the years were involved. In some schools, of course, teachers taught across year levels and were able to comment on children as they developed from one year to the next.

Year 3 and Year 5 teachers were asked to nominate three students each. The advice to the teachers was to choose students 'above level', 'at level' and 'below level', for each of literacy and numeracy. In many cases, teachers chose the same students for both literacy and numeracy, and in larger schools, teachers selected fewer students per teacher. In all, some 37 Year 3 and 33 Year 5 students were focus students in the project. Students were not directly interviewed in the project, nor were parents. Comments in the findings of this project about parents' use of Test data were reported by teachers.

Most importantly for the project, definitions of 'above', 'at' and 'below' level were not provided to teachers but schools were allowed to make their own judgments. 'Literacy' and 'numeracy' were also not defined. The focus of the project team was to explore schools' practices and schools' meanings.

Information gathering

The major information sources for the project were interviews with principals, teachers and support staff, Test outcomes for focus students, and student work samples provided by teachers. The logistics of data collection were managed by senior research assistants who interacted with schools and sector representatives to set up interview dates, interview processes, follow up teleconferences where distance was a considerable hurdle, and sector and participant ethical permission. They also gathered school, teacher and student classroom assessment artefacts in the form of photocopied work samples and documents or originals that were copied at the University and later returned to schools. The work samples were a combination of teacher generated tasks, commercial programs, diagnostic and standardised tests, and school report cards, as well as the external *Year 3 and 5 Aspects in Literacy and Numeracy Tests* reports for profiled students.

Visits

In total, 22 interviews were conducted during site visits, teleconferences and visits by sector representatives at Griffith University. Five of the case study sites had three research visits each over the course of the study and the remaining two case study sites each had two research visits, one in 2005 and one in 2006. The pattern of visits varied from school to school. For some schools, the initial visit to the school was to explain the project, distribute consent forms, and answer questions. The school then decided whether to be involved and the next visit involved the initial information collection. In other cases, schools had agreed to involvement and were ready to participate in interviews and provide student information on the first visit. Interviews were held as a group. In many cases the Principal, Deputy or Assistant Principal, Head of Curriculum or Learning Support Teacher participated with teachers for some or all of the interviews.

In 2005 Years 3 and 5 teachers provided information on teacher assessment of student literacy and numeracy performance in the same year as the students undertook external literacy and numeracy Tests, enabling meaningful comparisons of the information obtained from both sources. The interview data and classroom artefacts collected from Years 4 and 6 teachers in 2006 addressed the research questions on how teachers assess literacy and numeracy in their classrooms and how these teachers, the parents and schools interpret and engage with the external Test data obtained for students in 2005.

A map of the data collection techniques and research data is displayed in Appendix Three. The timelines for visits were initial visits in 2005 to obtain information about students and school practices and a follow up visit or teleconference to discuss the progress of the focus students in literacy and numeracy. According to the timing of visits, the students' Tests results may have been available for discussion at this time. In 2006, the focus of the visit or teleconference was discussion of the focus students with the new (if applicable) teachers and the Tests data.

Interviews

Following the theme of eliciting, rather than prespecifying, schools' frameworks and practices in literacy and numeracy, and in assessment and use of Tests data, the interviews were informally structured, with follow-up questions evolving from the discussion. In 2006, explicit questions on use of the Tests data were included. At the initial interviews, Years 3 and 5 teachers were asked to identify their chosen focus students and to discuss why these students had been chosen as representative, in the teachers' terms, as 'above', 'at' and 'below' levels. Teachers were asked to describe the children's performance in literacy and numeracy. From this teacher talk, the frameworks teachers were using for literacy and numeracy also evolved. Although not a direct question, teachers usually also confirmed how they validated their judgments of the students' performance, through their classroom observation, work samples, other testing and the Tests. Subsequent interviews with the same groups of teachers took place towards the end of the school year in November 2005 after the external literacy and numeracy Tests results were known. The purpose of these interviews was to determine the correlation between Years 3 and 5 teacher judgment of literacy and numeracy performance and external Tests data. Teachers were also asked to discuss their students and how they had changed over the interim period. In some cases, teachers noted surprises about students over the year. Student literacy and numeracy profiles were created from these data as a starting point for discussions the following year with teachers of Years 4 and 6.

Following the 2005 data collection visits, the 2006 on site visits tracked the profiled students in relation to their achievements in numeracy and literacy in class performance in Years 4 and 6. Teachers of Years 4 and 6 who were teaching the Years 3 and 5 focus students were interviewed about their use of the external Tests data on these students. They were also asked whether the data confirmed their in class judgments of the students' literacy and numeracy performance as well as their opinion of the consistency between the summarised student literacy and numeracy profiles, that had been compiled from the 2005 in class assessment information and external Tests data, and their own assessment information. Learning support teachers present at the interviews also commented on students' performance and the Tests data.

Additional interviews

During 2006, the project team also met with a small number of administration officers from the three school sectors in Queensland. The intention of these interviews was to gain a better understanding of how schools are encouraged at the district and sector level to use the Year 3 and 5 external Tests information. Further, these interviews helped to gather data about system aims for school use of external literacy and numeracy Tests data. The interviews also served to clarify, confirm or elaborate issues that were being identified from the school interviews. Those involved in the discussions were in positions with responsibility for management of the interface of schools and Tests data, whether through discussions with school leaders, report development or through management of professional development in the area. Their comments are not to be read as representative of their sectors at the State level, but assisted the project team in understanding common and different issues of practice, data access and school support. Various system school improvement, curriculum and assessment support artefacts were also collected from these senior officers.

Data Analysis

All interviews and teleconferences were audio recorded and transcribed verbatim with the permission of the participants (district and sector personnel, principals, teachers and their identified students). Thematic analyses were undertaken of the interview transcriptions to address each of the research questions. Extracts of teacher talk are provided in the Findings to substantiate the interpretations that the project team made about school and teacher assessment practice, perceptions of the Tests, and the nature of student achievement. As noted, student profiles were developed from the range of materials and comments available for each student.

Student work examples were examined against curriculum frameworks and teachers' comments and are used in the Findings to demonstrate descriptions of student performance.

Ethics

Ethical clearance was obtained from Griffith University and all relevant school system authorities. All schools, principals and teachers were informed about the purpose of the project, were voluntary participants and were advised that they could withdraw from the project at any time. All signed consent forms. Parents of students nominated by schools for special study were informed about the project and asked to sign consent forms allowing the project team to access their Test and school assessment data and work samples. Where parent consent forms were not obtained, information about student performance has been used in general terms to support the findings but are not explicitly discussed or represented.

Limitations of the study

As noted, the findings of this study were developed through interviews with teachers and principals in a small number of schools. While the schools are drawn from all education sectors, across geographic regions, school size and cultural and socio economic diversity of students, they are not intended to form a representative sample of Queensland schools. However, the consistency of information provided by principals and teachers over the period of twelve months, provides a substantial evidentiary basis for the findings and recommendations that have emerged from the project.

Ethical procedures require disclosure of the purposes of research to participants for them to be fully informed and volunteers. Such disclosure required focus on both teacher judgment of student performance and the Tests. Direct questions about the use of the Tests formed part of the interviews. It is possible, and perhaps evident, from the way the project was presented, that principals and teachers may have given primacy to external testing. This may well have directed the way teachers talked about their own classroom assessments, the focus many had on justifying their own judgments of student performance in terms of a range of other external test forms, and narrowing their focus in framing of literacy and numeracy to be more closely aligned to the Year 3 and Year 5 Tests. To have asked teachers questions about student literacy and numeracy performance, without investigating the Tests at the same time, may have elicited different discourses about their main assessment approaches, how they judge student performance, and their frameworks of literacy and numeracy in classroom practice. Only a further research study could demonstrate the influence the presence of the discussion of the Tests may have had on the information collected.

How teachers in Years 3 and 5 currently assess the broad range of literacy and numeracy outcomes identified in Queensland English and Mathematics syllabuses

Three main findings relate to this focal question and in what follows, these are introduced and then elaborated later in this section. First, teachers in Years 3 and 5 report relying on a wide range of practices, including teacher generated materials, standardised testing and commercial packages, including levelled readers, to assess student achievement in English and Mathematics. A finding of this study is that the range is motivated by, and reflective of, different constructs of literacy and numeracy, though in the main, literacy is synonymous in teacher talk with Subject English and numeracy, with Mathematics.

Second, the reported diversity of classroom assessment practice is suggestive of variable teacher knowledge about how assessment relates to planning, teaching, learning and reporting. More fundamentally, the diversity can be considered in relation to teacher knowledge about assessment techniques, assessment resources, standardised instruments, tests including external competitions that some schools elect to undertake, large scale testing initiatives and teacher designed assessment materials. Also at play in assessment practice is the influence of some current curricular and pedagogical frameworks, together with the residual influence of previously supported frameworks and materials. Often these coexist, though not always harmoniously, and combine with other 'tried and proven' locally prized activities forming a hybrid amalgam.

A third finding relates to the mix of assessment evidence that teachers routinely draw on to inform pedagogy and judgment. While some of this is explicit and therefore, available for scrutiny, other evidence remains 'in the head' or latent. More fundamental is the matter of the salience that teachers ascribe to both types of evidence, especially as each functions to inform efforts to improve learning. The interview data and the records of teacher assessment artefacts show that in practice, as teachers actively seek to collect and interpret assessment evidence to indicate outcomes achieved, they are also keenly 'kid-watching' (Goodman, 1978). In this work they orient to individuals and to the class group, mentally tracking progression from one task to the next, updating over time. They are also keenly tuning into not only the demonstrations of achieved outcomes, but also to learner engagement characteristics, especially as these are taken to indicate enhanced (or diminished) involvement in school learning. Essentially, it is this mix of latent and explicit information that teachers reported valuing for how it informed their efforts to maintain student engagement and progress literacy and numeracy learning over time.

Multiple constructs

Multiple constructs of literacy and numeracy informed teacher assessment practice in Years 3 to 6. Consistent across the interviews was the strong association of literacy with discipline 'English', the latter distinguished from the English in Years 1-10 syllabus (QSA, 2005b) as the official curriculum. Across the body of data, it was clear that while most teachers reported that they were aware of the syllabus, in the main it had limited influence over how they assessed.

Further, in discussing literacy assessment, teacher talk emphasised reading (decoding and comprehension) and writing (text production) as primary modes of focus. More specifically, the teachers emphasised evidence of operational or control aspects of language including spelling, grammar, and to a lesser extent, control of generic structures. The cultural and critical dimensions of the 2005 English Syllabus were mentioned rarely. In some sites, however, emphasis was given

to students' evaluative engagement with text, including comprehension and higher order thinking skills. Further, there were limited references to student assessment extending to working with information communication technologies, and no instance where higher order thinking skills and ICT's were combined in how classroom assessment occurred. This observation points to the continuing dominance of print and print related capabilities in assessment.

This is consistent with how teachers reported prioritising instruction in and assessment of what they referred to as 'the basics'. In the first segment below, for example, the teacher talked about 'concentrating on ... vocab and grammar':

Interviewer: Is there anything else that either of you would like to add about things that you're doing in literacy or numeracy or particular strengths or weaknesses of the students?

Teacher: I've been really concentrating on lots of vocab and grammar. I figure until my kids understand about full stops and capitals and adjectives and all that kind of stuff, then they're not going to be able to put any more than a really basic sentence together. And it's certainly changed the way I see things. ... just seeing how much further they've come in this term, I've started with a really heavy influence on that, I'll just keep doing that next year and maybe work directly from a novel, so they get that idea of characterisation...

While the emphasis on foundation literacy skills was common across sites, the teachers' talk also showed that typically, they decided teaching and assessment for a year level in the absence of knowledge of assessment information available from previous years. This occurred for two main reasons. First, there were cases where no information was available, due, for example to students transferring into the school, or where provision was not made within the school for collected evidence to be carried forward, say in folios, from one year to the next. There were also reported cases where teachers deliberately chose not to access portfolios, though they were available. Where this occurred, teachers reported that they preferred to make up their own minds, without undue influence from the judgments of other teachers.

Also of note is that the teachers reported making decisions about what literacy and numeracy assessment would occur, though in some cases they stated that they were not advised of what students are expected to be doing at particular year levels. In the first segment below, for example, the teacher states that she is 'not sure' about prior learning in report writing, for example:

We've started reports I'm not sure how much report writing they've done from Grade 1 to Grade 3. They've done a lot of sort of 'touchy feely' writing as I like to call it and they're quite good at that diary writing and general writing but I made a concerted (effort), and I talked to the other 5 teachers as well, because a lot of my kids can't write factually, can't write information reports, and that gets more and more important as you get further and further through your education, unfortunately that's just **the way the world works**.
(emphasis added)

The teacher's reference to 'the way the world works' suggests an interest in linking teaching, learning and assessment to literacy demands of contexts outside school. This motivation is also evident in the next segment where another teacher identifies the coverage given to literacy performance in a state newspaper:

I think editing is a big thing in Grade 5, just in general, getting the kids to read their work and fix it. I've been really focusing on that because sometimes I think they lose, well, just talking about the Sunday Mail with spelling and everything, I mean you want kids to use different vocab, you don't want them to just use words they can spell, but they need to let their ideas flow, fine. But they need to know that **in the real world** you've got to go back and read it again and fix it. And that's what I've been really trying to get them to do is to just, because I find sometimes in Grade 5 you can ask them 'How's that word spelt?' and they can spell it, it's just they've rushed, you know. And I don't think it's a fair indication

sometimes if they've done something quickly. It's just going back, proofreading, editing, I think is a skill. (emphasis added)

By contrast to literacy, numeracy was aligned more strongly with materials drawing on the Mathematics Syllabus (as distinct from first hand use of the syllabus itself). As elaborated in the next section, the teachers' reliance on particular packages (for example, *Go Maths*ⁱⁱⁱ) aligned with the Key Learning Area, was evident in how they discussed student progress and their efforts at monitoring progress over time. Additionally, the combination of the syllabus and the commercial package appeared to orient the teachers to the strands of the syllabus, their talk about assessment practice achieving coherence through use of the strands, related targeted outcomes and associated activities.

Diversity of assessment practices

Literacy, construed primarily as reading and writing, was assessed using a wide range of materials, packages and practices, as indicated earlier. (See Appendix Four for a mapping of literacy and numeracy of assessment practices, techniques and packages reported by the teachers in Years 3 to 6). Overall, teachers reported that they saw their core business to entail monitoring individual progress over time. To this end, their assessment work entailed collecting different types of material evidence. However, the relationship of the contributing elements to the total body of collected evidence, or indeed the motivation for collecting, was not traced to a commonly adopted framework or stated theoretical construct of literacy and numeracy. This was the case even though literacy/English (consistently taken as interchangeable) was commonly talked about as having reading, writing, comprehension and spelling as its core business, and numeracy/Mathematics, number, space, chance and data. This could reflect in part at least the carry forward of the *Aspects of Literacy and Numeracy Tests* programs on how teachers constituted the domains of English and Mathematics, and therefore, literacy and numeracy. The point is that with the exception of the Diagnostic Continua, there is a clear finding that the teachers did not routinely link their literacy and numeracy assessment practices to syllabus or other curricular policy documents, though some numeracy commercial packages tailored to the Maths syllabus were influential.

Beyond this, the teachers' talk showed that they relied on locally developed frames of reference for informing assessment practice. These included their prior evaluative histories and their knowledge of the types of assessment evidence they collected in previous teaching years, the evidence that peers collected in their current or other school sites, and their 'insider' knowledge of what could reasonably be expected of students at a given year level. A group of Year 6 teachers spoke of this as follows:

Interviewer: In Year 6, where do you form your standard? If you were going to say that that the group is very high or in the middle or low, from where do you get the standard? Is it from curriculum?

Teacher 1: Just compare it with what you've experienced with the children the year before.

Teacher 2: And then we do a test, which is put together here, there, just a selection of things that they've covered.

Interviewer: A Maths test?

Teacher 1: Yeah, a Maths test, taking bits and pieces from...this year we used the PAT (Progressive Achievement Test) because that's what we used last year and administered that to see how they went but see, depending on...the kids last year, it's just - the cohort really...the peer grouping.

While diversity of practice is not, of itself, necessarily a strength or limitation, some schools reported that they had taken the local decision to install mechanisms for achieving 'consistency', both in relation to assessment items and teacher judgment.

In one school, teachers spoke about the move to have common assessment items as a means of focusing on consistency of judgment using outcomes:

We're going through that at the moment. We're going through the outcomes based education and the quality teaching framework, and what did we have? The in service 'consistency of teacher judgment'. The whole point was how do you get consistency between teachers? We've got common assessment items in place across the school for exactly that so all the four streams and all the grades *we have common assessment items that all four classes do.*

In this segment there is a clear link between common assessment items and judgment consistency. Later the teacher elaborated the goal of consistency as enabling teachers to be clear about expectations at a year level. She stated:

And so that goal is to get that consistency there so somebody could come in to Grade 5 and it's there so they can say 'well what's an expected level for Grade 5? This is what the Grade 5 teachers think' and then get together with the other teachers and compare that so you can actually have something, cause it's been fairly subjective up until now I think. I don't know about anybody else but what you decide could be completely different to what somebody else decides.

2: Because personal expectations are different.

1: You might be looking at are they getting the structure right? Whereas another teacher next year might be looking at are they getting ideas? Are they being creative? You know.

T: Imagination.

Though the explicit mention of moderation is missing here, the teacher makes clear the need for teachers to 'get together with the other teacher and compare'. There is also the disclosure that what an individual teacher decides could be 'completely different'; leading into the view that teacher judgment can be 'subjective' or idiosyncratic.

In the talk segment, particular challenges to judgment consistency are associated with imagination and students 'being creative'. Yet, in assessing literacy, and in particular writing performance, most teachers across sites ascribed value to students' demonstrations of creativity, imagination, flair and humour, as well as control of other more traditional features associated with spelling, vocabulary, sentence structure and knowledge of genre. Consider for example, how a teacher characterised as a capable student:

... he's my top student in literacy, and he's very creative, he got the academic award this year for the whole of Grade 3. And just with his imagination he's very creative and he does a lot of reading, I think that really helps him with his imagination and creativity.

There is a clear association in this segment between the teacher's assessment of the student as 'top student' and her recognition of his creativity and imagination, these taken to stem in part at least from 'a lot of reading'. Where teachers referred to a student's initiating additional reading outside school, it was consistently regarded as having a beneficial effect on literacy achievement in schooling.

Beyond the influence of creativity, however, there were other occasions in the recorded talk where teachers mentioned both variable judgment practices and inconsistent decisions about the quality of student work. One such case was when a Year 4 teacher agreed to participate in the project and was asked to comment on the work of a continuing child who had been identified as an above standard focus student in Year 3. The Year 4 teacher made clear that he did not agree with the teacher of the previous year who judged the student's work to be above standard. This should not be read as signalling that the two teachers were unreliable in judging. Instead, it lends weight to the point made earlier, namely that currently teachers may arrive at different judgments of quality achievement in literacy and numeracy in the absence of explicitly stated standards.

Further, in the absence of provision of moderation opportunities at system level, there are no protocols in place to identify the actual criteria that different teachers use to make judgments about student performance. This is discussed in more detail below.

Across the body of data, moderation did not appear to be a common practice, so judgment tended to remain largely a 'private individual practice'. Further, even where moderation was practised in association with required assessment, such as in the case of the Diagnostic Continua, teachers reported that it tended to 'surface' the problem of inconsistency of judgment rather than correcting it, as can be heard in the segment below:

We currently have a number of new teachers in Yr 3-6. Each assesses in their own way and this has led to some difficulties. A couple are not familiar with the Diagnostic Continua so are finding it difficult to make valid judgments.

Taken together, the two talk segments show the continuum of assessment practice in schools: at one end are those schools attempting to self regulate by requiring teachers across streams at a year level to undertake common tasks in an attempt to achieve consistency of judgment. At the other end are those schools where 'each [teacher] assesses in their own way', with assessment understood as idiosyncratic, private practice. In the case above, account is given of factors combining, including being a 'new teacher' and a lack of familiarity with a mandated statewide framework. Further, in the second segment, we hear the teacher identifying how these led to 'difficulties' in arriving at valid judgments, even though as part of the Continua assessment protocols are documented and prescribed and moderation is required.

The difficulties associated with achieving judgment consistency in using the Continua, referred to here, were raised by other teachers also. For example, another commented as follows:

I've marked off very few things on the Continuum because I felt that they really weren't up to what had been marked off already for them. So the times that I've spent looking at the continuum, there wasn't really much more that I felt should be marked off for them.

Once again there is talk in this segment of how different teachers arrived at different judgments of literacy and numeracy achievement using the Continua. The point is that both teachers can be heard drawing attention to how the provision of stated performance expectations, say in the form of standards, do not necessarily secure consistency of judgment and reporting practice.

Observation evidence

A common characteristic of assessment practices identified across sites and across teachers was the heavy reliance on monitoring through observation. Such observation evidence, as mentioned above, typically remained latent or in the head, though recorded talk showed that teachers could readily recall first hand observations of student engagement with learning opportunities, at both the individual and class levels. The reported observation evidence was rarely recorded in written form, though it was through recourse to observed individual behaviours and interactions that the teacher saw first hand how progress was (or was not) occurring. Further, it seemed that the potential for linking otherwise discrete pieces of assessment evidence lay in the individual teacher's recollections, in many cases detailed, of how the students interacted with learning opportunities. Moreover, it was observation data that seemed most potent in the teachers' formative assessment practices and related efforts to improve outcomes for all learners. In the first segment below, a Year 4 teacher reported how she decided to read number work aloud to the student, working at the teacher's desk, to identify learning processes and areas of difficulty.

... she tends to copy off some other people so in her practice testing I saw that she didn't do too bad, but then when I asked her to come to my desk and explain it to me, the mental computation she can't do in her head she has to write it down so if something was, she had some practice questions in here where I would read it, for things like this, I would read to her and say she had to do $6+9$, she'd write it down and she couldn't do it in her head. She has to write it down on the test so she actually took a lot longer than what she was allowed

to have on the testing. So I think that's what slows her down. So it's mainly her number which I was focusing on. And I think the main factor of her not doing too well is the fact that she just says 'I can't do it' and she gets blockages and she just has the attitude that she's not very good at maths. But that has improved lately; she has got a lot better. And I've been getting her to go out with parents once a week, she'll go out and we'll focus on one particular topic and she'll go out and will get a little bit of help with that.

On this occasion, the entry point for working one to one with the student was the teacher observation that the results of in class numeracy testing were higher than anticipated. Then the teacher moves to talk about her action in deciding to read the number work aloud, recognising the student's difficulty in doing it 'in her head'. Strikingly clear in this instance is the fundamental contribution of observation in how the teacher tunes into students' learning needs, adjusts teaching and learning accordingly, including involving parent helpers, while concurrently monitoring attitude. This segment illustrates the multidimensional and responsive nature of teacher assessment in the classroom, also evident in the next segment:

... she can be a bit careless at times, she wasn't particularly keen on being bright, she didn't like being bright and I spoke to her a few times about that last year and I think I've had maybe one discussion about that this year but I think the general attitude is changing and maybe she's starting to mature, maybe it's just an improvement, maturity, realizing well maybe life isn't so bad after all is you're bright and her friends are still going to accept her despite the fact that she does get lots of things right. ... She didn't want to excel, yet she'd put her head down and some incredible work would come out.

