REVIEW OF STATEWIDE ASSESSMENTS IN THE CONTEXT OF NATIONAL DEVELOPMENTS

Interim Report

Professor George Cooney
School of Education
Macquarie University

Released June 2006
Executive Summary

Background
The impetus for the Review of Statewide Assessment Programs in the Context of National Developments was the inclusion in the Australian Government’s Schools Assistance Act Regulations 2005 of the requirement that “school authorities will implement before 1 January 2008 common testing standards, including common national tests in numeracy and literacy. Under the Regulations the relevant authority must ensure that the school administers a common national test before 1 January 2008 to each child who:

i) attends the school; and

ii) undertakes at the school a standardised assessment in reading, writing, spelling and numeracy at years 3, 5, 7 or 9”.

The overall aim of the review is to examine the implications of these regulations for the NSW statewide assessment programs; in particular to determine how best to shape the national assessment regimen so that it reflects the twin aims of achieving accountability while maintaining the richness of assessments for teaching and learning enhancement in a such a way that avoids unduly burdening teachers and schools.

This interim report describes the progress of the review. While this interim report contains a summary of the findings from my research to date, the focus of the report will be on the nature of the NSW and national testing programs. What makes a “good test”? What is necessary for a testing program to lead to improved teaching and learning strategies in our schools, and hence to higher achievement levels of our students in numeracy and literacy?

There is no disagreement with the proposition that the purpose of assessment, whether school-based or formal testing, is to provide information that will support teaching and learning, and lead to improved student outcomes. Curriculum reforms in New South Wales during the past decade, which have seen the introduction of
outcomes based syllabuses and standards based reporting, have stressed the importance of a close relationship between assessment and curriculum. Recent reforms in pedagogy including the *Quality Teaching Program* have highlighted the importance of consistency between assessment and teaching strategies, and the importance of timely and effective feedback to students.

Recent research indicates that the quality of statewide assessment programs can be judged under three headings: planning, collecting and using data. Planning involves clarity of purpose, resourcing and sustainability; collecting the data involves test development, administration and marking; and using the data involves communicating information to parents and schools.

My conclusion drawn from the review of existing research is that statewide assessment programs can be used to inform teaching and learning strategies in schools and thereby lead to improved student outcomes provided certain conditions are met. These conditions, in brief, are that:

- the assessment frameworks are based on a well-defined learning and achievement continuum so that the tests are related to what is taught in schools, what students learn, how they learn and the standards that they are expected to demonstrate,
- the tests items will allow the achievement levels of all students to be accurately determined on a common scale against standards, and
- parents and schools will receive timely and appropriate feedback that will identify strengths and weaknesses and which will support teaching and learning

All state and territory assessment agencies argue that their current statewide testing programs can be regarded as best practice as they possess the qualities specified by current researchers and practitioners in educational measurement. Despite differences in state and territory curriculum structures, in all cases it is the local curriculum which determines the content and range of skills assessed in their tests and the expected standards to be achieved, which is consistent with best practice in test design. All jurisdictions also assert, with some justification, that their tests are
comprised of high quality items that allow the achievement of all students to be accurately determined against standards; their teachers are actively involved in the development, administration and marking of their tests; and informative and timely feedback is provided to parents and teachers to support teaching and learning strategies in their schools. Examples have been provided of the way data from statewide testing programs have led to improved learning outcomes of students, particularly students in disadvantaged areas.

It can be argued that because the school curriculum in New South Wales is operationally defined by a set of syllabus documents which contain specific outcomes, the link between curriculum and tests is closer than in some other states. Consequently the specifications of the New South Wales tests reflect more closely what is taught in schools and the standards the students are expected to reach. Test items are linked to specific syllabus outcomes and vary in difficulty so that the achievement levels of almost all students can be accurately determined on a common scale. Because of the large range of ability and the need to make the test accessible to all students there is, however, some difficulty in having enough difficult items to accurately place high performing students. This is likely to be a major problem for tests carried out on a national scale.

NSW, Queensland, Victoria and Western Australia all provide information to schools on disk or through a data warehouse but the amount of information varies. New South Wales appears to be regarded as “best practice” because of the range of analyses that schools can, the close links with syllabus outcomes and the provision of links to specific teaching strategies developed by the DET Curriculum Support Directorate. Future embellishments are likely to include on-line access to data and support material.

Although the state assessment authorities acknowledge that the proposed national tests are likely to be of high quality, several general concerns have been expressed that the testing program may be deficient in some respects.

The first general concern is the purpose of the testing program and the nature of the
test itself. If the primary purpose of the testing program is accountability with the national tests serving as benchmark tests, the tests may not be able to accurately determine the achievement level of all students across the ability range. Currently the NSW tests contain items with a range of difficulty so that a student’s achievement can be determined accurately and reported against a set of standards.

The second general concern is the way the test specifications have been developed. Best practice dictates that the curriculum, as defined by a syllabus or a set of outcomes or a learning statement, determines the content and range of skills assessed in the tests. In the case of New South Wales individual items are tied to specific syllabus outcomes. In contrast, at the national level, general agreement has been reached on the test specifications so it can be argued that the tests themselves will define, in broad terms, a national curriculum. National statements of learning are being prepared but their impact on test development will be minimal in the first instance because learning statements and assessments are being developed in parallel.

In all cases, however, it is the local curriculum which determines the content and range of skills assessed in their tests and the expected standards to be achieved, which is consistent with best practice in test design.

A third general concern relates to teacher involvement in test development, administration and marking. This aspect cannot be underestimated. It has been argued that, to a large degree, teacher involvement helps give the existing statewide tests their credibility and is important for the professional development of teachers.

A fourth general concern relates to reporting and feedback to schools. There is no dissent from the view that schools must be provided with timely, informative and accessible data from the tests that will help schools improve the teaching and learning strategies. State assessment agencies will need sufficient data to allow them to report student achievement against standards and to identify areas of strengths and weakness in the curriculum. To achieve this objective, test items need to be linked to the outcomes of state curricula as they are in New South Wales and
the tests must be administered early in the school year. A late administration will not allow sufficient time for schools to be able to adapt teaching and learning strategies for their current cohorts of students.

The above summary does not imply that national testing programs cannot, per se, lead to improved teaching and learning in schools and hence to higher levels of student achievement. **What the research demonstrates is that for large-scale tests to lead to improved student outcomes a set of conditions must be met.** These conditions are discussed more fully in this report.

It is likely that the national tests will regarded by teachers and parents as important because of reporting requirements at individual and school level, in which case the tests could influence what is taught and how it is taught. Ideally large-scale tests would be developed from an existing curriculum but as this is not possible in the Australian context. It is therefore essential that conditions and strategies are in place to minimise any distortion of the teaching and learning programs in New South Wales schools as a consequence of the national tests.

The final chapter in this report addresses