The segment above provides further evidence of the teacher's reliance on observation as an anchor point for gauging learning progress and shifts in attitude over time. To date, while observation is a widely recognised assessment practice, only limited research has been undertaken into its possible contributions to recorded assessment evidence. The teachers' heavy reliance on observation for informing teaching, as shown in this project, provides a basis for considering further various ways in which teachers might document and examine the nature of observation evidence that they routinely collect. Until this occurs, it is likely that such evidence will lack credibility and public confidence relative to other material evidence types.

Distinguishing collecting, updating and profiling

Teachers reported routinely updating their assessment information, both in relation to material artefacts and the mental recording of progress over time. However, while information was updated through ongoing collection of information, a systematic documented approach to profiling was not reported. In effect, updating occurred as existing assessment evidence was superseded by more a recent collection. The absence of profiling as an assessment practice is of note, given that the collection of evidence in student portfolios was widely reported. This was the case even though the portfolios tended to have a shelf life of one school year. That is to say, teachers reported that they chose not to access the previous year's portfolio evidence when it was available, preferring instead to formulate their 'own initial impressions'. A primary concern for several of the teachers was the possibility that a portfolio collection including recorded assessment decisions may exert influence on how a teacher came to view a student's performance, especially early in a school year. In several instances, teachers reported that they preferred to make their own judgment, in the absence of information about other teachers' judgments from earlier years or indeed, earlier evidence collections, a point mentioned earlier in this report.

While teachers made mention of judgment in this context, there was otherwise only limited reference to the processes that they relied on to arrive at overall judgment of student achievement. Instead, priority seemed to be given to examining discrete items or a group of like pieces of achievement data, such as spelling, reading comprehension or number activities, the

focus being on what it revealed about the aspect of work being targeted. The aggregation of the data for gauging overall quality in relation to say, literacy/English or numeracy/Mathematics as discipline domains, did not figure in teacher talk. This suggests that the teachers did not routinely aggregate, and therefore did not confront the matter of how to combine essentially different pieces of data in a global judgment of quality. (See section later in the report on ‘Standardisation of task and/or moderation processes needed to be developed to assist comparability of teacher judgments’ for recommendations.)

Standards

Three findings are clear about the teachers’ practices in working with standards. First, as outlined earlier, for this project, the teachers were asked to identify students above, at and below standard. No teacher reported this to be an unusual activity. Instead, their discourse showed a reliance on ‘level’ as a notion for talking about achievement at a point in time and progress over time. This was the case even though some schools indicated that no students were below the expected standard. All schools could identify students who were above standard. We did not set out to calibrate standards and therefore cannot confirm consistency in levels or the dimensions of performance on which schools were making judgments. Instead, the point is that teacher talk about expected achievement at year levels is well established.

While teachers routinely talked about student achievement in relation to at, above or below standard, in their discourse they did not explicitly connect their judgments to stated outcomes at syllabus levels. This indicates an apparent disconnect between the teachers’ use of the term level, and formal curriculum levels. This is not to suggest that the latter have limited utility for teachers. Instead, it points to how the provision of formal curriculum documents, of itself, does not regulate teacher practice. The teachers involved in this study indicated that they knew of both the English and Mathematics 1-10 syllabuses, but as they talked about assessment, recourse to the syllabuses was not to the fore, as mentioned earlier.

The limited advice to teachers about the assessment evidence to be collected and how to judge it using stated outcomes and strands could, in part, explain this. The disconnect could also be accounted for in terms of teachers’ reliance on their own locally developed ‘in the head’ standards, as discussed in more detail next. To address this teachers could adopt the view of assessment as inquiry (Delandshere, 2002), their role being as key agents inquiring into learning, collecting and analysing data about learning. Currently, however, this project highlights how, despite syllabuses, broadly speaking, teachers adopt a technicist approach to assessment. That is to say, the focus is more on the selection and use of instruments, and less on interrogating the quality of the information that they produce and the insights that it yields about student learning. As discussed elsewhere in this report, teachers report limited professional dialogue occurring around data, a consequence being that typically, assessment information is not interrogated for stated (and unstated) standards. This situation reflects the current absence of support systems for teacher judgment at regional, state and national levels.

Mention is also made of how, in the main, the teacher assessments did not give priority to student self assessment. Essentially, assessment was the primary province of the teacher, and the students collectively were positioned as the objects of assessment (recipients of set activities). The exceptions to this were students perceived to be high achievers. This group tended to be given more opportunities to extend and to self nominate work activities. Further, more demanding work was often set for this group. This could reflect the teachers’ valuing of moving students ‘beyond’ level, though the reference to ‘beyond’ was not used by the teachers themselves. In the first segment below, the teacher talks about how extension activities are provided to a Year 3 student, while in the second segment a second teacher at the same site refers to providing extension literacy activities, especially relating to reading.

Numeracy:

And this week we've been doing volume and capacity and measuring the volume of things in space, the whole week's work he had done by half way through maths time today. It's frustrating for me because he comes to me and he's got it all right. Everything, he can just read it, does it, read it, do it, and it's right. So I'm trying to think of ways to extend him. So I mean one of the quick things, I'll buzz Year 4, 'what are you doing this week?' and grab a Year 4 volume and capacity thing, so he's still doing the same topic in maths as everybody else but he's actually expanded above where he is now. Because the Grade 3 work is just far too easy.

Literacy:

...for instance for her, I have her on a reading assignment, and a few other kids actually. So they don't do the readers, I mean they started that way, but then I said here there's an assignment and they worked on that over a number of weeks, self paced where they just do different activities. Part of it is read 3 novels, I mean some kids couldn't do that in a set amount of time, you have kids that don't like reading. So I have her doing different things, and other kids, I have other kids who enjoy reading and read novels. So yes I try to make them in certain areas do different, not that I'm putting them up to Grade 6, I'm just giving them different types of things. And she enjoys that mostly

In both segments, the talk shows the agency of the teacher in maintaining engagement and ensuring that learning and assessment accommodate individual difference. Teacher agency, however, was in some cases reliant on individual student uptake of opportunities, even when these were not directly made available to the student, as is evident in the segment below. Here we hear the teacher talking about an above level student who took an 'overheard' suggestion and in taking it up, demonstrated abilities relating to drawing out main points and transferring information to new contexts:

But ...with ..., ... the reports were then made into oral presentations as well, and I sort of dropped the hint of doing some PowerPoint's and that sort of stuff, you know the classic presentation techniques so people aren't all staring at you, which she picked up and ran with straight away. Like I didn't make a point of it in class, I just suggested it to one or two kids, she overheard it, went sounds like a good idea, came in with a PowerPoint presentation, didn't have the whole presentation typed in, you know some of the kids wanted to type the whole thing in, she just had the main points beautifully set out. Stood up, the information transference of information report to oral presentation was beautiful, like the structure was there, took the information and changed it into the fact that it was oral. A lot of kids just stood up and read what they'd written as a report. Yeah, I keep telling her she's going to be the first female prime minister of Australia.

The corollary of this was that below level students had fewer opportunities to self direct, though they were offered where possible learning support (working with teacher aide, parent helpers, learning support), both in class and in withdrawal activities. There were clear cases where the teachers identified students well below level, and it is worth noting that most of these were reported as included in the Tests. For these students, some schools reported difficulties in accessing, and hence limited use of, external diagnostic services. Overall, while all teachers reported that they valued specialist advice from allied health support, the students in the study falling 'well below level' (as established through teacher assessment) were not identified (in large scale testing programs and related school reports) as sufficiently severe to be high priority for further assessments and ascertainment funding.

A second finding is that even though teachers could readily state their expectations of quality at a year level, their unofficial standards, typically remained implicit. The practice of providing statements of outcomes or standards and criteria for tasks was not widely used. This inevitably meant that the standards were not available to other parties, including students and parents. Moreover, the standards that came to be constituted and enacted in the practice appeared to be valued for how they were locally generated and therefore, locally relevant, susceptible to change

from year to year. This finding could be expected given that currently there are no official or authorised standards for literacy and numeracy at national or state levels beyond the specification of benchmark as the minimum acceptable level. Further, it is noted that syllabuses are silent on the matter of standards for literacy and numeracy achievement across the years of schooling other than as a requirement for the award of the Queensland Certificate of Education.

The third finding relates to how the teachers routinely ascribed value to evidence of learner engagement, construing this broadly in relation to on task performance features and attitudinal features. This inevitably meant that the scope of teacher assessment, and within it, teacher judgment, was wide, with both taking account of variables that traditionally fall outside official curriculum documents. In practice, this involved the teachers in combining their collected information about 'process' characteristics (that is, how students go about completing in and outside class activities) and 'product' characteristics (indicating the quality of completed or final versions of student work), with additional individual learner focused information. In the latter category was a stockpile of information that the teachers collected about student engagement with learning and more specifically, their growth characteristics in the domains of literacy and numeracy. These characteristics had potency in informing teacher assessment of each individual's progression, raising to teacher consciousness those variables that could be attended to in classroom interactions and outside this context. They surfaced in particular in teachers' descriptions of students' engagement with learning, one of which is offered next:

With ..., basically he needs very direct teaching and constant repetition if he hasn't got it. Some things he gets, some things he doesn't. And if he doesn't, he really wants to please, and he can be quite nervous and gets embarrassed very easily. So the teacher really has to win his confidence and for him to feel supported in the classroom or I could see his maths deteriorating quite quickly, well not only maths, anything.

In this segment the teacher can be heard 'Reading the student as the child' to identify engagement, relying on this as an element to interpret how learning is occurring.

The repeated references to the attitudinal and affective domains suggest the teacher's keen interest in tuning into ways of being responsive to the student, in part to win and build his confidence.

Here and elsewhere throughout the corpus of interview data there was frequent mention of engagement variables including confidence, risk taking, enthusiasm, willingness to do more, and talking up in class. Overall, a strong finding is that teachers give high value to their 'reading' based on first hand observations of student engagement with learning opportunities and interactions with resources and peers and teacher, updated over time as teachers accrue artefacts. In effect, teachers could be said to be involved with additional banking of learner observations over time in a recursive model. That is to say, the teachers reported that they were actively monitoring learner engagement to cue themselves in to interpreting the artefacts for what they disclosed about learning, and then used the artefacts to cue in to interpreting engagement in the learning. The assessment gaze of the teacher was therefore both point in time and over time, giving assessment both a social and historical perspective.

This is the richness that the teachers added: they made provision for looking at the student/child as a social identity in school community, and accordingly, assessed within a framework that was not formally documented in local materials, or recorded in syllabus or other formal curriculum documents. The contribution of this research is to make an opening for considering how this framework was constituted by traditional or expected process and product features, as well as an expansive repertoire of learner engagement features, perhaps reflective of the broader reform goals of the *Adelaide Declaration* and national policies. Further, the framework, as enacted by teachers, was dynamic, with each of the three components taken to be dynamic and interrelated. More specifically, it was this combination that was talked about as useful for informing future learning and teaching, as well as generating teacher insights into 'distance travelled'. As such, the

framework had both ‘retrospective’ relevance, enabling teacher knowledge about what learning has occurred and how, as well as ‘prospective’ relevance, useful for informing future learning and teaching.

While teachers routinely talked of students in relation to levels, it was reported that streamed classes make it difficult to have coherent standards for ‘below’, ‘at’ and ‘above’. Where streaming was practised, the teachers reported that the range of standards in some classes was narrow, as is shown below:

When I spoke to these guys about it [identifying at/below/above standard], what we’ve done with our classes, we’ve **low and high classes** so if we pick, for instance a high from the low class, generally that should be - that’s still a standard or even sometimes a low because for instance in Sara’s class, her highest is below average still. (emphasis added)

Here the teacher identifies ‘low and high classes’, referring to a widely applied practice of selecting students into what are referred to variously as ‘like ability’ or ‘journey groups’. Such grouping is reported in this project, and in a concurrent DEST funded project, to be motivated by an interest in managing the diversity of students’ abilities and needs. In principle, it allows teachers to tailor their teaching and assessment more effectively to match student need, in the understanding that students can progress into a ‘higher’ group, if improvement occurs. Typically, this practice involves teachers working with learning support, aides and parent helpers where available, with these parties also being involved in assessment activities. Further research is needed into the sharing of assessment information that occurs across these parties, given its associated challenges in collecting, compiling and interpreting assessment information contributed by various personnel.

The preceding findings about how teachers reported working with assessment standards have particular salience given national and state moves to report student achievement in relation to A to E standards. At one level, they point to a clear need for teachers to be inducted into research based knowledge about standards in particular, and their utility in judgment models in particular. Given the current reform moves associated with the development of Essentials and Standards as part of the Queensland Curriculum Assessment Reform Framework (QCARF), it is timely for consideration to be given to how standards are conceptualised and expected to be applied not only in the designated curriculum areas, but also as they take account of literacy and numeracy as cross curricular priorities. Also vital would be the provision of clear advice about judgment models as they relate to applying the standards, and the system checks and balances for ensuring judgment consistency. Given this, it seems that while the teachers’ assessment practices could be likened to the work of ‘bower birds’ (collecting what they took to be blue bits), the attention could now shift to aggregation (building the bower) and quality.

Teacher Assessment and Statewide Aspects of Literacy and Numeracy testing

While the relationship between teacher generated assessment and statewide cohort testing in literacy and numeracy is discussed below in some detail, it is appropriate to mention here how teachers saw the function of their assessment work as Test preparation. Overall, a continuum of practice was evident. At one end were those schools that routinely drilled and rehearsed students in test taking skills and strategies, with teachers talking about their roles as preparing students to be ‘good test takers’. In these sites, test preparation was a regular part of the pedagogy, this work extending to time management, working in fully supervised conditions, and ways of entering responses into the required Test format, alone and unaided for most students. One teacher commented on her work in assisting students to meet the demands of the literacy Test, and more specifically, the writing component, as follows:

Ok, what I’ve got for you, I filled out that sheet you wanted on each child. I tried to include a variety of things, so we’ve got, now where I put ‘demand writing task’ that was

getting ready for this horrific Year 5 Test that they have to do. I usually don't like forcing children to write within a time frame, I actually think that it reduces the quality of their writing. But anyway, that's what we have to do. So you've got quite a few of those because we were practising for the Test at the time.

As the teacher spoke, she pointed to a recording sheet that she had used to indicate the range of assessments undertaken in class. Following her reference to trying to 'include a variety of things', she then narrowed her focus to the 'demand writing task', with words such as 'forcing' and 'horrific' suggesting that the Test is not a valued insertion into classroom assessment.

In this classroom and others where test practice was routine, opportunities for test taking came from a range of sources, with the schools funding these in class additional assessments. In other schools, teachers made individual decisions about the nature and extent of test preparation that was to occur for statewide testing, with the *Years 3, 5 and 7 Aspects of Literacy and Numeracy Tests* being the only fully supervised examination conditions that students encountered.

When a Year 5 teacher was asked 'How much prep for the exam did you do this year?', she responded as follows:

I did a fair bit because when I came I realised that the literacy, the kids hadn't been exposed very much, if at all, to things like comprehension, punctuation, a bit of proofreading maybe, but not anything that would have prepared them whatsoever for sitting the Test, it would have been like a foreign language.

In this segment, the teacher discloses her sense that the students were ill prepared to succeed. Further, she is heard highlighting how the act of literacy test taking was not only unfamiliar, but alien, likening the sitting of the Test to 'a foreign language', a significant association given that 'the kids' were from indigenous backgrounds. However, when other teachers at the same school spoke of literacy test preparation, they highlighted the apparent futility not only of such preparation, but of sitting the Test itself, as indicated here:

We do a lot of, we just go over and over and over on basics and I didn't do any preparation for the Year 3 Test. I just, two weeks ago inspired my children to do some independent writing and to write one sentence is very, very difficult for them independently. So like, I did say I was hoping that none of my children would have to sit the Test because I can't see much point in that.

Of note is that in this indigenous school, while some teachers reported a lack of value in test preparation for literacy, they reported being more hopeful that numeracy preparation would be beneficial. In the segment below, for example, the teacher reported being better placed to give 'help on the Test with numeracy than ... with literacy':

Teacher: All my kids were good at numeracy.

Interviewer: Was that a surprise...

Teacher: No, because I tend to think you can give them more help on the Test with numeracy than you can with literacy. I think that plays a big part in it. And the other thing we did, was a school decision I think, I know I did it, is that we didn't just do it on the two days, we kind of like did it over four days in sections. So we didn't do all the numeracy on one day, all the literacy on one day. We segmented it over four days, and that also helped with absences and things like that as well.

A common teacher comment was the need to cover a year's work content prior to the testing, and perhaps clearer evidence of the washback impact of the Tests on classroom practice.

... especially with maths, there were a lot of concepts that I almost had to rush through just to make sure they had a basic understanding for these 3, 5, 7 Tests, because a lot of those concepts were things that I would have stretched out across the year.

... there has been, here, such a heavy emphasis on number, getting basic things in place, ... but for ... space ... I don't know if the others would agree, but I spend most of my time just consolidating basic concepts.

Consistent across all the teachers was the observation that a test situation involving timed writing and writing to stimulus pictures is difficult for some students, and especially for those who experienced difficulties doing what the teachers referred to as 'writing on demand'. This was the case for indigenous students and those for whom English is the first language, as evident in the teacher's comment below about a student with English as a first language in Year 5:

Teacher 1: Writing on demand, and that's what stumped.... He had to write on demand, in a set time.

Teacher 2: That's the problem with the Test though.

Teacher 1: Exactly, that's right; you know that's fine for somebody who's used to doing it. And also... they had those two pictures and they had to write to those two pictures and that really threw....

There is mention here of 'writing on demand' as being the cause for 'what stumped' the student. However, in the next segment, two teachers of indigenous students described how they were similarly 'stumped', though they accounted for this in terms of the lack of cultural relevance of 'those two pictures'.

Teacher 1: ...I thought that the Test, the writing they set for the Test for Year 5 was particularly poor choice. ... It was a narrative, but they had a picture of an old English sailing ship coming into a bay, and you had to use the word, something about how 'he swung his lantern toward the clipper or something.

... And the other one was something like an old shack being lifted up, and I can't remember, 'It was there one minute and gone the next' and that was-

Teacher 2: And I agree with (T1), none of the pictures, in the Year 3 pictures, the kids wouldn't have had a clue! They weren't culturally based.

Teacher 1: They weren't relevant in any way.

Teacher 2: They weren't relevant in any way they were just horrible and we had to do a lot of talking about it, and because our kids have a very limited imagination to start with, just so off the planet, it was terrible actually.

While these and other reported observations about the apparent causes for student difficulties with aspects of writing tasks in the literacy Testing program are different across the sites, common to all is the unresolved matter of how the teachers could have prepared students in such a way as to avoid their being 'stumped'.

Recommendations

1. The recently released *Literacy Framework for Action 2006-2008* (Education Queensland, 2006b) is timely, given impetus for increasing teacher awareness of the relationships of literacy to English and other Key Learning Areas. In drawing in part on a previous DEST funded national study of the literacy demands of curriculum (Cumming, Wyatt-Smith, Ryan & Doig, 1998), the framework can be used to spearhead research-policy-practice connections, providing vital support to identify, teach and assess literacy demands of curriculum areas. As shown in recent research conducted by members of the project team, this is key for securing success for all students, and especially those at educational disadvantage.
2. Teachers need assistance to document student assessment evidence, especially as this is included in portfolios. This would include professional learning about profiling

within and across the years of schooling. It could be reasonably expected that schools may choose to move to digital portfolios, enabling assessment information to have portability within a site and across sites in ways not otherwise possible.

3. More guidelines are needed about recording observations in ways that take account of the additional dimensions that teachers identify as important for, and part of, student progress.
4. Teacher assessments of literacy and numeracy achievement, and opportunities to achieve consistency of judgment within and across sites would benefit from descriptive statements of standards. These could be developed to show what is expected of students at different stages of education, or at least minimally in the years when benchmark identification is occurring so that benchmark performance can be situated within a continuum of standards. Further, this initiative would permit a more transparent match of Test item development to the statements of standards, and in this way serve to inform teacher pedagogy and classroom assessment.
5. Further research is needed to document the multidimensionality of teacher assessment of the type outlined above.

Comparison of teachers' judgment of students' literacy and numeracy achievement with the results obtained from the *Year 3 and Year 5 Aspects of Literacy and Numeracy Tests* program and similarities and differences

Comparison of teachers' judgments of literacy and numeracy and the outcomes of the *Year 3 and Year 5 Aspects of Literacy and Numeracy Tests* can be undertaken on two dimensions — the literacy and numeracy constructs and elaborations that are the focus of the respective assessments; and, the nature of the judgments themselves.

In the previous section, we discussed the frameworks used by teachers in their discussion of their assessments of literacy and numeracy, noting that they were not consistent across sites, nor necessarily within sites, and that links to state literacy and numeracy or syllabus frameworks were not evident. The following section explores the constructs of literacy and numeracy that evolve from consideration of documents informing the *Year 3 and Year 5 Aspects of Literacy and Numeracy Tests* program.

Literacy, numeracy and the *Year 3 and 5 Aspects of Literacy and Numeracy Tests*

The primary source of information on the constructs underlying the Tests is the Queensland Study Authority website. For example, the information provided to parents is that the Tests:

contribute to the improvement of children's learning by providing information on their achievement in aspects of literacy and numeracy in Standard Australian English.

Aspects of Measurement and Data, Number, and Space will be assessed in the Numeracy Test. Test items will be multiple choice or open ended, some of which will be based on materials in a magazine. Some items will require the use of calculators and measurement materials. Other items include mental calculations, written calculations and problem solving.

Aspects of Reading and Viewing, Spelling, and Writing, will be assessed in the Literacy Test. Reading and Viewing will be assessed through multiple-choice items, which will be based on materials in a magazine. Spelling will be measured through dictation and proofreading tasks and through information from the writing task. Writing will be assessed through a written response.

The Tests are based on Queensland syllabuses and are designed to cover a wide range of children's abilities across the State. In the Tests there will be some items that most children will be able to answer. There will also be items that only a few children will be able to answer.

...

Parents/carers can help children feel comfortable about the Tests by pointing out that they will cover content that the children do every day in class. (QSA, 2006a)

This information indicates that the Tests should be reflecting everyday classroom learning. The numeracy Tests are linked to the four of the dimensions of the Mathematics KLA syllabus. The Queensland mathematics KLA dimensions are Number, Patterns and Algebra, Measurement, Chance and Data, and Space. On the face, Patterns and Algebra do not appear to be included.

The link for the literacy Tests to either the English KLA syllabus or literacy demands or reading, viewing, spelling and writing demands across syllabuses is not so clear. Viewing is constructed as multiple choice items in response to written texts.

The draft English syllabus, introduced for trial in 2006 (QSA, 2005b), is designed to

... guide(s) teachers in developing a balanced curriculum and informed pedagogy ... in a framework that facilitates planning and teaching for effective language learning. (p 3)

As noted previously, the English key learning area has three strands:

Cultural: making meanings in contexts

Operational: using language systems

Critical: evaluating and reconstructing meanings in texts. (p 3)

Each strand has a distinct focus and each makes an equally important contribution to English as a key learning area. The strands are, however, interrelated and maintain the holistic nature of English.

The three strands together provide a framework for planning for learning and assessment in the English key learning area. The core learning outcomes are organised into these strands.

Table 3: Overview of the three strands

Core learning outcomes		
What students can do with what they know		
Cultural strand Making meanings in contexts	Operational strand Using language systems	Critical strand Evaluating and reconstructing meanings in texts
Cultural contexts and social situations influence the construction and interpretation of meanings in texts.	Textual resources are chosen to be responsive to cultural contexts and social situations.	Knowledge, values and practices of groups (discourse) influence the selective construction and interpretation of texts.

(Source: English Syllabus Years 1 to 10, p. 5 (QSA, 2005b))

Further, the English syllabus has three substrands: Speaking and Listening; Reading and Viewing; and Writing and Shaping. The syllabuses are designed in Outcome levels that are representative of developmental paths rather than specific year levels of schooling. However, the English syllabus gives as a guide that students in Year 3 and Year 5 would typically be expected to be demonstrating Level 2 and Level 3 outcomes respectively. While this is a new English syllabus, it does not deviate in focus from its predecessor.

Clearly, the *Aspects of Literacy Tests* have limited construct representation in terms of the English KLA syllabus. The focus is mainly on aspects of the three substrands in terms of Listening (Dictation but not comprehension), Reading and Viewing (using written texts as noted previously), and Writing (with single draft time). Spelling is also a focus of the literacy Tests through various item formats.

More general information about the nature and purpose of the *Aspects of Literacy and Numeracy Tests* presents different purposes.

The 2006 Tests cover aspects of literacy and numeracy in standard Australian English that will allow reporting against the Queensland curriculum and national benchmark standards. It (sic) will not cover all aspects of literacy and numeracy.

The context of the items will relate to aspects of literacy and numeracy across the key learning areas, and will assess students across the range of student abilities.

The specific purposes of the 2006 Queensland Years 3, 5 and 7 Tests are to:

- *collect data on the performance of Years 3, 5 and 7 students for reporting to parents/carers and schools and for systemic reporting*
- *accommodate the assessment of students against national benchmark standards.*

...

The 2006 Tests have the following objectives:

provide parents/carers with information on their child's progress in literacy and numeracy in relation to:

- *the English syllabus, and related materials*
- *Mathematics syllabus, and related materials*
- *syllabuses in other key learning areas*
- *and enables (sic) parents and carers to compare their child's performance with that of students in the rest of the state*

provide teachers with information on their class's progress in literacy and numeracy, in relation to:

- *the English syllabus, and related materials*
- *Mathematics syllabus, and related materials*
- *other key learning areas*
- *which helps teachers determine areas where the class and individual students need attention*

provide schools with information on their progress in literacy and numeracy in relation to:

- *the English syllabus, and related materials*
- *Mathematics syllabus, and related materials*
- *other key learning areas*
- *which helps schools to reflect on some strengths and weaknesses in their curriculum and teaching programs.*
- *help the QSA monitor changes between aggregate student performance of students in Year 3, Year 5 and Year 7 syllabuses in other key learning areas and the MCEETYA-approved benchmark standards thus assisting each authority to determine educational priorities, allocate resources and meet responsibilities regarding educational accountability*
- *help the QSA monitor the aggregate performance of specific target groups within each cohort, that is boys and girls, Indigenous students, students from a language background other than English (LBOTE), students from urban and rural settings*
- *help curriculum developers to analyse the literacy and numeracy demands of their particular key learning area*
- *to inform the QSA and the Queensland Minister for Education and Minister for the Arts of student performance standards and trends in relation to aspects of literacy and numeracy in the Queensland English and Mathematics syllabuses and syllabuses in other key learning areas. (QSA, 2006b)*

This Overview differs from the advice to parents in two ways. First, it states more clearly that the aspects of literacy and numeracy to be assessed are theoretically intended to be curriculum literacies and numeracies. More importantly, however, this advice to the general public, and most importantly schools, talks about reporting children's progress against standards and other students, emphasising the accountability purposes of the Test to external authorities and monitoring of school performance and cohort progress. The significant purpose of improving student learning is represented only by the provision of information to teachers to determine areas where 'individual students need attention'.

Other information on the general website of the QSA presents a further synthesis of frameworks and purposes. The materials included are

- A Framework for Describing Reading Items
- A Guide for Selecting Stimulus Materials
- A Framework for Describing Spelling Items
- Aspects of Numeracy test Framework
- Presentation to Principals – 2006. (QSA, 2006c)

Starting with the last document, the first slide on the PowerPoint presentation to principals on the purposes of the Test indicates the overall purposes of the Tests are to improve student learning and accountability, in that order. The presentation also notes that the ‘benchmark’ is a statistical entity that is imposed on every state and territory Test result. Without being present for the presentation to principals, it is difficult to recreate the discourse that might accompany each slide. However, the presentation is a web resource without additional commentary to be taken at face value. In this presentation, the information provided on the literacy and numeracy constructs of the Tests is:

What do we test?

- *Literacy – knowledge of the socialisation of language and texts*
 - *how language represents thinking*
 - *how texts represent and communicate that thinking*
- *Numeracy – knowledge of the socialisation of mathematics in everyday practices*

The next slide elaborates literacy testing as Reading and Viewing, Spelling and Writing. Numeracy is later defined as:

The manifestation of practices and dispositions that accurately, efficiently, and appropriately meet the demands of typical everyday situations involving number, space, measurement and data.

The following numeracy slide elaborates Strands of the Numeracy Tests as Number, Measurement and Data, and Space. The representation appears to be that the *Aspects of Literacy and Numeracy Tests* are aligned with Queensland literacy and numeracy policy statements and definitions that are far broader than English and Mathematics curriculum. However, the bridge between the Tests and such broader and universally recognised definitions of literacy and numeracy is not clarified.

The remaining documents on this website location are resource documents to assist teachers and principals to prepare their students for the Tests. All these resources appear to be guidelines developed to assist Test developers and in some cases modified to provide retrospective explanations of the frameworks used. While understanding the theoretical frameworks for item or stimulus development may assist teachers in the breadth of curriculum teaching in their schools, the material is both complex and at times difficult to relate in practice to the KLA syllabuses.

The ‘Reading’ document is quite complex. It is difficult to see how it directly assists teachers with Test preparation and understanding the framework of literacy that is being assessed in the reading and viewing. It discusses text, task and reader, Samson’s six categories of thinking, the metalanguage of the Queensland English syllabus including contextual and textual features, literal, inferential and creative comprehension using Sanders’ framework, and reading purposes and processes based on the Progress in Reading Literacy Study (PIRLS) outcomes, all in five pages. Given that reading is assessed through multiple choice items in the *Aspects of Literacy Tests*, the lack of coherence between these guidelines to teachers, the current English syllabus framework and issues of curriculum literacies may confuse rather than inform teachers or assist them in preparing students to complete the Tests successfully.

By contrast the ‘Spelling’ framework document provides a simple singular synthesis of progression in the complexity and normal acquisition of spelling patterns. Many teachers may find this a useful spelling instruction resource independent of the Tests.

The ‘Numeracy’ framework document indicates the purposes of the Testing program to be first accountability and second the improvement of student learning. While reference is made to the Queensland *Numeracy Position Paper*, it is noted that the basis for the content and development of Test items will be the *Queensland Years 1 to 10 Mathematics Syllabus (1987)* as well as subject matter from other key learning areas. The document notes that all KLA syllabuses integrate numeracy as a cross curriculum priority and ‘contain statements of literacy and numeracy demands and contributions of each learning area’. This document is explicit about its role as guidelines to Test developers or item writers.

This document has been prepared to assist Test development for the 2003 Queensland Years 3, 5 and 7 Testing Program. (p. 2)

The numeracy framework document draws on the Mathematics Syllabus to provide frameworks of Mathematical Processes, Mathematical Concepts and Thinking Processes. The remainder of the document indicates the specifications of the items, such as ‘at least 15 per cent of items must be written at benchmark levels’. Clear links to the Mathematics curriculum could be used by schools and teachers as a checklist for their own instruction, as test preparation. Links to the broader numeracy definition and policy and curriculum numeracies are not clarified.

The *Guide for Selecting Stimulus Material* has similarities to the Numeracy Framework. It also appears to be guidelines to Test developers or item writers. One hopes it is not an expectation that teachers would be developing their own such materials and multiple choice items for student practice. There are components of the guide that could be helpful in guiding students about the structure of effective writing, however, aspects such as ‘typographical layout’ are not applicable for students completing the Tests. The overlay between this document and the English or other KLA syllabus is not clear.

This website address indicates further materials are being developed as resources for teachers. What is not clear is how these are situated within the current KLA syllabuses, the benchmarks and literacy and numeracy policies. It would seem critical that teachers should know the links between the resources being developed and syllabus and other expectations. Simpler documentation as a resource for teachers and principals is essential.

Two final, but most significant, resources provided by the Queensland Studies Authority are the *Test Reporting* handbooks (QSA, 2005a) and the *Test Preparation* handbooks. For the purposes of this report, we considered the 2006 Handbook (QSA, 2006d).

The 2006 Test Preparation Handbook (QSA, 2006d) repeats the purposes of the Tests as accountability and improvement of student learning, in that order. It states that all items must comply with a number of documents including the previously discussed framework guides:

- Literacy Test Frameworks (Queensland Studies Authority, 2003)
- Numeracy Test Frameworks (Queensland Studies Authority, 2003)
- Years 1 to 10 Mathematics Syllabus (Department of Education, Queensland, 1987)
- English in Years 1 to 10 Queensland Syllabus Materials: English Syllabus for Years 1 to 10 (Department of Education, Queensland, 1994)
- Literacy Benchmarks Years 3, 5 and 7: Writing, Spelling and Reading (Curriculum Corporation, Victoria, 2000)
- Numeracy Benchmarks Years 3, 5 and 7 (Curriculum Corporation, Victoria, 2000)
- Queensland School Curriculum Council key learning area Years 1 to 10 syllabuses.

Further, the ‘student, class and school reports provide objective performance information for parents or carers and schools that can be used to:

- help diagnose student performance

- monitor growth of individual student performance over time
- compare the results of the individual students with those of the rest of the state
- assist with future planning.’ (p. 11)

This is the first statement that prioritises individual student learning and indicates a diagnostic role for the Test information. No further information on the curriculum content or definitions of literacy and numeracy informing the Test is provided.

The *2005 Test Reporting Handbook* (QSA, 2005a) indicates that the Tests ‘measure the full range of student abilities in literacy and numeracy across the state’ and further that the Tests ‘strictly adher(e) to relevant/appropriate curriculum’ (p. 7). The ‘overall purposes of the ... Tests are to account for, and to contribute to the improvement of, students’ learning in aspects of literacy and numeracy in standard Australian English’ (p. 3). Further definition of the curriculum assessed is not provided. However, the *Test Reporting Handbook* provides full information on the rating criteria for Writing, and identification of the match of each Test item to curriculum. Information is organised according to the Test topics of *Reading and Viewing*, *Writing*, and *Spelling* for literacy, and *Measurement & Data*, *Number*, and *Space* for numeracy and an Overall Numeracy result.

The Handbook provides links for the *Literacy – Writing* marking. All links are to English syllabuses and English materials. No reference is made to other KLA syllabuses. The criteria for assessment—*Contextual Factors*; *Text Structure*; *Grammar, Vocabulary, Cohesion and Punctuation*; and *Spelling*—are linked to the learning outcomes for ‘story writing, grammatical features and spelling’ for the English resources, and the standards are linked to the relevant levels of the trial English syllabus (p. 22).

Overall, examination of these documents provides evidence of conflicting and possibly confusing advice to parents, principals and teachers about the nature and purpose of the Tests. The Tests were originally introduced for accountability purposes, albeit with the goal of improving student learning as a whole, under the National Plan. The original test administration was a sampling study. The form of tests being used are most suited to such a purpose where the unit of analysis is a school, class or population subgroup cohort, not the individual child. The purpose and the value of the tests appear to have become confused with the decision to move to population cohort administration and individual reporting of the tests. Hence, the terminology of purpose shifts from document to document as to whether the dominant purpose is accountability, improving student learning as a whole, improving school or class instruction, improving curriculum, or improving learning for the individual child. Only one document refers to a ‘diagnostic’ assessment role to improve student learning.

The documents provide conflicting information, or unclear information, about the literacy and numeracy constructs informing the Tests. Emphases that are constant relate to the form and process of the testing—that the Tests are paper and pencil tests on a single administration, address only aspects of literacy and numeracy, should be considered in conjunction with other assessment information, and assess a range of student ability. The dominant explanation of the informing literacy and numeracy constructs appears to be links to the English KLA syllabus and the Mathematics KLA syllabus respectively. Little mention is made of the literacy and numeracy benchmark statements except through nomination as a source document for item development. Curriculum literacies and numeracies are mentioned in some of the documents. The Resource documents do not have clear links to the English KLA syllabus. No document makes a clear link to specific aspects of curriculum literacies and numeracies that are assessed in the Tests.

The PowerPoint presentation to Principals is concerning. On the face, it appears to represent the Tests as linking to broader definitions of literacy and numeracy that reflect the underpinning constructs of modern curriculum. However, there is no evidence in the other documentation that this is the case. International research evidence does demonstrate that students who engage with a rich curriculum also demonstrate successful achievement on more narrowly focused standardised external testing (CTAV, 1992). Such research evidence is often used to demonstrate

that teaching to the test and limited curriculum focus are not necessary to achieve good test outcomes. However, such educational outcomes do require that the rich curriculum encompasses the underlying basic skills components that are being assessed. To state merely that a curriculum is intended to be ‘rich’, without demonstrating links from tests to the curriculum, is not sufficient to inform principals and teachers or to ensure breadth of classroom instruction and high level Test outcomes.

The impression is that the Aspects of Literacy and Numeracy that are assessed are indeed narrow. It would be better to represent this information more clearly. If these are valued essential and basic learning, then this should be the representation of the Tests. This will need to be further reexamined with the mooted introduction of the national literacy and numeracy tests.

What this analysis confirms is that while teachers lack consistent literacy and numeracy frameworks, the materials around the literacy and numeracy Tests are equally diffuse and at times misrepresenting the nature of the Tests. We noted the lack of offered commentary from teachers regarding curriculum literacies and numeracies. Despite the discourse of some of the official documents, there is no evidence to support that the literacy and numeracy Tests address these areas. In the undertaking of this project, the inclusion of discussion about the *Year 3 and Year 5 Aspects of Literacy and Numeracy Tests* may have corralled the way teachers talked about their own classroom practice. Only another study of classroom assessment practice, that does not discuss the Tests, could determine this. However, there may be concern that already the Tests are affecting the literacy and numeracy focuses of classrooms to be more narrow to match the apparent focuses of the Test. Ongoing research may be needed to monitor this impact.

RECOMMENDATIONS

6. Documents related to the *Aspects of Literacy and Numeracy Tests* should be reviewed to ensure consistency of information regarding purpose.
7. Documents related to the *Aspects of Literacy and Numeracy Tests* should be reviewed to ensure consistency of information regarding constructs and to make more explicit the links between the documents and all KLA syllabuses.
8. Representations of the Aspects of Literacy and Numeracy assessed through the Tests should be simplified and be made more transparent in terms of the KLA curriculum. This will need to be revisited with any introduction of the national literacy and numeracy Tests.

Unpacking the literacy and numeracy constructs of the *Aspects of Literacy and Numeracy Tests*

The previous discussion highlighted the official account of the aspects of literacy and numeracy measured through the Year 3 and 5 Tests. A fuller analysis of components of some official documents provides more detailed information on the aspects of literacy and numeracy actually valued by the items and marking.

We asked teachers to indicate the qualities they felt described different qualities of literacy and numeracy performance, and through these elicited the constructs of importance to teachers. Similarly, we have undertaken a brief review of the language used in the analysis of past performance on the Tests, provided in the 2005 Handbook (QSA, 2005a). Through these item and assessment criteria descriptions the most explicit public explanation of the informing constructs of the Tests are obtained. The following discussion considers the feedback provided to schools and teachers regarding student performance on the previous year Test.

Language used in the analysis of the Writing Task

Among the qualities that were considered to differentiate the highest standards of writing under the criterion 'Contextual factors' were control (maintenance of third person or first person), awareness of the reader (including humour and figurative language) to engage their audience, and confident control of narrative voice. Lesser quality writing included limited response to the stimulus task. Again, for 'Text Structure', the best writers were those who showed control and precision, and character development. Poorer quality was 'rushed or incomplete endings'. For 'Grammar', clause structure and connections differentiated quality of writing, although verb structure was less an issue. High standards for 'Vocabulary' involved 'control of a vocabulary that has both range and depth' used with precision (p. 27). Use of vocabulary was seen to be conservative, with limited use of complex words, because of the impact of spelling assessment through the writing.

'Cohesion' related to the presence or lack of 'tight cohesive links' between ideas (p. 28). 'Punctuation' included good use of paragraphing, but other aspects of punctuation (not specified) tended to be more poorly undertaken across the cohorts. Finally 'Spelling' in writing showed that some students were essentially 'pre-phonemic' with words presented with letters and letter strings. More developed spellers could spell correctly multiple syllable words or words with commonly used prefixes and suffixes.

Overall, for writing, the markers' comments were that most writers were able at Year 3 to:

- ◆ respond to the stimulus with variety and imagination
- ◆ build in the words from the stimulus
- ◆ construct simple and compound sentences
- ◆ attempt direct speech
- ◆ write an introduction
- ◆ (in many cases) write a simple plot.

The markers also noted that the very best of the Year 3 scripts showed quite advanced story structure.

and, at Year 5, to demonstrate:

- ◆ good creativity and ability to develop mood and mystery in their stories
- ◆ some good planning on the part of the better students
- ◆ some good vocabulary. (p. 29)

Language used in describing the Spelling task performance.

The feedback lists the eleven (11) words used from the dictation to analyse spelling and their description in terms of *phonemic, morphemic and etymological* knowledge as well as *letter-pattern* knowledge. Students were asked to correct identified misspellings (four items) or to identify and correct misspellings (four items) based on common errors. The general feedback was that students in Year 3 were able to spell words where 'sounds map directly onto letters', that is usually single syllable words, and standard letter patterns. They were less successful in changing tense and plural endings.

Language used in describing the Reading and Viewing tasks

Every item for Reading and Viewing is described in terms of intended task demand. These include *explicit retrieval, inference* (word meaning from context), *interpretation* and *integration of ideas and information*, and *examine, evaluate and respond to content*.

The classification of items accorded with the framework of comprehension provided in the *Framework for Describing Reading Items*. Examination of the Test item descriptions for Reading and Viewing shows links to the following level and outcome statements from the English syllabus Trial Level Statements.

Cu 1.2 form simple inferences from pieces of information in close proximity in texts

Cu 2.2 make simple inferences from information that is closely related in the text

Cu 3.2 make some inferences about characters from directly stated descriptions and actions

Op 3.2 make meaning of resources that develop subject matter, signal relationships and organise and link ideas in written and multimodal texts

There were no clear links to other aspects of reading and viewing described in the English curriculum. On the face, cultural or critical aspects of comprehension were not a focus of the assessment, although ‘cultural’ understanding may have been an implicit requirement for success. The format of the stimulus materials may have provided a ‘camouflaged’ cross curriculum linkage (Cumming & Maxwell, 1999) but more explicit linkages were not clear.

Language used in describing the Numeracy tasks

The discussion on feedback on the numeracy Test performance is clearly linked to the mathematics curriculum. Many items are multiple choice although some require constructed responses. The language of the feedback is in terms of standard mathematics such as number facts, and geometrical shapes (two and three dimension). There was discussion of the use of language of chance and positional language. Year 5 feedback focused on calculator use for number, area, quadrilateral shapes and perspective (pp. 31-33). The description for each item is clearly linked to mathematical content of the KLA Mathematics syllabus and levels. An observation is that the items represent only a small component of possible outcomes, with a heavy representation of number facts and computation. Again, while camouflage in item presentation may have been present, curriculum numeracies were not apparent.

Consistency of teacher judgment and student judgment

How teachers talked about the features they considered in judging the dimensions of teacher talk had many similarities to, but also differences from, the constructs used in marking of the literacy Tests. For students described as ‘above’ level, teachers used terms such as fluent and mature, but without elaborating what these meant. However they also used the terms of the Test marking, including appropriate genre, sentence structure and ‘nuanced’ vocabulary. Imagination and creativity and engagement of the reader also figured. For students described as ‘below’ level, teachers focused much more on component parts of reading and writing, including decoding skills and phonemic knowledge. In the Tests, these only emerge in the Writing, not reporting of Reading Comprehension.

What was noticeable was the additional dimensions on which teachers reported, as discussed in Section A. Teachers made many comments about the processes and strategies that the students used for their success, as well as their attitudes and engagement with their activity. It was interesting to note that in comparing the characteristics of the above level and below level students, for one student a factor was ‘going off on a tangent’ and producing interesting work, for another student ‘going off on a tangent’ produced confusing work.

Overall, without the added dimensions the focus in writing, and reading as represented through writing, were similar for teachers and the Test markers.

For numeracy, the differences are more striking. We have noted that the numeracy Tests are very related to the Mathematics syllabus and strands. The student performance is reported as a single overall score although schools receive more detail about areas of performance. When describing the difference between ‘above’ and ‘below’ level students, the teacher talk centred around not just the degree to which a student had mathematical knowledge but their speed in acquisition of new knowledge and conceptual understanding, accompanied by information on memory strengths and concerns. Again, as for Literacy, teachers talked about students’ strategic processes as well as their approach and attitude to engaging with mathematics. Perhaps in contrast to Literacy, teachers

focused less on the mathematical elements when describing the difference between students and more on processes and learning speed.

These discussions of the dimensions of understanding teachers can bring in the identification of student learning needs prompt a call for much more specific and useful diagnostic information on student strengths and weaknesses in both literacy and numeracy to support teacher judgments. Diagnostic tests that are context and level specific are needed. Such tests need to reflect not just whether a student can answer a specific question but also to identify processing aspects of students' work. Research in the late 20th century was exploring dynamic assessments and tests that focused on cognitive, metacognitive and information processing aspects of performance (Burns, 1992).

Table 4: Teacher Descriptions – Continuum of descriptions from high to low achieving performers

Literacy	
Above	Below
<p><u>Reading</u> fluent independent reader</p> <p><u>Reading comprehension</u> can do a re-tell make inferences</p> <p><u>Writing</u> very good writer eloquent vocabulary mature vocabulary nuances formal language in report writing structure of report writing volume and quality sentence structure</p> <p><u>Processes and strategies</u> far deeper level more detailed problem solver only have to tell once</p> <p><u>Character</u> quirky put her head down and some incredible things would come out excellent creative thinking</p> <p><u>Imagination</u> goes off on a different tangent and captures the reader's attention insightful comments model sense of humour very quick witted flair initiative</p> <p><u>Attitude-approach</u> brave very confident loves books</p>	<p><u>Reading</u> inappropriate intonation struggles to read words with more than 3 letters difficulty in reading problems with decoding does not apply decoding strategies difficulty in reading and comprehension needs a lot of work on sight words and initial letter sounds and alphabet no initial sounds no sight words or initial sounds illiterate</p> <p><u>Reading comprehension</u> low comprehension difficulty with comprehension, finds it hard to understand the questions and forming inferences about texts</p> <p><u>Writing</u> struggles to spell words with more than 3 letters trouble developing past one sentence difficulty with sentence construction finds it hard to structure a proper sentence writes whole stories without punctuation finds writing tedious so it takes ages to write things down poor spelling has difficulty with Spelling very short answers basically copying from the text spelling...can't read what written even forgets what she's done. ... she'll forget what, she can't even read her own writing back to you.</p> <p><u>Processes and strategies</u> lack of risk taking "safe" person, conservative with her spelling because she doesn't try to use sophisticated words she can't spell slow in getting things going</p> <p><u>Character</u> Goes off on a tangent off the head type stuff</p>

<p>loves to write well focused perfectionist <u>Other</u> transfer from report to oral presentation neat, tidy excellent verbal communication</p>	<p><u>Attitude-approach</u> struggle with confidence really tried hard in writing just doesn't want to do it lack of writing more than skill quiet <u>Other</u> learning difficulties doesn't know how to do anything needs a lot of one on one (has forgotten) last year's (work) challenging behaviour to go along with his learning difficulties handwriting is unreadable</p>
At	
<p><u>Reading</u> <u>Reading comprehension</u> <u>Writing</u> can do simple structured writing structures sentences well <u>Processes and strategies</u> takes time to complete tasks struggling to get through the class work <u>Character</u> (less) creativity not very detailed <u>Attitude-approach</u> (needs) motivation and a good environment does take pride in it plods along keen and interested an avid reader however, comprehension is an issue Depends on her interest/motivation level</p>	
Numeracy	
Above	Below
<p><u>Knowledge of area (number, space)</u> can utilise number facts to concentrate on problem solving rather than using working memory to work the number facts out <u>Conceptual understanding</u> <u>Processes and strategies</u> catches on to concepts utilises strategies/tools modelled. speed catches on very quickly can clue on very quickly to concepts when they're introduced works quickly her intellect, able to handle and grasp well <u>Attitude and approach</u> confidence never comes and asks how to do something</p>	<p><u>Knowledge of area (number, space)</u> difficult to apply common measures <u>Conceptual understanding</u> 'quadrilateral' ... just doesn't compute trouble with abstract ideas and concepts takes time to understand something new <u>Processes and strategies</u> needs more practice than brighter child gets blockages can't work answers out in head needs that one on one instruction to explain everything <u>Attitude and approach</u> intimidated by the boys (who are all fairly good at maths) confidence (lack) social issues attitude affects application in numeracy</p>

(catches on very quickly) then tends to rush through things, not read questions properly actively participates <u>Memory</u> memory	not really interested <u>Memory</u> memory problems great difficulty remembering and processing Number facts
At	
<u>Processes and strategies</u> (thinks) mathematically tries to apply strategies <u>Attitude and approach</u> engaged student, will attempt everything (consciously engaged) the whole way through the lesson quite on the ball	

Table 4 provides a summary of the types of descriptors that teachers used when elaborating on the characteristics of students they had selected as ‘above’, ‘at’ and ‘below’ level. We compare these descriptors with the types of descriptors provided in the Test Reporting Handbook on differential standards of student performance.

Teacher judgment of literacy and numeracy achievement and the Test outcomes

We asked the teachers participating in the project to identify students in their classes they believed to be ‘above standard’, ‘at standard’ and ‘below standard’ for each of literacy and numeracy. We did not specify a definition of ‘standard’ and have used teachers’ descriptions of their justifications for choice to provide insights into their literacy and numeracy assessment constructs.

Overall, teachers’ determinations were consistent with the outcomes for the *Aspects of Literacy and Numeracy Tests*, although used by teachers, as we note in the following section, as confirming of their own judgment and seen as limited information:

What the Year 5 Tests have confirmed is basically where they were and the way I thought they were moving, and confirmed my judgments that I made during the year on the students as well.

Pretty high standard girl. Our testing kind of validates that that’s exactly where she was.

It is consistent but at the end of the day it still doesn’t give you much detail.

I guess what the Year 5 Tests have confirmed is basically where they were and the way I thought they were moving, and confirmed my judgments that I made during the year on the students as well. And also the changes that I’ve seen in the students over the year, I think they’ve been fairly well confirmed by the Test, except for, I’m not sure about the writing.

I think the Year 5 Test probably confirmed what we’d been writing on reports

(System level) ... you can call up the Test results say in Year 3 and you get an individual report for the student and the front page will come up with just the list of students and I’ll say ... pick a student, tell me one that you might be concerned about and they’ll pick that student and you do the report and sure enough it’ll be the one that’s below the state mean and right down.

... my experience is it’s very confirming you know to teacher professional opinion.

There was one notable exception. A child whose low achievement at school is of high concern gained a very high achievement level on the Numeracy Test. This child had very limited achievement across the curriculum, especially literacy and number. However, his Test result for numeracy was very high. The result for Measurement and Data was beyond the high range, his

Number was beyond the high range, and Space was above the seventy-fifth percentile. The result surprised school and teachers. The teacher had to explain to the family that the results were not indicative of work in class. No clear reason for the result could be determined. Work samples for the child confirmed very limited demonstrations of knowledge.

Teachers were most effective at identifying children who were 'below level' across all schools. In some cases, these children were identified as below level on both literacy and numeracy. The discussion of the teachers and work samples showed many children with serious learning difficulties and possible learning disabilities. This occurred across almost all schools regardless of sector or location. Many of the children were below the benchmark levels or between the benchmark and twenty-fifth percentile range. Some children were beyond range at the lower level. Teachers' comments on children who were below levels for Year 3 for Literacy included:

Touch of dyslexia, visual processing difficulty. Spelled know as 'kown' and 'on' even though word spelled out.

Reads from cues like beginning Grade 2 reading. Oral language is very good. Good at retelling events.

Identified as learning support not special needs. Asked to write a narrative, didn't know what a narrative was, wrote about hair being cut.

SEMESTER TWO SPELLING LIST TEST		Name: _____
1	up ✓	21
2	Jet ✓	22
3	gap ✓	23
4	fun ✓	24
5	wass	25
6	bast ✓	26
7	wet	27
8	drag ✓	28
9	hid ✓	29
10	leg ✓	30
11	That ✓	31
12	no ✓	32
13	ky	33
14	is ✓	34
15	we ✓	35
16	WOR	36
17	st ✓	37
18	dso	38
19	Wor+	39
20	wer	40

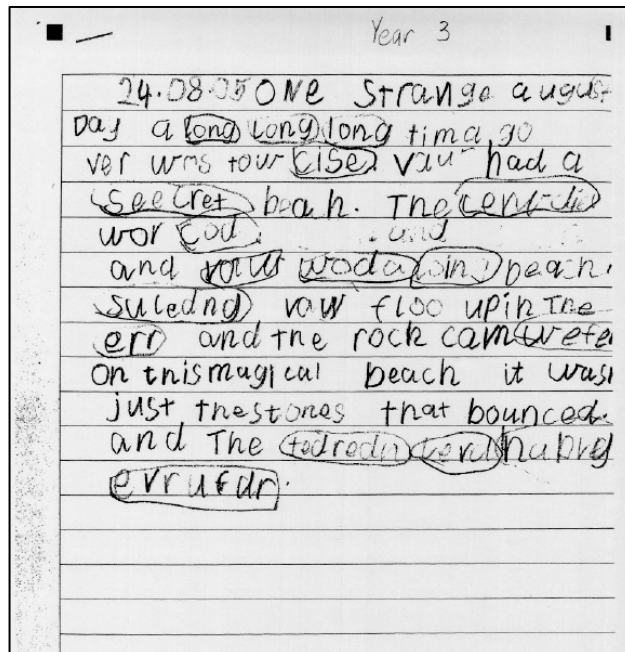
Sample spelling from Year 3 Semester Two Spelling List Test

Just started to use initial sounds in writing. Beginning to use digraphs. (ai, ew). They're completely in the wrong spot, but using them. Writing the word blue heard the bl, and had the ew.

Can do simple structured writing

Quality of writing very poor, probably not even half a page. (on Test)

Finds hard to structure a proper sentence, writes whole stories without punctuation
Phonetic choices (for spelling) don't always make sense.



Sample spelling from work sample for 'children' (cise, cenrdie, tedredn), 'they' (vaw), 'there' (ver)

For Year 5 literacy, the comments showed a small number of children still struggling with basic elements of writing and spelling:

Sentence construction very basic. Very little idea of phonics and phonemes, developing patterns.

Finds writing tedious and takes ages to write things down. Difficulty understanding comprehension questions.

Reading and spelling difficulties. Writes with a lack of space between words.

Sometimes he can't read back his own writing because he's just invented words.

Struggles to read and spell words with more than 3 letters. Oral language transposes into writing. Has fantastic ideas, which is really frustrating, getting down in writing is a real struggle

For Year 3 students considered below level in numeracy, comments included:

... needs assistance with patterning. Works with Learning Support teacher. Can't do the basic number facts with any regularity.

... memory problems, trouble with abstract ideas and concepts.

(As practice) I would read to her and say she had to do $6+9$ she'd write it down and she couldn't do it in her head she has to write it down on the Test so she actually took a lot longer than what she was allowed to have on the Testing. So I think that's what slows her down. So it's mainly her number which I was focusing on that brings her down a lot cause of the mental computation she can't do to well

For Year 5 students, the teacher talk was similar:

... not as quick at picking up concepts. Basic number facts are not too bad.

Reading and writing difficulties are affecting progress in maths. Can't read the books used in maths. Words like 'quadrilateral' too difficult. Number not very good, measurement and data not too bad, spatially brilliant. Doesn't know times table, doesn't know simple addition, subtraction with zeros.

For one student whose Numeracy Test result was higher than expected, the teacher commented:

... number over state average and I just know that that's not her. If doing straight algorithms, just written down, no problem at all. If there are words involved then she hits a brick wall.

For another Year 5 student, the teacher commented:

I don't think her results are really standard of what she does. I think it's because it's mostly multiple choice. She made it above national benchmark and in the grey area for Reading and Viewing. She reads ok but it's her comprehension that would be the concern. Not really sure how she did as well as this.

These comments show that the teachers were addressing similar aspects of literacy and numeracy as the Tests when considering the performance of these low achieving students. There was a tendency for students to appear to perform at a slightly higher level on the Numeracy Test than teachers would have expected. In some cases, the comment was that multiple choice formats may assist children to demonstrate higher standards than expected.

Two major differences emerge in the information teachers could provide about these students and the information from the Tests. Firstly, teachers generally sited the students' learning in a progression, talking about areas that were improving, despite the identification of difficulties by the Test that they recognised. Teachers also elaborated contexts where the children could demonstrate knowledge and have strength, such as oral presentations, PowerPoint presentations that are not assessed through the Tests.

I look at their sentence structure, their vocabulary, their spelling, all those sorts of things like the genre itself and how it's set out, you know I look at all those sorts of things. Perhaps I look at more than what they do for that, the writing task itself. But I mean they're not, I shouldn't say that they're bad writers, they're not bad writers, it's just not their strength. The three of them have all come along quite significantly in their reading and in their comprehension. ... has definitely improved, though he's still not as strong a reader as, or ... by any means. But he's still improved which to me, that's an achievement in itself, as long as they're improving. He's definitely stronger mathematically.

For children that teachers described as 'at level', consistency with the Test outcomes was high but slightly more varied. A small number of children were rated at below the 25th percentile, and a small number of children were rated more highly than the teachers expected—towards or just above the 75th percentile. For Year 3, the comments included:

... good reader, can read whatever Year 3 things we've put in front of her.

... needs some help with some things, but she can work things out for herself, but it's a struggle. No particular strengths or weaknesses in literacy.

... ready recall of number facts is of an excellent standard. She uses different strategies to solve number problems.

... got some problems mainly in maths. Doesn't think good at maths.

For Year 5, comments included:

Reading and viewing was good. Obviously this means that reading has improved somewhat because was in the grey area, actually made that middle area, not quite state average.

Strengths are in speaking and listening. Spelling's still a concern but picking up some skills

... I've been using a chunking strategy with them and getting them to become more aware of common chunks, ... 'and' and the 'ch' and the phonics stuff.

Good writer, punctuation, paragraphs, does have to edit.

... takes time to take on something new ... gets it eventually but she just takes a bit more (time).

Number facts are good.

... usually struggles with new concepts but once got it, got it. Have to redo a couple of times.

Middle of the road, tries really hard. Didn't do as well as I would have liked.

There was tendency for teachers to indicate that students described as 'at level' could be capable of higher standards of work if they put in more effort or if conditions around them were more supportive. Often students 'at level' and even students 'above level' were described as working without self monitoring and editing, with impact on the external Tests.

Very consistent. Produced work that is good but not outstanding, if pushed could probably do better as most average students could.

... effort needed.

... sometimes does not want to apply too much effort.

He's one of those kids who could achieve a lot better if he had (1) the support at home and (2) the motivation to be a motivated learner ... and he is brilliant if he has had a good night at home, he'll come in, he'll sit down, he'll do his work, he'll get it right. You know there are a few areas he's not by any means at the top of the class, he's in the middle, he does need a lot more guidance to go back. ... Maths-wise he tends to grasp things very quickly, but if things at home over the last couple of days or whatnot, or there's one other particular child that he bounces off in the classroom, if that child's having an out of focus sort of a day, then ... that impacts a lot and especially with the standardized Year 3 Testing, on that one particular day he mightn't have been having a good day. So you might not have been able to read every third word that he wrote because he just wanted it to be finished and not be hassled anymore. ... If he has the attention span on that day to listen then he could do quite well, and be achieving for the outcomes that we expect them to achieve

For children whose results were slightly lower than expected, teachers made comments such as:

... performed better in class than on benchmark Tests. Narrative writing not rehearsed.

I found that she did pretty well except for, I was quite surprised that she didn't do well in number, which I thought she would have done a lot better in. But her rounding also, I focused a little bit on her rounding, she tends to, if she's going round the number 568 she'll round it to 600 instead of 570 so that brought her down a little bit, but she generally did pretty well.

As one teacher commented:

... even though you look at the grading masters they use ... the students you think are writing well, don't always perform well on writing in the Test. And students that you think aren't doing very well on writing often perform much better. I think a lot of teachers do tend to look at the content, the ideas they're using, but they look at the structure of the sentences, whether they're using complex sentences or not, that sort of thing.

Overall, teachers found fewer ways to describe children who are performing at level. They are seen as solid workers, consistent, with some skills and strategies and good basic knowledge. Again the teacher talk in describing performance was consistent with the focus of the Tests.

For the higher achievers, as discussed earlier, teachers used a more diverse range of descriptors of achievement. Almost all the students identified by teachers as above level were at the 75th percentile or higher on the literacy and numeracy Test results.

One exception was a child whom the teacher described as having

Excellent verbal communication, reads fluently, really good comprehension, really good critical analysis.

who sat the first Literacy component when ill, did not complete the Tests, and was recorded at below benchmark level on Reading and Viewing. Another child who received a rating around the 75th percentile was also considered to have performed at lower than expected standard due to illness. However, the teacher did identify that spelling may need to be addressed for the child.

General comments on children identified as 'above' level, and identified similarly through the Aspects Tests included:

For reading, reading fluently, comprehension not a problem. Oral skills like talking to a 28 year old. (Year 3)

Writes and writes and writes. Spelling incredible as writes so much (Year 3)

Really interesting way of thinking. Give a problem won't give up until solved.

For one or two children the teachers had considered of a very high standard, the results were slightly disappointing,

Test outcome below expectation based on in class assessments and testing. Should have spelled more words correctly in the Test.

For some of the children identified as 'above level', their ratings were beyond the range at the higher end. Some of these children had been described by teachers as:

Writing is mature, vocabulary nuances. Thinks critically and competently. Has a whole separate list of words that I even have to check the meaning of just to challenge. (Year 5)

In summary, from the talk provided by the participating teachers in the study, teachers have the capacity to identify students who are at risk, and students who are achieving at a high standard relative to other students. Teachers' judgments of students who were achieving at a more 'average' standard were also consistent, although the reasons for students' achievement at this level varied from child to child.

The segments of teacher talk above show also that teachers looked to the *Aspects* Tests to see if they confirmed their own judgments. If the Test outcomes differed from the teachers' expectations, teachers felt confident in their own judgments as more informing and representative of a child's day to day performance in the classroom. They noted that the Tests were information gathered on one day, and that for some students illness or attention span on the day affected performance.

I think that the most valid indication is her work in day to day class; I don't think the Test really gives us an indication on her. (Student identified as below standard who still showed as low on Test)

Only in one instance did a teacher consider that the Test provided some information of an area to be addressed of which they were unaware.

The discussion also shows the consistency of the teacher talk with the item focus of the Tests. As we have noted, it is not possible to separate from the direction the project focus may have given, to identify whether it was because the Tests focus on the areas that teachers focus on as important for student learning, and representative of the learning development in literacy and numeracy, or whether the Tests have already had an impact in focusing teachers' attention to the

type of learning assessed by the Tests. In the extreme, the concern that research evidence confirms is that such Tests become by default the curriculum with the subsequent narrowing of what is offered within the classroom and the demise of other important aspects of literacy and numeracy such as oracy—a strength for many children on which teachers were developing teaching interventions, problem solving, or publishing.

Despite the degree of similarity of teacher talk about students and the Test focus, teachers commented on the narrowness of the Tests, and expressed some concern about the possible impact of the Tests on teacher practices in the classroom.

RECOMMENDATION

9. Further research to monitor the impact of the Tests on classroom literacy and numeracy instruction and assessment should be undertaken.

Teachers' comments on the 'benchmark' and overall standards

One argument that supports the use of external standardised Tests is that often teachers underestimate or overestimate students' achievement. With only one or two exceptions for individual children, most of the teachers in the study felt that the Tests gave a more positive indication of children's standards than they themselves held. Contrary to commonly held opinion, it would appear that teachers did hold high expectations for students' work. What was also apparent was that teachers did not have a high regard for the level of knowledge that needed to be demonstrated to achieve the 'benchmark' level. For them, many students in serious need of instruction, achieved above the benchmark standard. For these students, the danger was complacency on the part of parents or the students that their work was going sufficiently well.

Well, overall the Tests weren't that surprising. I was a bit more enthused when I saw some of the kids' results than I actually thought they were going to achieve.

... I just think what they're making the benchmark and what they say is reasonable ... it's really low now ... the dotted lines is that national benchmark and some of those are ridiculous ... if that is the national benchmark in Writing, we really are in trouble I think.

So you can be above the national benchmark, which sounds great and be way below average.

... being cynical, the benchmark is supposed to attract funding. So if you make the benchmark (really low) so there are very few people below it then you haven't got to fund it as well. (Doesn't mean these) kids are literate ... there's no help for them.

I'm still not overly confident in their writing abilities, despite what the 3, 5, 7 Test said.

... because his Spelling result was towards the high end and I actually, after doing the ... spelling test at the start of the year, his spelling is average, it's not – sort of – at age. I don't consider it to the high end myself.

... there are a few that, I guess, ... surprised me because I thought their results were better than what I would expect.

Teachers' comments on Tests format impact and preparation on student outcomes

Teachers' comments about the performance of students, and possible variation, included the perceived impact that the narrow focus of the Tests, the format of questions, particularly multiple choice, or the one off assessment may have had on students' demonstration of literacy and numeracy strengths and weaknesses.

... they might (think) 'oh I know what to do, I've worked out this problem' but if (they) make a mistake in ... calculations (they) might shade in the wrong number. ... It doesn't give a really clear indication of their problem solving ability.

... sometimes I think the way the questions are worded ... that it's almost as if they're trying – not trick the children but ... sometimes I feel that they're not just asking what the children know ...

... data in a certain context is useful but it doesn't cover all of the outcomes we're supposed to do ... in Literacy and numeracy ... I know the Test can only be so big ... sometimes I think parents and teachers become very stressed about the outcomes of that. ... if a child has had an off day or whatever and you've got a pretty talented child, you might get a very upset parent.

... she (high achieving student) had real trouble with (the) Year 5 ... Testing because they only had that set time, ... and she was struggling when we did the practice testing, she struggled. You know, I'd go 'five minutes' and she'd go 'Mr. ...' mate, sorry it's a horrible way but it's the way they're going to test you.

In many cases, the Tests were seen to have negative impact on the students as learners, regardless of their level.

She, well, ... does tend to get very worked up over...she does-a bit of a perfectionist and I think the Test situation last year did throw her even though the teacher prepared them fairly well beforehand, you know got them sitting on their own doing little things and do little tests, to get them used to it. I think in her case it probably could be a considerable factor.

I had one parent (exempt) their child. ... in a round about way when I finally got out of her what was actually wrong with this boy ... she just said, 'it just destroys his confidence' because he wants to see the results and it destroys his confidence.

We find it quite cruel that they are put through ... because they ask us questions all the time and we can't answer them, and then ASD kids and there's nothing we can possibly do. ... There were special considerations that we had to go through the principal with and they had extra time to finish off the questions or we give them an aide to read the question to them ... We always made sure that we had two people in every room.

The best we did for a lot of them was just move them to another room and took a lot of the kids that are low and just moved them so they weren't around that other cohort, but yeah, it was tough.

And a lot of it was just letting the kids know that just to do their best, and obviously we will look at their results but we know what they can and can't do.

The pre-writing Test briefing settings may also advantage or disadvantage classes according to group number.

But as I said, I was discussing this before, when you've only got three students and you sit down to do that writing task, and you've got those how ever many minutes it is at the beginning for planning and discussion, you can certainly achieve a lot more with three students than you can with a class of 25. So that to me, you know, we had that extra advantage, well I think it was an advantage. So I mean what they wrote I thought was quite good, in fact, I don't think ... even finished his story, and yet he still came out quite well, like on the little blue continua that they give his writing is very much on the average line.

Despite the efforts of the Test developers, as we have noted, cultural bias was still an issue for some material.

How teacher assessments can be used to provide additional information on students' performance in literacy and numeracy.

Discussion in the two previous sections demonstrates teachers have considerable knowledge about student achievement in literacy and numeracy and that there is considerable congruence between teacher talk about literacy and numeracy achievement and the outcomes for students on the external Tests. Against the syllabuses, the literacy and numeracy frameworks of both are narrow. However, the discussion also demonstrates that teacher assessment may have a richness that could enhance the systemic collection of information on students' literacy and numeracy performance.

Teacher assessments could be used to provide additional information if two necessary and related preconditions were established. First, as noted earlier, teachers collect considerable information, both artefact and observation based, and portfolios are used in many sites as composites of collections, though not currently used from year to year to update student progress. More structured portfolios could focus teacher attention on the nature of information that is being collected, and permit a more focused and systematic approach to profiling student achievement over time. While the notion of updating was implicit in teachers' talk, there is a need for explicitly connecting teacher understandings about portfolios, evidence range, profiling and selective updating, so that progress over time can be monitored. The potential of this approach is for the teachers, students and parents to have access to a body of assessment evidence, collected over time, showing progression (regression or stabilising) over time. A portfolio approach for assessing, judging and reporting student achievement in literacy and numeracy would require teachers to engage far more directly with these domains as cross curricular priorities. As stated earlier in this report, the recourse currently is to English and Mathematics, even though the most recent literacy framework and related policy, consistent with relevant research (Cumming, Wyatt-Smith, Ryan & Doig, 1998) emphasises the need to make explicit the literacy demands of curriculum, especially as these are related to assessment requirements (Wyatt-Smith & Cumming, 2003).

Second, we have noted that teachers may judge students at risk and yet the students are not identified through the Statewide *Aspects of Literacy and Numeracy Testing* programs. This observation has led some teachers to describe the benchmark minimum standard as 'too low'. One teacher spoke of this as follows:

Well, with ..., I would agree with his literacy ratings (referring to the previous year teacher selection of the student as below expected standard). He certainly is below average and would be below the benchmark for reading and viewing and his writing is below average. I'm amazed at what the benchmark is there, he's above the benchmark but the benchmark must be a long way down. Spelling he's ... yeah, well below average for spelling based upon the tests we've already done this year and numeracy he's below average ... yes, he's below average overall because he's in the lowest Maths class, we have five streamed [classes] in Year 6.

The teacher can be heard here making consistency checks between her assessment of the student's achievement in literacy and numeracy, and the report of his work on the *Aspects of Literacy and Numeracy Testing* programs. The teacher's surprise – I'm amazed – at his writing being above the benchmark is then offset by the view that 'the benchmark must be a long way down', her view staying firm that the student was below what she deemed a reasonable minimum standard.

Further, currently, teachers report that where they are given information about reported Test data, as it is returned to the school, the assumption is that they use it to confirm (or otherwise) their school performance. That is to say, the schools and teachers regard the reports as a type of feedback on the quality of their provision. In this situation, it is the Test evidence and subsequent reporting that is legitimated. Missing however, is the provision for schools and teachers to have a feedforward function which could occur if teacher information, also legitimated, could be used in identifying need. There is a strong case for this, given that the two sources of data are different, and yet in many cases, teacher and system reports of achievement are consistent, as reported earlier. To this end, schools could adopt the more systematic and focused evidence collections, mentioned above, and these could be used in conjunction with Test data. This has potential for achieving interactivity between the two data sets to inform both internal instructional planning as well as identifying students at risk.

Processes and issues

How to validate in school collections

This issue is at the heart of current Queensland Curriculum, Assessment and Reporting (QCAR) reform directions in assessment and accountability in Queensland.^{iv} While this discussion could extend to current decisions about juncture years and the apparent underpinning conceptualisation of Essentials and Standards, of more direct relevance is the foundational matter: What will be taken to count, at both system and local levels, as evidence of achievement?

As we have written elsewhere, there are continuing, strong tensions between system and site validity: the former values the primary need to prescribe and regulate to achieve uniformity across sites, washing out local variation, whereas the latter, site validity, values local responsive as a primary indicator for ensuring quality assessment, especially for learning (Freebody & Wyatt-Smith, 2004). A structured portfolio approach, as suggested above, with core and optional components specified for school selection, may provide a useful way forward for working to satisfy system needs for validity and local needs for responsiveness. The question is whether this would serve both validity and reliability purposes, offering an assessment approach across schools that could address the purposes of comparability, focusing on quality assessment and teacher judgment using standards. While it is recognised that moderation has resource implications, it is clear from existing research that any efforts to promote cross site consistency of judgment, whether through standards reference as in the Senior years of schooling in Queensland or validation efforts associated with a regulated task focus and standards, need to be supplemented by moderation opportunities, if a system is to have confidence in how standards are implemented, both within and across sites.

A related issue is the policy stance taken on what teacher assessment evidence can be carried forward for providing the system with additional information about literacy and numeracy achievement. Currently, literacy and numeracy demands remain embedded, and hence not explicit, in draft statements of Essentials being prepared under the QCAR agenda, having implications for how they come to be assessed. If teachers are to provide useful information, or if teacher assessment information is to have a utility for system reporting purposes, then the contribution of literacy and numeracy in both the Essentials and Standards need to be made explicit to make clear the connectedness of curriculum, literacy and numeracy.

For early and middle years, teacher assessment talk is multidimensional and may be drawn from ongoing classroom observation, as well as material copies of student work, including drafts, final copies and to a lesser extent, digital data. While we have noted the need for a systematic approach for documenting and validating observation, a residual question is the contribution of such multidimensional evidence – including engagement features – in what a system would choose to use and value in quality assessment. While this report makes clear the value that teachers ascribe

to such knowledge as central to their efforts in progressing learning, typically it is made available in local school reporting, and then only to a limited extent. The question is whether such knowledge represents an underutilised resource for system monitoring and/ or reporting purposes, especially as it could be used to inform the identification of students at educational risk. As noted previously, these added dimensions are reflective of the original national policy agendas.

Further, for teacher assessment and judgment of literacy and numeracy to have a utility for system purposes including reporting, attention should be given to how standards developed for Years 3, 5, and 7 and the forthcoming Year 9, include and make explicit the demands of these cross curricular priorities. As discussed elsewhere in this report, currently teachers formulate judgments of literacy and numeracy achievement in the absence of explicitly defined standards, and therefore, variation in expectations both within and across sites should be expected. By extension, there can be no public confidence currently in teacher judgment of literacy and numeracy achievement, not due to limitations in teacher practice, but instead, due to a shortfall in information about required practice, especially as this relates to standards, and of professional development of teachers on such matters. In this regard, there is a clear case for teachers to work with standards across year/outcome levels showing the expectations of quality across the years of schooling, P-12.

RECOMMENDATIONS

Based on the above, it is recommended that the QSA

10. Trial a structured portfolio approach to collecting evidence over time. The trial could include sites where only final versions or products were counted as assessment evidence, as well as those sites where other evidence types were included, such as observation data.
11. Undertake further research into the prevalence of teachers reporting the literacy and numeracy benchmarks to be lower than the standard they count as suitable for a year level. At issue here is the utility of the benchmark reports for alerting systems to students at educational risk. If the benchmark is shown to be not functioning as an accurate alerting mechanism, then teacher input into the identification of at risk students at Years 3, 5, 7 and 9 is vital.

Interpretation and use of the *Year 3 and Year 5 Aspects of Literacy and Numeracy Tests* data and information by teachers, parents, schools and systems.

Interpretation and use by the Queensland Studies Authority

The Queensland Studies Authority has responsibility for development of the *Year 3 and 5 (and 7) Aspects of Literacy and Numeracy Tests*. System representatives noted improvement in the content and breadth of items over the last few years. The source documents that were analysed previously indicate the care taken in developing specifications for the Tests. We have noted that the information contained in these documents is contradictory as to the relationships between the measurement focuses of the Tests and the broader English and Mathematics syllabuses and to curriculum literacies and numeracies.

QSA personnel did not form part of the information sources for this project. However, QSA documents did. In the analysis of source documents, we noted that the Overview for the *Aspects of Literacy and Numeracy Tests* (2006b) indicated that key purposes of the Tests were to

help curriculum developers to analyse the literacy and numeracy demands of their particular key learning area. (QSA, 2006b)

... help the QSA monitor changes between aggregate student performance of students in Year 3, Year 5 and Year 7 syllabuses in other key learning areas and the MCEETYA-approved benchmark standards thus assisting each authority to determine educational priorities, allocate resources and meet responsibilities regarding educational accountability. (QSA, 2006b)

The responsibilities undertaken by the QSA to assist the system authorities to monitor and report on student performance were not examined in the project. Similarly, the project did not explore the extent to which the QSA gathers information from the system authorities to explore the validity of the Tests, after administration. We did not explore the extent to which the QSA reflects on Test content retrospectively to analyse the match to curriculum content and student performance. The project did not explore the follow up with developers of the different KLA syllabuses regarding the literacy and numeracy demands of their curriculum. However, it is recommended that these are important components of test development, use and interpretation and should be part of systematic practice at the QSA level.

Interpretation at the school level

Interpretation

The major interpretation of Tests results by principals and teachers related to students' overall profiles and relative performance. As noted previously, the major interpretation of Tests data by teachers was to confirm their own judgment. Teachers regarded Tests data, not so much as with suspicion, but as a limited piece of information about student achievement, strengths and weaknesses in literacy and numeracy, 'one piece of the puzzle'. Lack of precise match of Tests outcomes, as far as they were interpretable from the 'bubble' and other information provided such as performance relative to class cohort, to teachers' own judgments did not undermine their confidence in their own judgment. Unlike their use of the Year 2 Diagnostic Net, and even other externally used assessment pieces such as commercial reading comprehension standards, teachers did not use the *Aspects of Literacy and Numeracy Tests* to calibrate their own judgment. If the results did not confirm their own judgments, they provided reasons for discrepancies. Only for one child

did a teacher believe the Tests data had increased their knowledge of the child's strengths and weaknesses, noting that Spelling may be an area that needed more work than had been realised.

Teachers considered, and appeared to demonstrate, that they already knew students with difficulty and their areas of concern. Teachers of children in Year 3 and even Year 4 made considerable reference to the Year 2 Diagnostic Net and continua. The work that has been undertaken with the Year 2 Diagnostic Net, understanding of continua and moderation, have led to these becoming an externally located and much broader confirmation of students' achievement, across all standards. The discussion of teachers at all schools confirmed the strength of engagement with the Continua:

The question I asked him is actually on phase C of the continua and it's 'If I went to the shop and I bought something for 80c and I gave the lady \$1 what would I get back?' and the usual thing they say is 'change' because that's what you need them to understand that there is change, and he said 'oh you'd get 20c back'. And I thought wow, so I did some change with him and he really has an awesome concept of mental computation or money.

... going back to the continua, they should be all through phase C, totally through phase C, ... and working with, if not completed D.

While teachers regarded the Tests data as a minimal reflection of student achievement based on the format and the one off testing process, few comments were made about the underlying literacy and numeracy constructs that were assessed, the relationship to English, Mathematics or other KLA syllabuses, or to curriculum literacies and numeracies. The exceptions were comments about the nature of 'demand' writing and the impact on students' achievement, and the weighting of Space in the Numeracy Test, that many teachers considered did not reflect the weighting of Space in the curriculum for children of that age, or the focus on Space in their own teaching.

As discussed in the next section, teachers in the project engaged to differing degrees with the item level data from the Tests. In all cases, the items were interpreted at face value as linked to and representing curriculum areas.

Use of Test data by individual teachers in their classrooms

Teachers' use of Tests data for their own classroom practice varied. Some looked to see how their children had 'gone', some looked for gaps in their instruction:

(Teacher) would rather look at the class profile, look across that summary sheet that says every child in that group got that Test item wrong, that's obviously a weakness for that group of kids of the majority of them got it ... so you look for commonalities, that's what you teach.

... I found it interesting ... to see how well they went, to know if ... that's an area that they hadn't done so well on so I've got to make sure I cover that better in the future, prepare them better so...

... we get the reports ... how many errors, how many did spell well, how they compared across the state ... I would prefer... (that) this goes in their file, we do have access to it. I find this a quicker, useful way to actually look at... cohort and sections ... like number, space, measurement ...

Considerable information about student progress and achievement is now documented in folios for each child in schools. The folios may pass through year levels with the child, although their use, as reported earlier, appears limited. The Tests data are usually included in the folio as part of a child's record. Some schools reported using these folios:

... with every year, at the end of every year, at the beginning of the new year ... the teachers (from either side) would get together and this sort of information would be shared if we

felt that it was valuable at that time because a Year 3 might say to us well this is how they went in the Testing so that information is used, if they feel that it's valuable.

A folio of work samples ... goes up ... Literacy and Numeracy, there are...certain work samples that go from year to year and are passed on to the teachers at the beginning of each year which provided a look at how they've done and that's set down now as a policy as to what we put in that folder.

One problem is that the folios are becoming voluminous and not as useful to or used by teachers as intended, and, as noted earlier, do not necessarily provide a coherent profile of a child rather than a collection of 'blue bits'. Several teachers in Years 4 and 6 reported they did not access the fairly simple forms of their students' data from the Year 3 and 5 Tests, until or unless they wanted confirmation of concerns with the child's achievement.

... in terms of the Year 3 Test, I glance at them at the start of the year but then I use my own feel of the kids and experiences (with) them through the start of the year to see how they've gauged from the start of the year more so than really using ... print out from the Year 3 Test.

... about in week six (I finally) have a look at the Test data and by then I was able to look and think oh yes, I agree with that, I agree with that, I could predict that. So by then I had a guide within myself of what, like I knew already where the children were so by that stage, I ... find it more interesting having taught the children after the year then seeing how they've gone ...

... Personally I try not to go to the (past) teacher's notes or past things; I try not to do that. I think in term one especially you've got to give them time to settle in, find their way in their new setting because it's different from year to year

Engagement with Tests outcomes at the item level reflected the degree to which school principals provided the data and the extent to which easy guides to manipulating the data to make interpretations existed. A major issue is the form of provision of data to schools. For example, while Education Queensland schools can access the Data Warehouse, individual reports to schools are provided in hard copy. To investigate item trends requires schools to reinput such data manually to an appropriate software program.

... so the schools get their report from QSA with performance on individual items on all the Tests but it only comes in that paper format so if ... schools want to have a look at how they might be going from year to year in particular curriculum areas based on those, which are only one test, but if they want to have a look at that it's quite time consuming and difficult to do, so they have to be quite motivated to do that. It's not something that you can download and get a report on. ... unless a school is prepared to devote the time to do all the data entry and do the analysis there's limited use of that.

There's a report that comes out ... that gives you individual children but besides it gives you their responses to the (individual) Test items. There's nothing descriptive, there's no paragraphs to summarize their performance or anything like that. ... if it's correct it's a dot, if it's incorrect it gives you A, B, C, D whatever if you really wanted to laboriously go back ... I wouldn't imagine any teacher would want to go through that long winded process of pulling out every Test item.

The presence of a significant other person in the school to assist with this, such as the learning support teacher, affected the degree of engagement.

Sharing information across whole school

Discussion with principals and teachers showed more sharing of information from the Tests for comparative purposes, although the nature and extent of sharing varied even for our small group of schools. For each purpose, the extent of use of the Tests data was linked to an individual assuming significant responsibility. First, schools would look at their comparative performance with the state and, if available, like schools. These discussions were tempered by the role the principal was willing to play.

... We also have a look at those when they come out, we sit down with ... the principal and he sort of shows us through and we can compare ... very often we are beneath the state average but when you look at the development that's been made (from) Grade 5 to Grade 7. ... I don't know that all schools would have access to that. Is that (our principal) sharing it with the staff? The principal has to log in to get into the corporate data warehouse ... it's very much on a case by case basis what teachers (in different schools) do get. (At our school) basically if (the principal) sees it, we see it ... and he really takes the time talking through it and giving his position on it ... and when we have those professional discussions, obviously we identify things that we might want to work on and what we thought worked ..., they did better in this ... what would be better in the future, you know all that sort of thing.

Principals and teachers found value in looking at the changes in cohorts over time:

(seeing the progression) from Grade 3 to Grade 5 and then from Grade 5 to Grade 7 (was useful).

It validates your job. ...when I went through (the growth graphs for school) he worked out what the average growth was for the state ... and our average leap is bigger than the average for the state but while our kids may not necessarily be up there with the others they've grown more.

A second major use of the Tests data across schools was to change programs of instruction. While implicitly the focus of the changes was to ensure fuller coverage of the curriculum and addressing gaps identified through the Tests outcomes, the overall implicit intent of these changes seemed to be to enhance Test outcomes for the next cohort:

... I think if a child is noticeably weak ... in some areas or noticeably bright in other areas that (information) might be passed on but no, as I said earlier we tend not to use this data in the following year level but we do tend to have a look at it to see how we can improve that particular year level the following year, like the 3 and 5 and 7.

...that written report that we got this year that ... identified the general weaknesses within the grade level, like you know in their writing style, the things that might have been missing in both Literacy and Numeracy, that was really helpful as a class teacher but it's also really helpful as a support teacher because then we were able to take that information to a staff meeting and generally discuss what we wanted to target in learning support with kids. ... we've actually changed our programming within the school ... from that data ... I'm working with children in Year 3, 5 and 7, working on the kinds of areas this school has traditionally been weak in and I'm going to do that till August and then I'm going to start working with the 2's, 4's and 6's in readiness for next year so that everybody's (skills) like comprehension and problem solving, those kinds of things are strengthened ...

... (in one area) we're still above state average but it's not much above so we look (to see) if there's any areas as a whole school that are lacking that we should target during the year and we've done that for English, I mean we've got ... a really fantastic (English Test outcome for school) especially in the Year 7 with the writing because that was an area that (the teachers) were really, really focusing on due to previous results so and there's been a marked improvement (lately). ... over four to five years we've worked together, we've tightened and tightened and tightened it so that now what I'm doing is leading very well to what she's doing.

Some changes that were hoped to have impact on future outcomes provided evidence of much deeper approaches to curriculum change prompted by Test results but based on complex teaching approaches and evidentiary bases.

... in the past all of our Maths programs have started from a 2-d theme and then gone into 3-d whereas research has shown ... that it's much better to go from a 3-d theme, because that's what we see all the time (to a 2-d and maybe that's where the problems have come from ... we may not see the results in our Test results this year because we've only been doing it for six months, but over another year or two when kids start coming through who started on a 3-d theme and then progressed into a 2-d, maybe we will start to see some improvements.

I'll certainly be working a lot more on reading comprehension, I'd like to push that a lot higher. They're writing, I've introduced a new thing since the Test was done, I'm daily writing with them, which I believe has made a big impression and has also made a big improvement on their spelling because the sentences that we write, we also have them as dictation at the end of the day. So it's on the board in front of them all day and then the board's turned over and they have to write that as dictation, and I've found that the kids have really pulled up their act as they've worked through those. ... it's (a change in) the way I've been doing it. I've used the two different methods, one where I just write up a fairly complex sentence on the board and we analyse how the sentence is put together, what it means, and then they have to write one using a similar format. We've also done where I put up a short sentence and we've expanded it, adding phrases, clauses, adverbs, adjectives that type of thing. We've also taken sentences from storybooks we've been reading, and looked at those, looked at how the author has constructed those sentences and then tried to write one modelled on it. So I've used a few different formats, they've all been about looking at good sentence structure and trying to model it and copy it themselves. ... I think we've improved our writing considerably using that, and their spelling, their actual spelling in their writing is improving.

... (as) a teacher generally, if I had Year 5 again I think it would be helpful cause I'd remember ... this is an area that they cover within the Test, I've got to make sure that I don't just forget about that, I've got to make sure I cover it ... it just improves my way of teaching something. I think, all right, I know I've got to attempt that from a different angle and try and get the children to understand it differently or teach it more, r...I mean Number's something that's taught a lot but I think Measurement and well, I noticed that Space is a big area that is-the children haven't got so much a grasp of ...

In some cases, the discussion showed that the form of the Tests was having direct impact on the classroom content, with instruction becoming tailored to match Test content.

More likely as a year level...particularly if most or all of the same teachers are teaching that year level the following year and by the time we've been getting this data we will know who's teaching that particular year level so we'll have a look and see how we went this year and then we'll look at where perhaps the kids were pretty bad or pretty good ... what we thought they might do and improve on that for the following year.

... as a whole we might say well we need to work on our writing because there were a lot of kids who didn't do so well in the demand writing aspect or we might say well the reading comprehension, and this is what we've noticed in the past, both of these areas, the reading comprehension is so different from stuff that we've done in the past that we really need to do a fair bit of work in that area. We might look at different aspects of Numeracy, I know we've picked up on the fact that we've been weak in Space in some years in the past and we've looked at that and worked on that, they've been gradually picking up in that area. So that's usually what happens. It does tend, up until now, I can't say what's going to happen in the future but it has tended up until now to be a 3/5/7 issue...whether that's right or wrong (who knows).

Identification of learning support needs for individual children

Although not frequently part of the teacher talk, some use was made of the Tests data for looking at the individual needs of children. Often, to maximise effective use of the limited support resources available in a school, students would be grouped by need through examination of the area or item data and student performance.

Now what tends to happen here a little bit is that we got the summary of all of the children in a tabular form and we find that very useful and from that summary then we make our own tabular format to identify who needs help with what. So we've got our target group, so we can plan our strategies from there, basically. We find that the summary (part) is very useful to us as opposed to the individual ... from that table (we will) target groups of children who need help and we're thinking too, we're going to be doing ... some rotation ... so that we also are able (to provide some more time) with the children.

We will (devise) some kind of numeracy program because basically those kids (have) no number at all.

... having a support plan put together for her. I grabbed their Year 3 Tests, and in Year 3 (on) the *Aspects of Literacy* she was right down, couldn't have gone any lower in writing, spelling, reading and viewing, and in the indicators nothing basically on the dictation items and basically all she could do was locate specific information in written and visual text and identify the purpose in a set of instructions in a recipe. Pretty much that was all, and understand an idea in a specific part of a poem, that was about all she could do

This was usually the role undertaken by Learning Support teachers, who emerged in the project as individuals who were proactive in using the data as much as possible to identify students with learning needs and to work with teachers in planning supportive learning interventions at the start of the school year.

We actually make sure we get (the Test data) at the start of the year ... we wanted that data because basically it's one of those ways that we can pick up on kids and I mean just this year alone I would have picked on, just say off the top of my head forty kids that were below benchmarks unless I had that data in front of me I would have had (to wait) to see if a teacher referred that child, to see whether the teacher picked it up but because I had that data in front of me, oh look at this reading score, there's something going on here. I can then go and grab that kid and say I have-this child didn't do well in the Year 3 Test, can I grab them for a quick reading test, ... and ... I think every single one bar I think two that I tested ... and those two were like borderline and I just thought we'll monitor them. Every single one was like a definite.

One comment indicated the interpretive pressure that can be placed on individual item responses on the Tests.

... it is a lot of work to identify what the demands of the (children's needs are) so (Learning support teachers are) inferring backwards from the items.

Reporting to parents

Principals and teachers commented on the student results on the *Year 3 and Year 5 Aspects of Literacy and Numeracy Tests* that are provided to parents. No teacher said that parents had indicated the data were useful or informative about their child's achievements. Teachers reported that the information to parents was confusing, caused some parents some stress who visited the school with concern. The minimal information provided, with the removal of item performance and descriptions, was regarded as less useful than before. Some teachers were protective of the information and felt it was important to discuss the children's reports with parents rather than sending them home.

(My class Test reports) aren't just going home because I want to see as many parents as I can, so I can explain the Year 3, this is their first like systemic test type of thing, so I would like to talk to my parents about what it actually means.

... it doesn't really tell you much and even (for a parent) previously ... have all of the Test items ... parents don't even get that any more ...

Parents can't understand the graph ... some ... whose children were all below average but because they were above that black thing (benchmark) they think their kids are (are going fine).

That's fine if (there is) a conversation between the teacher and the parent but what about the parents who don't come up, the parents you never see ... you can't have that conversation.

Schools noted that from 2006 they will be the mechanism for identifying students who should receive the Federal tuition vouchers and the organisation of provision. They were generally not aware of parent engagement with the vouchers, although some teachers described the barriers one family had met to accessing the funds.

... he qualified for (the voucher based on) his Year 3 results and we ... helped the parents to go through all the process to get the funding, of course they couldn't find a teacher cause John lives ... and eventually we got (an appropriate tutor identified) and they agreed to accept her but by the time they got that organized it was September last year and the parents had put the forms in before Easter last year and then they turned around and said you've got to do it all before the end of November ...

Publication of results for school

Pressure to work with the results has increased for schools with the Australian Government requirements for publication of school outcomes.

... I look at that information (how school does in Test) because I've got to publish it on the web site ...

However, concerns were voiced by teachers, principals and some systems that public reporting makes no consideration of the differing school intakes of students and distance travelled. Suggestions were made that if such data were to be used, appropriate entrance assessments were needed to demonstrate children's development and value adding. Within the schooling sectors there was more recognition of such diversity, and looking at the practices of schools that were outperforming or underperforming in comparison to schools with 'like' intakes.

...this is what this information is saying about your school, ... what are you doing about this, how does this translate down to your year level discussions and/or dialogue and how are your staff responding to this and what are you doing to ensure that your staff respond to this now ... we generally put it in the hands of the principal as a school set of data and we do a display (of) trends (for) a district set of data, we've put up the entire primary cohort and put that on display (albeit) with schools coded. ... we've given the schools a take home folder with your letter is x and you've just discovered you're a bookend in the district and now you're either at this end or that end of the shelf ... we're trying to pick up those as exemplars of (good) practice (as well as examining those that appear to be underperforming).

With more testing, national testing, and increased curriculum testing, such pressures on schools will only increase.

Summary

The information provided by principals and teachers about their interpretation and use of the outcomes of the *Years 3 and Year 5 Aspects of Literacy and Numeracy Tests* was very revealing. The most prevalent onsite use of the Test data was to direct instruction at the classroom level for future cohorts. In some cases, even Learning Support teachers were directing their attention to this goal. While this may achieve the goals of the National Plan to ensure all children are literate and numerate, the goals will only be achieved to the extent that the Tests are validated assessments of these constructs. It is evident that the Tests outcomes are becoming more high stakes for schools, particularly with the publication of school outcomes in public forums. There is an underlying sense, even in such early implementation of public reporting and school comparisons, that the Tests outcomes are becoming the focus, not the quality of the student learning experience to attain high outcomes.

There was also evidence of deeper curriculum planning to address learning needs. Such curriculum planning overlapped syllabus and Tests contents. For literacy, the instruction was on writing and enhancing sentence structures and spelling, aspects of literacy. For mathematics, the discussion tended to focus on Number, as needing more improvement, and Space as needing more attention. Such aspects clearly relate to the Mathematics syllabus.

Less evidence was provided by the principals and teachers about the use of the Tests data to inform the learning of individual children. Some references were made to use of the Tests data to plan interventions with learning difficulties. No references were made to use of the Tests data to identify children who needed extension activities. However, having made these comments, in general the impression given by teachers was that they knew or could identify student learning difficulties and were already offering many extended learning or assessment opportunities to students who were high achieving.

The lack of usability of the item level information may be a factor in planning for classroom or individual instructions. The lack of clear content validity and curriculum connection and diagnostic capacity of the Tests reduces their usability. Further, the form of provision of Tests data and lack of assistance in manipulating the data hamper the usability of the Tests feedback in schools. The reports to parents are considered almost an insult to parent at best, or confusing and worrying, at worst.

Schools indicated limited use of data for tracking the progress of individual students. They were aware of the school tracking data and reflected positively on such evidence of growth. However, tracking of individual children through the use of unique identifiers did not seem to occur. Further, the schools did not seem aware of the potential for systematically examining such Test data, in conjunction with their own assessment judgments, to examine the effectiveness of different forms of intervention activities. Student achievement, needs and interventions appear to be operating as independent events, unable to inform future planning and resource allocation, within or across schools.

Finally, the degree to which schools did share school wide Tests performance data and classroom and individual student data was directly related to the leadership undertaken by the Principal or by a critical other, such as a Learning Support teacher.

Many of these factors can be addressed to enhance the value of the *Years 3 and Year 5 Aspects of Literacy and Numeracy Tests* to schools, albeit as one piece of the puzzle in a folio of student achievement.

System level discussion with administration staff from the three schooling sectors

Many of the comments made by the teachers led to further questions on the role of the systems in assisting schools to interpret and use Tests data from the *Years 3 and 5 Aspects of Literacy and Numeracy Tests*, and the provision of data and professional development to schools. We held informal discussions with administration staff for each of the three schooling sectors in Queensland: government, Catholic and independent. These individuals were volunteers who were prepared to talk with us about the Tests and school use. They did not provide comments that are to be interpreted as official policy, nor as representative statements of their sectors. However, their comments were very useful for siting the discussion of schools, for showing the types of support that are made available to schools, and for discussion of further resources that could be available to schools.

How system administration staff interpret and use the *Years 3 and Year 5 Aspects of Literacy and Numeracy Tests*

Interpret

All administration staff regarded the *Year 3 and Year 5 Aspects of Literacy and Numeracy Tests* as a now incontestable part of the curriculum Australian and state education policy domains—such testing is required, reporting to the community and parents has to occur, schools receive information that should be used to improve learning and student outcomes. Because of federal and state legislated requirements, the administration staff acknowledged the need to assist schools in all these activities. The system administration staff did not make comments on the literacy and numeracy constructs underpinning the *Years 3 and 5 Aspects of Literacy and Numeracy Tests* or their relationships to KLA syllabuses.

System administration officers focused on two aspects of the data: ‘means’ for comparing school performances, and the patterns of responses to items to examine curriculum and teaching matters. They used these to provide comparisons, where possible, for ‘like school’ groupings.

Use

Administrators from Education Queensland indicated that their priority purpose was to engage with principals to enable teachers to use the data. Staff visit schools and discuss outcomes with the principal. Accountability was a major focus, although a ‘big stick for underperforming schools’ (determined through ‘like school’ data) was an underlying theme.

The administrators from the Independent school sector were not able to develop ‘like schools’ comparisons as they do not receive system data. While this is currently under investigation, the major role the Independent schools administrative staff could play was to support schools in their use of the data. Workshops with teachers analysing their own school data have been a major activity, often working with an independent consultant who has developed software to help data interrogation.

The representatives from the Queensland Catholic Education Commission noted that the issue was still under consideration as to the appropriateness of system level analysis. There is not one Catholic school system in Queensland but five diocese and several independent Catholic schools. However, given the large number of schools that are aligned with the Queensland Catholic Education Commission, some limited population analyses are undertaken.

... we don’t also necessarily have socio economic data or postcodes and things attached to that data (so) it’s not easy for us to do analysis of things like comparisons with like schools, for example. ... (what we can do is) reasonably limited. We do do some correlation to some

other census data that we can import and compare with it but it's ... pretty basic analysis \. We do do some analysis at that level ..., particularly in terms of the Indigenous population. We certainly do a bit of gender analysis, a little bit of trend data, but again very much at that global level not looking at school groups or cohorts or certainly not individual students, even though we have the data down to the individual student.

While some analyses were undertaken comparing different population groups, item level information was used at the system level as federal funding could be used to direct programs for schools. Such reports were prepared for senior management including leadership teams and area supervisors.

We do use it to inform our decisions about professional development and focuses within our Literacy funding So we'll be saying, well we'll focus on Writing because relatively that was an area of weakness perhaps in a particular year or across a few years. ... the data is much more useful from that sort of cohort level or item level than it necessarily is in terms of student level.

The expectation of school planning in this sector was that school planning could demonstrate how student learning data in literacy, numeracy and core learning outcomes, including Test data as one part of the picture, was used to inform teaching.

Use of Years 3 and 5 Aspects of Literacy and Numeracy Tests data

All sector commentators confirmed that school practice in using and interpreting the Tests data varies, in their views dependent on the activities of the principal, who for each sector was seen as the critical leader.

... some schools are at a very high degree of sophistication, they do their data analysis, they have workshops with their teaching staff on well, this is our data, what do you think it says, how will we plan for the next year, down to the people who never open the corporate data warehouse, they file their reports they get from QSA as soon as they come and if I ask them for them they don't know where they are and they wait for me to tell them what their mean score was to write on their AOP (Annual Operation Plan).

The change of school for students at the end of Year 7 in Queensland provides a systematic barrier to provision of data on children from Year 7 to Year 8. While many Catholic and independent schools operate from early years to Year 12 on the one campus, and many new Education Queensland schools are being developed in this way, most current government high schools are still physically separated from primary schools. Information does get transmitted to high schools although not necessarily systematically.

I had actually given the high schools doing their triennial school review this year an analysis of their cohorts for the last two years on the Year 7 Test and what they looked like. The problem with that is, is that the system only provides that to them after the end of February when the census comes in ... but by then it's halfway through first term and they've basically already got the kids and already ... put them in class or ... got their own idea about them so it's ... the logistics of getting that information.

However, the usefulness of the data for individual children and planning was considered limited for all due to the timing of the Test data.

I think the schools would consider the reports they get on individual students, more often than not these days they include it as one piece of data in a portfolio. They conceptualise it that way, which I think is a good way to deal with it rather than putting it in as a report card so that it almost looks like an alternative to the report card, whereas a piece of evidence in a portfolio accompanying a report card is a better positioning of it, but in terms of analysing the data, partly because of that time lapse, its usefulness to inform an

individual student's progress is a bit limited. It can be a good check and balance though with other data, of course.

Education Queensland has the advantage of being a clearly identified schooling system. The representatives indicated more systematic use of the Tests data available with reports collated and shared with schools at district level, within cluster groups and at individual school level. Trends over time were a specific focus for looking at performance. There was a sense that schools were becoming more 'data smart' and no longer apprehensive about such data but interested in using it for school and instructional improvement. The value for individual teachers was in the link to their own classroom assessment practices and outcomes.

However, from the discussions there was also the sense that the system administration staff are making the Year 3, 5 and 7 Tests more high stakes, particularly in the state school sector, and are using these as a means to identify schools that are underachieving.

The 'data guru'

The theme of the critical value of a 'data guru' in an individual school occurred across discussions with the administration staff from the different sectors, reflecting the discussions that had occurred with staff at individual schools. Effective use, even minimal effective use, in a school was seen to be dependent not just on the principal's leadership but also on the presence of a person with interest and 'know how'.

... somebody in the school who has the motivation and the expertise to manipulate this data and do something with it.

(use of the Test data is) varied across schools ... linked very closely to the technology skills of people on staff. If you've got someone who's interested in data analysis or interested in computer programming and being able to store information that way, well then those schools might have made advances but unfortunately in some schools there isn't really anybody who's got a bent for that

Recommendations

12. Principals and curriculum leaders need to undertake and be able to undertake stronger roles in examining implications of data for planning for students' learning gaps for Years 4 and 6, and also pedagogy for Years 3 and 5 where appropriate.
13. Schools should discuss the Test data across schools in the context of other assessment data and consider the constructs of literacy and numeracy assessed by the Tests.
14. Data should be provided to schools in electronic forms to assist in data interrogation.
15. Strong representations should be made federally to adjust testing to early in school year so that the original purpose and value adding to student learning can be realised.
16. If schools develop longitudinal databases of student achievement, they should incorporate intervention histories to explore patterns of effective instruction.
17. Professional development is urgently needed to assist schools in using data effectively for school planning and individual instruction. A 'data guru' is also needed. These are discussed further in the later section on Resources.

Issues for students identified as at risk or in danger of being at risk by either teacher judgment or the *Year 3 and Year 5 Aspects of Literacy and Numeracy Tests* data or both.

T1: ... nearly all (25 to 27) of (the children in class) have learning difficulties to some extent. ... the Year 5 Test definitely showed their results...

T2: (It) definitely reflected that they were learning difficulty kids ... definitely still highlighted that they're low kids (from Year 3 to Year 5)

Interviewer: So the Test is a confirmation especially for at risk students?

T1: Definitely.

T2: Absolutely.

From analysis of teachers' judgments and the Test outcomes, even for this small group of schools and students, the indication is that students identified by the Tests as below the respective literacy or numeracy benchmark standard, as defined by the Tests' cut scores, were clearly identified by teachers as experiencing literacy and/or numeracy difficulties.^v These are the students for whom schools gain funding for intervention and for whom the federal tutor allocations are available. However, from the discussions with teachers and our examination of the student work samples, a number of students, even within this small sample, were considered by teachers to be highly at risk, possibly below a qualitative interpretation of a benchmark standard, yet obtaining a Test outcome above the benchmark standard. Students exhibiting very limited achievement in school activities were often rated between the benchmark and 25th percentile. While a proportion of children must necessarily rank below the 25th percentile (that is, 25 percent), the discussion with teachers and work samples tended to confirm that these children were not achieving in literacy and/or numeracy.

The image shows a student's math worksheet with the following content:

Using and Extending Mental and Written Addition Strategies

Work out the answer in your head. Record the steps you used.

a. $54 + 37 = 110$
 $30 + 30 = 70$
 $4 + 7 = 11$
 $= 110$ x

b. $36 + 48 = 80$
 $6 + 8 = 13$
 $30 + 40 = 70$ x

c. $67 + 15 = 12$
 $60 + 10 = 70$
 $7 + 5 = 12$ x

d. $78 + 35 = 113$
 $70 + 30 = 100$
 $8 + 5 = 13$ ✓

e. $86 + 57 = 132$
 $80 + 50 = 130$
 $6 + 7 = 13$ x

f. $234 + 57 = 291$
 $200 + 0 = 200$
 $30 + 5 = 80$
 $7 + 4 = 11$ ✓

g. $256 + 128 = 384$
 $200 + 100 = 300$
 $50 + 20 = 70$
 $6 + 8 = 14$ x

h. $328 + 257 = 585$
 $300 + 200 = 500$
 $20 + 50 = 70$
 $8 + 7 = 15$ ✓

i. $165 + 227 = 392$
 $100 + 200 = 300$
 $60 + 20 = 80$
 $5 + 7 = 12$ x

2 © Check Unit 30, Unit 2 1/3 Date _____
 yr 3 below

Sample mental and written addition strategies of 'below' level Year 3 student

Yr 6 'below' 27-3-06

Subtraction review

1 Use counting on to answer these.

a 39 - 35	b 40 - 38
c 88 - 84	d 97 - 93
e 71 - 68	f 79 - 71
g 68 - 63	h 45 - 39
i 60 - 56	j 52 - 49
k 91 - 86	l 63 - 58

2 Use counting back to answer these.

a 78 - 2	b 65 - 3
c 56 - 10	d 83 - 20
e 44 - 11	f 73 - 23
g 65 - 12	h 77 - 21
i 33 - 12	j 41 - 3
k 52 - 4	l 95 - 14

Counting On

Example:
75 - 68 =

Count on from 68
'69, 70 and 5 more'
The answer is 7

Counting Back

Example:
81 - 32 =

Count back 30: '71, 61, 51'
Count back 2: '50, 49'
The answer is 49

3


a $\begin{array}{r} 84 \\ -23 \\ \hline \end{array}$	b $\begin{array}{r} 79 \\ -55 \\ \hline \end{array}$	c $\begin{array}{r} 63 \\ -33 \\ \hline \end{array}$	d $\begin{array}{r} 47 \\ -12 \\ \hline \end{array}$	e $\begin{array}{r} 98 \\ -24 \\ \hline \end{array}$
f $\begin{array}{r} 60 \\ -27 \\ \hline \end{array}$	g $\begin{array}{r} 42 \\ -15 \\ \hline \end{array}$	h $\begin{array}{r} 37 \\ -18 \\ \hline \end{array}$	i $\begin{array}{r} 83 \\ -46 \\ \hline \end{array}$	j $\begin{array}{r} 91 \\ -23 \\ \hline \end{array}$

4

a $\begin{array}{r} 137 \\ -24 \\ \hline \end{array}$	b $\begin{array}{r} 514 \\ -204 \\ \hline \end{array}$	c $\begin{array}{r} 815 \\ -712 \\ \hline \end{array}$	d $\begin{array}{r} 246 \\ -35 \\ \hline \end{array}$
e $\begin{array}{r} 625 \\ -155 \\ \hline \end{array}$	f $\begin{array}{r} 907 \\ -94 \\ \hline \end{array}$	g $\begin{array}{r} 534 \\ -329 \\ \hline \end{array}$	h $\begin{array}{r} 560 \\ -80 \\ \hline \end{array}$

5

- There were 628 people in the hall. 98 were dancing. How many were not dancing?
- We had 352 fireworks. We used 246. How many were not used?
- Rachel has 170 birds. How many canaries has she if there are 134 that are not canaries?
- Luke has \$354. Heather has \$194. How much more money does Luke have?
- I have 253 coins. How many more do I need to have 600 coins?
- When the ship began sinking, 325 life-jackets were given to passengers. If there were 515 passengers on board, how many were not given life-jackets?



Number Study p24-31

Sample counting on and counting back subtraction strategies of 'below' level Year 6 student

Therefore, while part of the purpose of the Year 3 and Year 5 Tests is the identification of the percentage of children who have achieved the 'benchmark', the educational value of the statistically derived benchmark score, and the diagnosis of children's learning difficulties, may be questionable. As the introductory segment indicates, children who were diagnosed as having serious difficulties in Year 3 continued to be diagnosed as having serious difficulties in Year 5. While they may have made some learning gains, their core learning needs were still not able to be, or had not been, addressed.

The QSA emphasises that the Tests are tests of the full range of ability, within the psychometric parameters of construction. The relatively brief times that can be allocated for such tests for young children means that relatively few items of a very restricted nature must represent a broad range of ability. The questions to be considered are 'how many items must a child be able to answer to be rated beyond the benchmark' and 'to what extent might a child be rated above the benchmark by the probability of guessing'. We were not able to examine Test items and students' response patterns to determine whether students identified as 'above benchmark' were demonstrating knowledge that was indeed 'above benchmark'. A reason for some students with difficulties in literacy and numeracy 'fluking' the Tests may be the use of many multiple choice items for the Tests. Some children of lesser ability may gain higher ratings simply by guessing. The guidance in the Numeracy Framework resource document (QSA, 2006c) indicated at least 15 per cent of items should be at benchmark level. If the Test developers made these items mostly multiple choice, students have a high probability of guessing responses and achieving benchmark level.

Concern has been expressed among educators and policy makers that Australian children's performance on international tests demonstrates high average achievement, highly comparative performance for our highest achieving students, but often demonstrates a longer 'tail' of students

who are achieving at lesser standards (AEU, undated). While there may be many reasons for such a tail, when examined comparatively to other highly achieving countries,^{vi} the test data do indicate that there is a higher proportion of children for whom educational programs and interventions should focus than the percentage of children who do not achieve the 'benchmarks'.

The concern then, is the extent to which focus on the 'benchmark' achievements for the states distracts from more significant educational agendas of ensuring quality learning achievements for all students, especially those students whom teachers can identify as at risk, confirmed by work samples, who populate the lowest 25 per cent of classroom performance. Teachers were able to identify these children and quantify and elaborate their learning difficulties and needs. In our data, the Test data appeared to confirm their level of learning relative to other children.

And a lot of it was just letting the kids know that just to do their best, and obviously we will look at their results but we know what they can and can't do.

As noted previously, Year 3 and Year 5 Tests may not be as effective an indicator of students with difficulties as teacher judgment, supported by considerable evidence from day to day interactions.

Teachers were concerned that children needing additional support could achieve the benchmarks and either (i) were not eligible for additional funding and support or (ii) could be considered to be making adequate progress as evidenced in the Tests. The discussion with teachers indicated that resources to assist students, such as use of learning support teachers in and out of classrooms, teacher aides, parent assistance, programs, were fully utilised to the extent that they could be provided. However, the necessary prioritising of support to meet the students ascertained as needing most assistance, and in many schools, the low priority afforded many children with difficulties to get special support assessments, meant that many students at risk could be receiving very little additional support in the classroom.

In conjunction with this, teachers spoke of limited access to and use of external diagnostic and allied health assessments to ensure appropriate diagnosis and intervention. Barriers were the previously mentioned lengthy waiting lists, or lack of necessary parental approval. The consequence is that teachers and learning support teachers are trying to provide the best intervention available but without full information and assistance. The expectation would be that the children identified by teachers as 'at risk' but not attracting additional funding or support will continue to be identified by Tests at every year level to be at risk. They are unlikely to succeed in schooling and have the potential to be early school leavers, with all the consequential employment, health and associated crime risks.

Teachers were almost resigned to this outcome for the child where the school recognised that its potential to intervene for a child was very limited.

... the poor little fellow is almost illiterate and that continues to be of grave concern. We've tried to make contact with dad with no success at all and when ... been away, I've actually had a couple of letters from dad and dad's writing is very poor, he has a lot of spelling errors but I've also noticed that words ... (for 'on') he wrote 'no', and there were a couple of words like that which makes me wonder whether ... dad's possibly ... dyslexic. ...

(response to query if diagnosed dyslexic) ... No, because you see you actually need a medical diagnosis for that and ... only lives with his dad and even though we've made several requests to dad we've never seen him.

The child is now in Year 6. The parent is unlikely to activate the federal voucher program. There are few years available to intervene in this child's learning. With his current learning history he is unlikely to attend school beyond Year 9.

Funding and resources

Many teachers raised the issue of the value to students who are clearly having severe learning difficulties of being required to complete the *Aspects of Literacy and Numeracy Tests* even with in Test support. They noted that the impact on children was generally negative. Several teachers discussed the mechanisms taken to assist children, including isolating them from other children while completing the Tests. Teachers talked of their own distress, and of children's distress, that they could not provide assistance to children's queries. The percentages of children exempted from the Tests (reported as below benchmark) or absent or withdrawn (by parents) from the Tests (not reported in population base) varies from state to state. Queensland's exemption and withdrawal rates for children for Reading and Numeracy of 4.6 per cent and 4.0 per cent respectively, are the lowest in Australia and well below the Australian average of 6.8 per cent and 6.5 per cent.

However, money flows for children under the benchmark. The nature of such funding and links to benchmark performance appear to have changed over the last quadrennium. In 2006, Queensland provided \$91.68 'per incident' to a child in the lowest 15 per cent in the state on the Year 5 *Aspects of Literacy Test*,^{vii} while overall, Queensland received over \$62m DEST funding for the Literacy, Numeracy and Special Learning Needs Programme.^{viii} These funds are to be targeted to strategies to assist students who:

- Are not achieving or are **at risk of not achieving** a national benchmark standard of literacy and/or numeracy or other appropriate standard of achievement; and/or
- Require additional assistance to reach an appropriate standard of achievement. (Emphasis added)

The Reading Assistance Voucher programme of DEST provides a large sum of specifically targeted intervention funding, costing \$8.8m in 2006-2007 and 2007-2008.^{ix}

Queensland has developed a funding strategy following the Year 2 Diagnostic Net. This identification results from a process of teacher judgment against a developmental chart, with moderation processes. There is clearly scope to extend the diagnosis of children in Years 3 on to a process involving teacher judgment that is more informative than statements of 'lowest 15 per cent' on the *Year 5 Test*. In an era that promotes 'evidence based' policy, from the findings of this project, such a statement has limited import for identifying students at risk, and most importantly, their areas of learning need. The QSA and the systems' authorities need to redevelop ways to engage with what counts as evidence across all years of schooling, to investigate more educationally sound diagnosis and intervention of student need.

Many children need considerable instructional intervention and individual attention. Our conclusion is that the children below the benchmark, as indicated by the Test outcomes, are not the only children needing specialist assistance. Systems should examine the extent of learning support provision available to schools, and more provision of within school diagnostic processes. Supply of such resources was clearly less available to the schools in our project than demand. There is concern also that such support tends to be made most available in the early years of schooling and gradually less available as students progress through the year levels. However, the project clearly shows that problems are persisting through Year 5 and at least into Year 6. Research into the upper years may show similar pattern.

Schools are caught in the process of gaining funds for students at risk and improving student learning.

You're damned if you do and damned if you don't, if you perform too well then you don't get the funding. The funding supplies the teacher aides you put in to implement a lot of the program or to support the teacher and the program, if you don't have the program

then you get poor results. If you get poor results, then you get the funding, then you can implement.

Overall, there needs to be a clearer educationally based procedure for identification of appropriate in school and out of school support needs for students at risk. Considerable financial resources appear to be available. Their expenditure to assist students at risk in the essential areas of literacy and numeracy needs to be more transparent.

RECOMMENDATIONS

18. Teacher judgment and supportive evidence should be used in association with Tests data to identify children at risk not picked up through the external Tests. This reinforces the discussion in the section on how teacher assessments can be used to provide additional information on students' performance in literacy and numeracy and recommends that processes should be developed to allow teachers and schools to feed forward into student appraisals at the system level.

Standardisation of task and/or moderation processes needed to be developed to assist comparability of teacher judgments

From our findings about the 'idiosyncratic' nature of literacy and numeracy programming, and the apparent range of understandings of standards in each school, we propose that consideration be given to establishing cluster formations within regions in the first instance. This could be the basis for working in teacher teams, with curriculum leaders, to explore the features of quality assessment, and to collaboratively generate and share tasks and accompanying statements of standards in moderation or peer review activities. The motivation for this is to use assessment as a lever to improve learning, targeting the critical relationship between curriculum and literacy and numeracy.

A major conceptual shift is required for teachers to move to an understanding that literacy and numeracy demands are involved in all assessment activities. One way to move assessment practice forward, and to thereby influence and improve curriculum planning, teaching and learning, is to involve teachers in examining tasks demands, extending to making explicit the curriculum knowledges and literacy and numeracy demands that require teaching for students to achieve success.

This can be done, as shown in other research undertaken by members of the team, without enforcing rigid adherence to assessment task templates. Of more use is professional development inducting teachers into the principles of quality assessment practice, building on their current strengths in observation and consultation. Areas that need attention to assist comparability of teacher judgments are teacher knowledge about:

- ◆ how different types of assessment information, collated in portfolios, can be collected and updated over time using profiling;
- ◆ judgment models as they involve the use of stated standards;
- ◆ moderation or peer review models providing opportunity for teachers to review classroom work as evidence of progression, focusing in part on the match between student work and standards, and also on task design issues. Key to the latter, as indicated above, is the dual focus on assessment demands as being inclusive of curricular knowledges and curriculum literacies and numeracies.

This project presents the case that such provision is a necessary development for improving school accountability through more cross site consistency of reporting and for fulfilling the national and state goal of improving learning opportunities and literacy and numeracy outcomes for all children.

Queensland teachers in Years P-10 report in this study and in previous research (Wyatt-Smith et al., 2003b) limited opportunities to have conversations around quality within schools and often very limited opportunities across schools and within districts and clusters. As discussed earlier, moderation processes and practices are well established in the senior years (11 and 12), having been established as developed and installed as part of the state's move to standards based education dating back to the early 1980s. Further, there is some experience in moderation as practised in relation to the Year 2 Diagnostic Net, though this was variously reported, in some cases as negative, in the discussions about cross site reliability of judgment.

We note that teachers themselves did not initiate a call for standards for literacy and numeracy or moderation. Further, in some sites, when a possible move to formulate standards was raised by the project team, teachers in the main did not anticipate that they would be beneficial. Instead, they raised the concern that the standards may be used as a further 'accountability' mechanism, as though the latter was not only separate from, but perhaps a distraction from their primary concerns with teaching and learning, reinforcing the cultural position noted by Ingram et al.

(2004). However, when moderation was raised, it was seen as valuable opportunity for literacy and numeracy discussions to clarify expectations, confirm judgments and to critically reflect on the achievements and assessment activities of their own and other teachers' students.

While schools had developed Whole School Literacy Plans, there was no evidence in the teachers' talk or reported practices of how these linked to, or were informed by, discussions of literacy development across year levels. As yet, there is no requirement that schools develop Whole School Numeracy Plans, and no mention was made about school numeracy policies. As discussed above, the resources used, and especially schools' reported use of packages (in particular *GoMaths*) appeared to set standards by default.

Moderation is known to be valuable for teacher professional knowledge development. As recognised, it requires time, training, and is not resource neutral. If there is a policy requirement that teachers use standards to inform their practice, and report achievement against stated standards, there needs to be a review of moderation models and the functions of stated standards in each of those models. For example, as evident in the existing research, moderation can be undertaken in a range of ways including blind review, face to face with classroom examples, online or a combination of face to face and online, the latter involving bringing teachers together for simultaneous sharing of cross site examples and discussion. It is timely for moderation practices to be re examined as they could apply in P-10, and support the accountability reporting requirement for teachers to use stated standards. Digital forms of moderation, as examined elsewhere, have the potential to break the confines of narrow geographical boundaries, bringing teachers together for 'discussions' of standards not otherwise possible. In this regard, we note that rural and remote student performance on the current literacy and numeracy Testing programs is a reported concern, and teachers in rural contexts are often more welcoming of standardised assessments so that they can anchor their students' performance. Technological moderation, where feasible, has the potential to address this issue and enable the performance of students in isolated communities to achieve identified standards, concurrent with delivering professional development for their teachers.

Consideration of the resources needed to install moderation need to be situated in relation to the research finding that the provision of standards, in themselves, will not guarantee consistency of teacher judgment. As a move in maintaining public confidence in education, there is a clear need to support teachers as they work to 'deprivatise' judgment, examining for the first time in the state how their judgments are informed by and align with expected standards.

Drawing on the first set of findings, however, this project makes clear that the nature and purposes of moderation need to be clearly established, making space for all relevant data about student achievement. In relation to literacy and numeracy achievement, this would involve teachers looking at the contribution of and coherence between both the reported assessment data of literacy and numeracy achievement in large scale testing programs, and teacher generated and controlled assessments of literacy and numeracy in all Key Learning Areas. In the latter category, the multidimensional nature of teacher assessment information, extending to student engagement features, discussed earlier, should be overlooked or regarded as threatening 'objectivity'. Instead, it should be explored as offering new insights into learning development and growth in ways that have not been captured in statements of standards developed to date.

The main challenge therefore is to situate both literacy and numeracy Tests data and classroom data in relation to stated standards, using the exercise to inform teachers' work generally, and expectations about quality in particular. If standards become the linchpin centrepiece, then the two data sets could be used to generate meaningful indexes of achievement of students in Queensland, relative to national and international achievement data.

RECOMMENDATIONS

In proposing recommendations here, two main domains of QSA work in developments are recognised. First is work on statements of Essentials and Standards, as well as common assessment tasks and locally endorsed tasks, as part of the QCAR initiatives. Second is the current review of syllabuses with a view to repackaging in 2008, as well as the revision of the English 1-10 syllabus. It is also recognised that the QSA is the industry partner in an Australian Research Council Linkage project with a focus on moderation models, teacher judgment models and consistency. Given these directions, it is recommended that the QSA:

19. examine the literacy and numeracy demands embedded in statements of Essentials and Standards;
20. identify ways in which such demands can be made explicit for teachers, and in turn, inform classroom practice;
21. review the statements of Essentials and Standards for their implications;
22. identify ways in which both common assessment tasks and locally endorsed tasks convey to users clear information about both the curricular knowledges and literacy and numeracy demands involved in successfully completing the tasks.

Processes or resources to assist teachers, parents, schools, and systems to make more effective use of information provided by the *Aspects of Literacy and Numeracy Tests* and to assist in teacher judgment of literacy and numeracy achievement

Resources to assist with effective use of the Tests

Overview

Our discussion shows that information about the Tests and the outcomes of the Tests in terms of student achievement and also cohort performance are used as different units of analysis for different purposes. However, it is important that there should be a coherent theme underpinning such uses. Tests such as the *Aspects of Literacy and Numeracy Tests* offer two dimensions of information—achievement about those taking the Tests, and performance on and validity of the information being tested, at the item level and ‘bubble’ level. In an earlier discussion, we explored from the data the ways in which the sectors, schools and teachers indicated they used, interpreted and responded to the information provided by the Tests. In some cases, this was the result of explicit questioning about their Tests data use. However, first, we need to discuss issues that interpreting and using test information raise in general.

Overall, data interpretation should be collatable from the individual student to the broader class and cohort of interest. While the original purpose of the Tests as sample tests was to provide accountability and cohort performance in order to inform student learning and progress, the implementation of the literacy and numeracy Tests for full populations has necessarily shifted this focus. Whether appropriate or not, the Tests are seen as indicators of individual student achievement on aspects of literacy and numeracy that are at least retrospectively identified as valued or significant. Interpretation of the Tests outcomes must focus on usability from the individual student as unit of analysis. If this is not achievable or meaningful, then by corollary, provision to students and parents of individual reports is not meaningful.

Similarly, patterns of response across items for students, classes, schools and cohorts of interest should be able to be examined and interpreted, if meaningful. Why is Numeracy identified as three substrands but reported as a single score? Is any Test outcome for a student a composition of performance across a number of important items or is it, as one would more traditionally expect with a standardised test, a score based on the summation of the items developed as representative of areas but with no single item able to carry much import for learning and instruction? Are the distracters for the multiple choice items logical distracters? Major test developers such as ETS in the USA are developing multiple choice tests where the choice of distracter by a child will provide insight into a common conceptual error or computational error held by the child. There is a need to make more explicit to Tests outcome users the degree to which interpretation of outcomes on single items is meaningful and to allow performance across the facets of literacy and numeracy to be explored to effectively inform instructional planning and interventions.

Understanding and interpreting the Tests outcomes

This project has not examined the psychometric validity of the Year 3 and Year 5 Test developments, and has referred only to curriculum and classroom practice commonalities. However, discussions about the use of data with principals, teachers and system representatives

indicate that there should be more information provided about the nature of the Tests and scaling undertaken. There is a propensity for both system people and school staff to take the Tests at face value, and to interpret a direct correlation between scores on individual items, total scores and bubble placements.

How meaningful are the individual items, how representative of the intended constructs for assessment are they, how do they relate to and reflect the benchmark and other literacy and numeracy standards? The QSA sources state that the benchmark indicators are cut scores imposed on the literacy and numeracy scales, a statistically developed standards referent point. The typical student performance that would result in a student failing to achieve the benchmark is not known. Teachers expressed surprise that some students whom they considered to have very severe learning difficulties had reached the benchmark.

Student bubbles are reported in terms of the performance of the cohort, a norm referenced point, but also against a scale that theoretically spans Year levels. Children may be reported individually as off the relevant Year cohort while the state analyses show the progression of cohorts on the common scale, based on common and calibrated items. A descriptive interpretation is not provided for this scale.

The scale against which individual student performance is reported for a cohort is also dependent on items correct and their statistically derived item difficulties. How well a student can do can be limited by the difficulties of items included from year to year, under such a scaling approach, although theoretically the student may be achieving at a level Years higher. This type of information is not provided to parents. The central indicator on reports to parent is a 'mean' for the cohort. The question is whether the representation to parents of student outcomes in such a simplistic manner, against a notional benchmark without measurement error, is statistically appropriate, whether reporting on a single scale without a performance description is educationally appropriate, and whether more information about the construction and calibration of the Tests should be provided to teachers and principals. Much more information about the nature of the Tests construction and interpretation of outcomes should be provided in the interests of transparency.

RECOMMENDATION

23. More transparent information on the construction and validation of the Tests should be provided to schools and parents.
24. More descriptive information on the scale of performance should be provided to schools and parents.

Clearer guidelines of content, focus and curriculum links

We have noted that specifications of the content of the Tests regarding the relationship to English, Mathematics and other KLAs and the demands of these syllabuses in terms of curriculum literacies and numeracies need to be clarified to assist in interpretation and use of the *Test* information. This then is the most essential resource development that needs to occur.

The current Handbook of Reporting is an informative resource for principals and teachers. However, it was not reported as broadly used, despite the enthusiasm of one system representative.

Do you know what I've always found really, really good is the *Test* reporting handbook and it often surprises me how few teachers actually ever delve into that.

The QSA should examine the nature of information provided in the Handbook and ways to disseminate the information in more immediately accessible and interpretable form, with the data.

RECOMMENDATION

25. The *Test Reporting Handbook* should be rewritten to be simpler and promote increased usage.
26. Information about the *Test Reporting Handbook* should be circulated more broadly to teachers and schools.

Forms of data for system and school analysis

The Queensland Studies Authority provides system level data to Education Queensland and to the Catholic Education systems, but not to Independent Schools. Education Queensland makes data, and some forms of analyses and graphical representations, available through the Data Warehouse. How widely these were shared by principals varied from school to school. Schools reported receiving their school's data information in hard copy. This made analysis of data by schools very time-consuming if they wanted to undertake their own analyses, for example, chart student growth.

RECOMMENDATION

27. Provision of all Test data to schools should be in electronic form to enable much more systematic use of the Test data.

Provision of assistance in interpreting and analysing data

During the data collection for the project, three resources to assist systems and schools in analysing and using data were identified. Firstly, when data are provided to schools electronically, software tools and guidelines to analyse the data can also be provided electronically. It was reported that such assistance is already provided to schools in New South Wales by the NSW Board of Studies.

(Following NSW example) ... we used to get printed and a CD and so the basic reports were done plus it gave you the tools to manipulate what you wanted to do with that. For example, I was saying to C, we could then go into this and flag particular students that we were interested in, so we were interested in tracking our Reading Recovery kids, so we could flag them as a group and then we could do-the tool would then, after we'd flagged them, and we'd just call them anything you like, it would then do an analysis comparing the Reading Recovery kids with their cohort with the state or whatever we wanted and that was really useful because, you know you might want to track-we might want to track refugee kids, we might want to track students who didn't start Year 1 at this school compared to students who did, you know. That can be an interesting analysis to do. So a tool...I just think it's got to be cheaper for QSA or some central authority to develop this sort of solution.

... the type of data analysis that comes to (a school in New South Wales) on the CD from the Board of Studies with the test results is very easy to use. A lot of the analysis is already done ... it's just a matter of clicking on buttons to generate graphs of particular ... analysis so you can analyse kids by gender and Indigeneity and ESL classification and I think you can even do things like date of birth ...

Smith (2005) describes the data management and assistance provided to schools in NSW, through the *School Measurement, Assessment and Reporting Toolkit* (SMART). SMART provides the means for principals and senior teachers in NSW schools the data analysis functionality to, among other aspects,

- create Custom Groups of students such as class groups and students involved in special programs such as Reading Recovery
- produce tables in PDF format, including school summary information and the ability to regenerate detailed reports on individual students for the information of parents and teacher
- analyse groups of students across the performance bands including the students in custom groups – providing evidence of improved student learning outcomes
- access to information concerning the performance of students on individual items at the group for school levels, including patterns of student responses and additional distractor information detailing the reason why students chose particular options. (p. 41)

New South Wales is also developing this toolkit to allow schools to include teacher assessments of students.

The school-based assessment module will form the nexus between external and internal assessments and has the potential to see the consistency of teacher judgements and hence the significance placed on school-based assessments, greatly enhanced. (p. 42)

Such a toolkit would assist the implementation of many of the recommendations in this report.

Tracking children and cohorts within schools was a common identified need for principals and schools. At present schools can examine the overall trends for their school cohorts from Year 3 to Year 5 and from Year 5 to Year 7. Principals and teachers would like simple access to a means to examine an individual child. Within Education Queensland, the allocation of individual Student IDs will assist such monitoring. In the other sectors, developing a method of giving students unique identifiers could assist in monitoring.

If the school got a graph showing ... the (individual) child at Grade 5 and then what their result had been at grade three, I think that would be really helpful.

RECOMMENDATION

28. Software tools should be provided to schools to assist in analyses and interpretation and to ensure that data analyses can be shared between principal and teachers. As a starting point the software developed in New South Wales and software already in use in some Queensland schools should be examined for broader applicability.
29. All school sectors should explore the potential for schools to systematically track student achievement longitudinally by using a student-specific ID and create a longitudinal database.

Professional development for principals and teachers

Test implementation in Queensland is still a recent phenomenon and the understanding of principals and teachers about the nature of the Tests data, their potential for improving learning, and simply how to access and manipulate the data are varied across schools. Professional development of the most basic kind was often requested. Principals and teachers felt there was still limited understanding of the potential of the data, let alone processes for its analysis. Professional development about the Test and data use was a common request across systems.

...this is the corporate data warehouse and here is how you log on to, you know this some way you can manipulate if you really want to know about the Year 7s are going with this, that and the other so ...

Allocation of specific time on student free days for principals and teachers to work with Tests data at the beginning of the following year, to plan for individual student intervention as well as cohort Tests preparation, was seen as beneficial.

Time at the beginning of the year to sit down and analyse the data.

(Professional development should address) the issue of processing, being able to use the data and the understanding, interpretation of responses wisely (to intervene for individual children) ...

...(so) we could use this more productively as one piece of the puzzle about what we're doing even as a state and a system better than we do.

The request was usually for people time, in small groups, to assist in development of principal and teacher understanding. As one principal said, professional training on CDs is not always effective, people time was preferred.

CDs that are released ... (are not) always so helpful (you can't ask it questions). ... just some people time where people actually talk to you about the process would be much more helpful. ... meetings (with other teachers) after school (for example, cluster).

Professional development for principals and teachers is clearly a needed resource. Effective professional development using other common resources has been occurring. The Brisbane Catholic Education Commission has published a special issue of its journal *Curriculum Matters* about using data towards the middle of 2005, as mentioned.

The Independent Schools Queensland and Brisbane Catholic Education Commission have held workshops for school staff about the Test data and its interpretation. BCEC is working with a private consultant who demonstrates a software program that schools can use with their own data. It is recommended that such programs are investigated by the QSA and programs suitable for use by schools could be documented on the QSA website or in newsletters to schools.

The follow up professional development that we have identified from the project is the provision of assistance to schools, principals and teachers about the use of individual, or classroom, Tests outcomes to develop appropriate intervention strategies for children, especially for children at risk. Such strategies require first that the Tests can be grounded in curriculum, and that the outcomes can serve a diagnostic purpose. What does a low score on spelling mean? Do the items and forms of assessment provide guidelines for spelling strategies that are weak or nonexistent? What does it mean to be below the 25 percentile in Numeracy?

RECOMMENDATION

30. Professional development on using and interpreting data from all sources should be a priority for school leaders and teachers. The professional development should build on the models of effective activities identified in the Review of Literature. The professional development should consist of assisted workshops and in school planning involving face to face interactions.
31. Time should be allocated in schools for teachers to discuss Tests data, in combination with other activities to be undertaken to enhance the literacy and numeracy assessment practices in schools.

As we have noted elsewhere, the timing of the feedback following the Tests is a barrier to their useful interpretation by teachers. The recommendation is that if a move to earlier testing can be negotiated with the federal government, more use can be made of the tests.

... it often comes down to time and the timing of them. Often by the time we get the results it's November or late November in the year and realistically the people who are getting the reports, the teachers of the children currently, are probably finished writing any report card they were going to write so it's not as useful for them then. At the beginning of the next year the kids are swapped to a new teacher, the first thing that they look at is not necessarily that report.

Information regarding other forms of standardised testing

During our conversations with principals and teachers, it emerged that many were making use of a range of other external tests and measurement devices. As discussed earlier, it is not evident how such tests related to the Queensland syllabuses, to levels and standard of performance, or how relevant some of the tests may be to Queensland student cohorts in the 21st century. However, teachers did feel a need at time to back up their own judgments or to obtain basal information about students through such means. While several recommendations have been made in this report about enhanced practices to assist teacher judgment of literacy and numeracy, it is still possible that schools will continue to make use of external standardised tests from time to time, particularly for diagnostic purposes. The QSA, in conjunction with system authorities, can play a role in assisting schools through a brief review of standardised tests, taking as a starting point the tests identified as in use in schools through this project (Appendix Five) and provide a brief annotated bibliography on: the match of such tests to Queensland literacy and numeracy policy directions and to syllabuses; the recency and location of norms of such tests and relevance to Queensland school children; and on additional diagnostic tools and resources that might be of assistance to schools. We recommend that such information should be easily available, perhaps as a web resource.

RECOMMENDATION

32. The QSA in conjunction with system authorities develop an annotated bibliography about external standardised tests, including information on purpose, norm development and appropriateness for Queensland curriculum and students to be hosted on the QSA website.
33. In conjunction with the above, the website should include links to standards for ethical use of commercially available standardised tests.

A specific request from some schools was the need for 'screening' tools for identification of student strengths and weaknesses on entering schools, or for placement of students. A common corollary of this request was that it would benefit comparisons of school achievements if basal measures of students on school entry could be used for some value added analysis and interpretation.

... I think (what would be desirable) to start off as a prep would be a generic screener across Queensland as well, just so every single teacher is on par with, for instance, ... knows according to her Test, you know whether her kids are up and down. ... just from day one ... so that teachers know ... especially first year (who lack experience). ... I can sit here and say ... I know what a Year 3 student looks like but that's only because I've been doing it for so long whereas, if I stepped into a Year one classroom I wouldn't have a clue because I haven't taught that ... I think every single teacher needs that generic base to go from ... cause our clientele is very different as well. So you can't say our Year 3s are the same as a Year 3 class down the (road).

Within this project, we do not explore the educational benefits and disadvantages of 'streaming' of children, nor the issue of use of standardised materials with very young children. However, we

do note that if such processes are current practice, then they should be informed through the best available advice on appropriate resources.

Developing a role for a key staff coordinator or 'data guru'

As noted in earlier discussion, several schools and teachers noted the advantage of having, or the need to have, a person who was allocated the role responsibility of working with the Tests data, the 'data guru'. This we found resonated with some of the literature reviewed.

... if you get a (person) on your staff (who) just loves playing around with (the data), well that's just fantastic but a lot of teachers are daunted by it in the first instance ... you know problem is an issue at Year 3 and it's an issue again at Year 5 and, you know so those sort of patterns I think are very useful for schools in terms of looking at where they need to improve.

We recommend that all systems should identify a person within schools who can fulfil this role and allocate a responsibility to them accordingly. If necessary, specific support personnel may need to be available to clusters of small schools. External tests data will be a part of school life in the foreseeable future, with the potential to grow in quantity if various planned federal and state assessment and measurement policy initiatives are implemented. Such quantitative data is not going to go away in the near future, its significance for schools is likely to increase in the near future. All schools need to know that they have a staff member, or resource person, who is effective in dealing with the data and communicating about the data and implications with staff. Such a coordinator should participate in special professional development to ensure their understanding and data manipulation skills. The implementation of the 'data guru' would mean shared access to data in all schools, independent of the skills of the principal.

RECOMMENDATION

34. The position of 'data guru' should be developed in schools, or for clusters of schools, with appropriate official recognition. A 'data guru' should be a volunteer, not conscript, as enthusiasm is an essential ingredient of the successful 'data guru' in schools in this study and previous research.

Resources for parents

The project participants were united in considering that parents received little effective communication about the Tests, their student outcomes and the meaning of these outcomes. They considered that the quality of the available information had even deteriorated in recent years, with parents not even receiving the individual item data as previously. Whether the provision of individual item data to parents should be restored is a policy decision. However, the information given to parents should be more informative about what their child can do and their learning strengths and weaknesses. To promulgate in the 21st century that parents only need to know the relative place of the child, albeit broadly represented by the bubble, fails to support evidence based research findings about effective instruction, motivation, learning and assessment.

RECOMMENDATION

35. Feedback to parents should include meaningful descriptions of students' standards and levels in relatively simple English with links to the relevant curriculum. Further, parents should be informed about how the Tests data fit within the portfolio of evidence being collected for the student, and the aspects of literacy and numeracy that are represented.

Resources to assist principals and teachers for school based student assessment

Teachers identified needs to assist their own professional development as effective assessors. Several initiatives are underway in Queensland. However, teachers were more specific in identifying their own needs in order to develop common understandings of appropriate standards, and suggesting ways in which this could occur.

I've done a couple of the assessment workshops. That's what my kids are working on at the moment but I struggle with writing criteria sheets because it's something I've only done in the last couple of years.

Some schools had already initiated their own activities but indicated ways in which further assistance could be provided through the QSA or system authorities.

There's something else that we do do as a cluster, locally, is we...we tried to start it last year but we've had it going this year is getting together and taking certain samples of student work and comparing them...doing our moderation basically and saying, you know this is what we think is a average for our Year 5s or this is what we think is...we were working on that. The next meeting we're doing writing, taking samples of student writing along and comparing them. That's the local assessment project that we're trying to get going. (small rural school)

A consistent resource request by teachers to assist in their own judgment of student work was processes to understand and moderate expectations of students' work against common standards. This might involve the type of moderation activities the school above had established with cluster schools, or 'like' schools across the state given technologically based capacity to share PDF files, or simply the provision of more exemplars of student work.

I also think it would be very valuable if there was an in service or some kind of professional development where teachers actually learnt criteria based marking. So actually learning how, what is an A or a B or a C or a D? What is actually suitable, and not just looking at the classes but then learning how to moderate, then at a glance ... could bring me something and at a glance I can say well, that's a B or a D

.. some sort of folio that has a sample of ... what we think an average year 2 should be like in literacy or in writing or reading or whatever.

... to me in terms of...a lot of it's just writing samples but we need to see, you know in terms of Numeracy as well, reading, what, you know where children are at in the state of Queensland we need to know that and I really think, yeah, I think it would be valuable.

Finally, although not identified by teachers and principals themselves, we note that teachers need professional development in bringing together the often disparate assessment evidence they have about children's literacy and numeracy. What does this mean to them as a whole, how does it represent literacy and numeracy as a whole, how is it linked to syllabuses and across syllabuses? The *Test Reporting Handbook* advises principals, teachers and parents to see the Test data as part of the picture, or the puzzle as our teachers indicated, regarding a child's learning performance. The resources provided to teachers should go further and assist teachers in the development of their assessment skills and understanding of curriculum to demonstrate how these parts should be combined to form the whole picture.

Previous recommendations have been made about the need to trial cluster school assessment and moderation activities to encapsulate literacy and numeracy standards and expectations. In conjunction with these, there is still room for further workshop activities with teachers on specific assessment practices such as combining judgments and on the Queensland policy documents on literacy and numeracy.

RECOMMENDATION

36. Professional development workshops on assessment should attend to the nature of holistic judgment, the combination of various assessment components towards the whole, and bases in Queensland literacy and numeracy policies, English, Mathematics, and curriculum literacies and numeracies.

Impact of federal tests of literacy and numeracy on discussion and recommendations

Originally we had seven research questions. However, during the conduct of the project some aspects of school practice, and issues of literacy and numeracy assessment, arose. The discussion has occurred in the framework of individual external tests of literacy and numeracy developed at the state level with some coherence to the state curriculum.

The discussion has focused on teacher judgment of the literacy and numeracy achievement of selected students in Years 3 and 5, with follow up in Years 4 and 6, compared these with curriculum frameworks, the curriculum frameworks of the *Years 3 and 5 Aspects of Literacy and Numeracy Tests*, and the results students gain from the Tests. The discussion raised issues of compatibility of the frameworks of both the Tests and teachers' judgments with the frameworks of the literacy and numeracy curriculum in Queensland. The discussion showed that teachers were able to identify students at risk, or who were high achieving, and those in between. Moreover, the discussion showed that teachers provided considerable information about the nature of students' literacy and numeracy strengths and weaknesses, and progress, within the narrow framework of literacy and numeracy achievement, and across additional strategic and engagement dimensions.

The project would confirm the general understanding in Queensland, Australia and internationally that teachers are competent assessors of children's achievement. Further the project confirms that teachers have rich, although varying, frameworks of assessment. The gaps are the identifiable frameworks against which to make their judgments. Their ability to judge students in ways comparable across the state will be enhanced by further professional development, opportunities to engage in professional conversations around assessment, development of cluster projects, access to and sharing of exemplars, and clearer frameworks, both curricular and assessment frameworks, on which to make such judgments. The potential exists to enrich the information collected at the system level about children's performance.

The project findings provide some concern that if the outcomes of the *Aspects of Literacy and Numeracy Tests* should gain higher status, and become more high stakes, they have the potential to narrow further instruction in literacy and numeracy in schools. Project findings also reflect the continued difficulty of the timing of the Tests, and the feedback on class and individual student achievement performance near the end of the school year. Even with prompt feedback, this timing means limited effective use can be made of the Test outcomes.

From 2007, national tests of literacy and numeracy will be introduced. Such tests will need to reflect common literacy and numeracy curriculum elements. By import, their implementation will direct teacher focus to the elements measured by the tests. For some schools under pressure, the pattern that has occurred internationally will occur here. The tests will constitute the de facto curriculum, reducing the breadth and richness of learning for students who are most likely those most in need of breadth and richness.

If the national tests replace the state literacy and numeracy tests, then the recommendations for the state systems to value and prioritise teacher knowledge and judgment of students' strengths and weaknesses across a broader framework gain in significance rather than being devalued. It becomes more important to clarify and scaffold the opportunities that are provided to teachers to identify valued curriculum outcomes and standards of student performance. The national literacy and numeracy tests will then hopefully become a mirror of much broader learning and success for all students.

We repeat the recommendation that the state systems should be lobbying the federal government to undertake any proposed national standardised assessments towards the beginning of the year. The return of information to schools, students and parents can then inform further instructional planning and intervention for all students in a meaningful way, ensuring effective progress towards the original objectives of the National Plan. Further, we recommend once more, the need for further research to monitor the impact of externally mandated tests on classroom practice and the breadth of students' learning experiences.

Conclusion

Throughout the one year investigation of this project, specific aims related to the exploration of teacher judgment of student literacy and numeracy performance were addressed in terms of the implications for the QSA in improving literacy and numeracy assessment practices and enhancing QSA policy development. The research focused on

the capacity of teachers to make judgments against defined standards such as the national benchmarks, and whether their judgments can provide more valid (comprehensive) and reliable (stable) information on student capabilities

and addressed

the question of best use of standardised-test data along with school-based data. (QSA Project Brief, 2005)

The investigation of teachers' use of school based and system assessment information has provided key insights into the practices and frameworks involved in teacher judgment of literacy and numeracy performance, the comparability of such judgments to student achievement on the *Years 3 and Years 5 Aspects of Literacy and Numeracy Tests*, and the use of this information by schools and teachers to improve learning. A major outcome of the project is the evidence to indicate that, in the absence of official common stated standards for literacy and numeracy, the nature of the personally located frameworks and standards used by teachers for literacy and numeracy assessment in case study schools was idiosyncratic and on the whole did not reflect the complex and cross curricular nature of literacy and numeracy described in official policy and syllabus documents. While the study concedes that the framing of the project interviews, with emphasis on the confirmation or rejection of Tests results as they were compared to teachers' own judgment of student performance, could have influenced teacher descriptions of their literacy and numeracy assessment practices, the fact that such descriptions and work samples were focused on narrow dimensions of literacy and numeracy provides the impetus for recommendations for policy work to further enhance this area. The project provides a foundation for providing policy advice to teachers about the evidential base that would be useful to them in working with system data, as well as their own assessments, to arrive at defensible judgments. This is an important point regarding the improvement of professional learning opportunities for teachers including the design of explicitly stated assessment criteria and standards, a focus on depth and range of curriculum literacies, the profiling of student work and moderation processes.

The project data has also demonstrated the marked comparability of school-based teacher judgment and external Tests results. This finding suggests that the capacity of teachers to make judgments against defined standards of the national benchmarks is well developed, even though the project did not establish that the teachers were familiar with the verbal descriptors of the benchmarks. Beyond this, of particular note is that in many cases, teacher judgment of literacy and numeracy achievement, particularly for students judged 'below' level and 'above' level, exposed broader learning dimensions related to processing strategies and engagement than the information reported in the Tests results. Further, teacher identification of students as 'at risk' using school-based assessments uncovered a significant group of students who, despite achieving Tests results above the benchmark, were clearly struggling in aspects of literacy and numeracy. This points to the vital role that teacher judgment of school based performance of literacy and numeracy needs to play as an augmentation to system data. More specifically, the project exposes how such teacher judgment, especially in relation to students deemed to be 'at risk' is a necessary and more robust means of working towards the objectives of the National Plan and gathering information regarding student learning and accountability than the Tests reports in and of themselves.

Finally the project clearly indicated the problematic nature of schools' strategic use of the Tests information supplied by the QSA and highlighted areas how the school and individual teacher's use of Tests data could be enhanced to impact learning in a number of ways. The timing of the Tests and the subsequent acquisition of the Tests results at the end of the school year together with schools' capacity to retrieve Tests data information in its present format were observed as major barriers for schools in working with the data to improve student learning. A related barrier was the ill preparedness of schools to track systematically individuals and cohorts, using the data over time, in longitudinal analyses of student performance. There is a clear need for upskilling school leaders and system personnel in how to support schools seeking to optimise the potential of the existing system data, and to examine its coherence with locally generated assessment information.

References

- Ashby, J. & Sainsbury, M. (2001). *How do schools use National Curriculum test results? A survey of the use of National Curriculum test results in the management and planning of the curriculum at Key Stages 1 and 2*. London: NFER.
- Assessment Reform Group (ARG). (1999). *Assessment for Learning: Beyond the black box*. Cambridge, UK: University of Cambridge.
- Australian Education Union (AEU). (undated). *Background paper 3 - Performance Indicators of Student Achievement (PISA)*. <www.aeufederal.org.au/PE/BP3.pdf> (Accessed 18 September 2006).
- Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-71.
- Boudett, K.P., City, E.A. & Murnane, R.J. (2006). The "Data Wise" improvement process. *Harvard Education Letter*, 11(4), 1-3. <<http://www.edletter.org/datawise.pdf>> (Accessed 29 June 2006).
- Brisbane Catholic Education Office, Brisbane Diocese. (2005). Using data. *Curriculum Matters*, 4(4), 1-20.
- Broadfoot, P. M. (1999). Editorial. *Assessment in Education: Principles, Policy & Practice*, 6(2), 173-175.
- Brown, G. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301-318.
- Burns, M. S. (1992). Changes in cognitive strategies in dynamic assessment. *The International Journal of Dynamic Assessment and Instruction*, 2(2), 45-54.
- Cognition & Technology Group at Vanderbilt (CTGV). (1992). The Jasper series as an example of anchored instruction: Theory, program description and assessment data. *Educational Psychologist*, 27(3), 291-315.
- Cromey, A. & Hanson, M. (2000). *An exploratory analysis of school-based student assessment systems*. <<http://www.ncrel.org/policy/pubs/html/data/index.html>> (Accessed 29 May 2006).
- Cumming, J. J. (2001). *Valuing learning and learners: The forgotten dimensions of educational assessment worth remembering in the 21st century*. Paper presented at the 30th International Conference of the International Association for Educational Research, Rio de Janeiro, Brazil, May 2001.
- Cumming, J. J. (2002). *A framework for assessing and reporting personal learning*. Paper Presented at the Second Conference of the Association of Commonwealth Examination and Accreditation Bodies (ACEAB) Malta 2002.
- Cumming, J.J. & Maxwell, G.S. (1999). Contextualising authentic assessment. *Assessment in Education: Principles, Policy and Practice*, 6(2), 177-194.
- Cumming, J., Wyatt-Smith, C., Ryan, J. & Doig, S.M. (1998). *The literacy-curriculum interface: The literacy demands of the curriculum in post-compulsory schooling*. Final Report. Brisbane: Centre for Literacy Education Research, Griffith University, and Department of Employment, Education, Training and Youth Affairs.
- Delandshere, G. (2002). Assessment as inquiry. *Teachers College Record*, 104(7), 1461-1484.
- Department of Employment, Education, Training and Youth Affairs (DEETYA). (1998). *Literacy for all: The challenge for Australian schools*. <<http://www.dest.gov.au/archive/schools/literacy&numeracy/publications/lit4all.htm>> (Accessed 22 August 2006).
- Department of Education, Science and Training (DEST). (Undated). *National literacy and numeracy benchmarks*. <http://www.dest.gov.au/sectors/school_education/policy_initiatives_reviews/key_issues/literacy_umeracy/national_literacy_and_numeracy_benchmarks.htm> (Accessed 22 August 2006).
- Education Queensland (EQ). (1999). *Queensland State Education 2010*. Brisbane: Department of Education and the Arts.
- Education Queensland (EQ). (2005). *School improvement and accountability framework. Destination 2010 action plan 2006-2008*. Brisbane: Department of Education and the Arts. <<http://education.qld.gov.au/trategic/accountability/destination/index.html>> (Accessed 8 August 2006).
- Education Queensland (EQ). (2006a). *Rich Tasks 2005*. <<http://education.qld.gov.au/corporate/newbasics/>> (Accessed 22 August 2006).
- Education Queensland (EQ). (2006b). *Literacy framework for action 2006-2008*. Brisbane: Department of Education and the Arts.
- Freebody, P. & Luke, A. (1990). Literacies programs: Debates and demands in cultural context. *Prospect: Australia Journal of TESOL*, 5(7), 7-16.

- Freebody, P. & Wyatt-Smith, C. (2004). The assessment of literacy: Exploring the tension between 'system' and 'site' validity. *Journal of Educational Enquiry*, 5(2), 30-49.
- Fuller, M. (2000). *Teacher judgment as formative and predictive assessment of student performance on Ohio's fourth and sixth grade proficiency tests*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.
- Goodman, Y. (1978). Kidwatching: An alternative to testing. *National Elementary School Principal*, 57, 41-45.
- Green, B. (1988). Subject specific literacy and school learning: A focus on writing. *Australian Journal of Education*, 32(2), 156-179.
- Hayward, L. & Hedge, N. (2005). Travelling towards change in assessment: Policy, practice and research in education. *Assessment in Education: Principles, Policy & Practice*, 12(1), 55-75.
- Hodges, C.A. (1992). *Teacher judgments and standardized assessments*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.
- Hoge, R. & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59(3), 297-313.
- Ingram, D., Seashore Louis, K.R. & Schroeder, R. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *TCRecord*. <<http://www.tcrecord.org>> (Accessed 29 May 2006).
- Kellis, M. & Silvernail, D. (2002). *Considering the place of teacher judgment in Maine's local assessment systems*. Maine: Center for Education Policy, Applied Research, and Evaluation, University of Southern Maine.
- Kenny, D. & Chekaluk, E. (1993). Early reading performance: A comparison of teacher-based and test-based assessments. *Journal of Learning Disabilities*, 26(4), 227-236.
- Kirkup, C., Sizmur, J., Sturman, L. & Lewis, K. (2005). *Schools' use of data in teaching and learning*. (DfES Research Report 671). London: NfER, Department for Education and Skills.
- Lesley, A.R. & Nelson, M. (2004). How teachers' professional identities position high-stakes test preparation in their classrooms. *TCRecord*. <<http://www.tcrecord.org>> (Accessed 17 May 2006).
- Luke, A. & Freebody, P. (2000). *Literate Futures: Report of the review of literacy in Queensland schools*. Brisbane: Department of Education and the Arts.
- Matters, G. (2006). *Using data to support learning in schools: Students, teachers, systems*. Camberwell, Victoria: ACER.
- Meiers, M. (2005). Getting It Right symposium—Evaluation of the Getting it Right literacy and numeracy strategy in Western Australian schools. Proceedings of the *ACER Research Conference, 2005, Using Data to Support Learning*, 72-83. <http://www.acer.edu.au/workshops/documents/Proceedings_000.pdf> (Accessed 17 August 2006).
- Ministerial Council for Employment, Education, Training and Youth Affairs (MCEETYA). (1989). *Hobart Declaration on Schooling*. <<http://www.mceetya.edu.au/mceetya/default.asp?id=11577>> (Accessed 22 August 2006).
- Ministerial Council for Employment, Education, Training and Youth Affairs (MCEETYA). (1997). Information statement 6th MCEETYA meeting, Melbourne, 14 March 1997. <<http://www.mceetya.edu.au/mceetya/default.asp?id=11510>> (Accessed 22 August 2006).
- Ministerial Council for Employment, Education, Training and Youth Affairs (MCEETYA). (1998). *A review of the 1989 common and agreed goals for schooling in Australia (The Hobart Declaration) discussion paper*. <<http://www.mceetya.edu.au/mceetya/default.asp?id=11578>> (Accessed 22 August 2006).
- Ministerial Council for Employment, Education, Training and Youth Affairs (MCEETYA). (1999). *The Adelaide declaration on national goals for schooling in the twenty-first century*. <<http://www.mceetya.edu.au/mceetya/nationalgoals/natgoals.htm>> (Accessed 22 August 2006).
- Ministerial Council for Employment, Education, Training and Youth Affairs (MCEETYA). *National Report on Schooling in Australia 2004. Table c.1*. <<http://cms.curriculum.edu.au/anr2004/index.htm>> (Accessed 6 August 2006).
- Nisbet, S. (2004). *The impact of state-wide numeracy testing on the teaching of mathematics in primary schools*. Paper presented at the 28th Conference of the International Group for the Psychology of Mathematics Education.
- Perry, N. & Meisels, S. (1996). How accurate are teacher judgments of students' academic performance? *Working Paper Series*. Washington DC: National Centre for Education Statistics.
- Protheroe, N. (2001). Improving teaching and learning with data-based decisions: Asking the right questions and acting on the answers. *ERS Spectrum—Summer 2001*. <<http://www.ers.org/spectrum/sum01a.htm>> (Accessed 22 August 2006).
- Queensland Government (Department of Education and the Arts). (2005). *Queensland Curriculum, Assessment and Reporting Framework. Strategic Policy and Education Futures*. <<http://education.qld.gov.au/qcar>> (Accessed 8 August 2006).
- Queensland Studies Authority (QSA). (2000). *Studies of Society and Environment Years 1-10 syllabus*. <<http://www.qsa.qld.edu.au/yrs1to10/kla/sose/syllabus.html>> (Accessed July 20 2006).

- Queensland Studies Authority (QSA). (2001a). *Literacy position paper*. Brisbane: Queensland Schools Curriculum Council. <<http://www.qsa.qld.edu.au/yrs1to10/oia/papers/literacy.pdf>> (Accessed September 21 2006).
- Queensland Studies Authority (QSA). (2001b). *Numeracy Position Paper*. <<http://www.qsa.qld.edu.au/yrs1to10/oia/papers/numeracy.pdf>> (Accessed June 30 2006).
- Queensland Studies Authority (QSA). (2002). *The Arts Years 1-10 syllabus*. <<http://www.qsa.qld.edu.au/yrs1to10/kla/arts/syllabus.html>> (Accessed July 20 2006).
- Queensland Studies Authority (QSA). (2000). *Studies of Society and Environment Years 1-10 syllabus*. <<http://www.qsa.qld.edu.au/yrs1to10/kla/sose/syllabus.html>> (Accessed July 20 2006).
- Queensland Studies Authority (QSA). (2004). *Mathematics syllabus Years 1-10*. <<http://www.qsa.qld.edu.au/yrs1to10/kla/mathematics/index.html>> (Accessed on July 20 2006).
- Queensland Studies Authority (QSA). (2005a). *2005 Test reporting handbook*. <<http://www.qsa.qld.edu.au/testing/357tests/archive.html#2005>> (Accessed 24 July 2006).
- Queensland Studies Authority (QSA). (2005b). *English Years 1 to 10 syllabus*. <<http://www.qsa.qld.edu.au/yrs1to10/kla/english/syllabus.html>> (Accessed 24 July 2006).
- Queensland Studies Authority (QSA). (2006a). *Your child and the 2006 Literacy and Numeracy tests*. <<http://www.qsa.qld.edu.au/testing/357tests/parents.html>> (Accessed 24 July 2006).
- Queensland Studies Authority (QSA). (2006b). *Overview. Queensland Years 3, 5 and 7 tests. The background and purposes of the Queensland tests*. <<http://www.qsa.qld.edu.au/testing/357tests/overview.html>> (Accessed 24 July 2006).
- Queensland Studies Authority (QSA). (2006c). *Presentation to Principals 2006*. <<http://www.qsa.qld.edu.au/testing/357tests/overview.html>> (Accessed 24 July 2006).
- Queensland Studies Authority (QSA). (2006d). *2006 Test preparation handbook*. <<http://www.qsa.qld.edu.au/testing/357tests/publications.html>> (Accessed 24 July 2006).
- Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(1), 1-25.
- Scarino, A. (2005). Teacher judgments: Going beyond criteria for judging performance. *Babel*, 39(3), 8-16.
- Smith, M. (2005). Getting SMART with data in schools: Lessons from NSW. Proceedings of the ACER Research Conference, 2005, *Using Data to Support Learning*, 38-45. <<http://www.acer.edu.au/workshops/documents/MaxSmith.pdf>> (Accessed August 23 2006).
- U.S. Department of Labor. (1991). *What work requires of schools: A SCANS report for America: 2000*. Washington, DC: U.S. Department of Labor. <<http://www.ncrel.org/engage/skills/match1.html>> (Accessed August 5 2006).
- Whetton, C. (2004). *More or less? Current assessment issues in England*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, April 2004.
- Wyatt-Smith, C. (2001). Exploring the relationship between large-scale testing programs and classroom-based assessment: A focus on teachers' accounts. In H. Fehring (Ed.), *Literacy Assessment: A collection of articles From the Australian Literacy Educators' Association* (pp. 144-163) The International Reading Association.
- Wyatt-Smith, C., Castleton, G., Freebody, P. & Cooksey, R. (2003a). Teachers' qualitative judgements: A matter of context and salience. Part I. *Australian Journal of Language and Literacy*, 26(2), 11-32.
- Wyatt-Smith, C., Castleton, G., Cooksey, R. & Freebody, P. (2003b). Teachers' qualitative judgements: A matter of context and salience. Part II. *Australian Journal of Language and Literacy* 26(2), 33-42.
- Wyatt-Smith, C.M. & Cumming, J.J. (2003). Curriculum literacies: Expanding domains of assessment. *Assessment in Education: Principles, Policy and Practice*, 10(1), 47-59.

Appendices

- ✧ Appendix One: Survey
- ✧ Appendix Two: Survey Responses
- ✧ Appendix Three: Data Collection Map
- ✧ Appendix Four: Years 3-6 Literacy and Numeracy Sets
- ✧ Appendix Five: Range of Standardised Tests in Use in Case Study Schools

Acknowledgements

The authors would like to acknowledge the contribution of Ms Kelly Freebody's research assistance in this project.

ⁱⁱ Perhaps the first finding of the project was the reluctance of schools to engage with research that may be demanding to complete if there was no apparent benefit for the school unless they participated further. In 2006 a further 200 questionnaires were sent to schools with the support of the QSA, but with a similar small response rate. Appendix 2 provides a summary of the questionnaire responses provided by all schools

ⁱⁱⁱ See <http://www.origo.com.au/index.php?page_id=1811> (Accessed 22 August 2006).

^{iv} Readers are advised to see <http://www.qsa.qld.edu.au/> for relevant documentation.

^v The two exceptions of students regarded by teachers as inappropriately identified as below benchmark and above benchmark have been discussed previously. These were the cases of the student with serious learning difficulties, if not disabilities, who gained a Numeracy rating beyond the high range scale and the high achieving student who was reported as below benchmark due to illness on the day. She did not complete two of the three literacy tasks.

^{vi} For example, Australia's inclusive education policies and high attendance rate until high school may affect the nature of the population included in testing, compared to other countries.

^{vii} Literacy and Numeracy: Year 5 test Intervention Grant.

<<http://education.qld.gov.au/finance/grants/fund/garp/html/yr5st.html>> (accessed 6 August, 2006)

^{viii} Literacy, Numeracy and Special Learning Needs Programme, Appendix C, Table 12.

<http://www.dest.gov.au/sectors/school_education/programmes_funding/programme_categories/key_priorities/literacy_numeracy_initiatives/default.htm> (accessed 6 August, 2006).

^{ix} Literacy, Numeracy and Special Learning Needs Programme, Appendix C.

<http://www.dest.gov.au/sectors/school_education/programmes_funding/programme_categories/key_priorities/literacy_numeracy_initiatives/default.htm> (accessed 6 August, 2006).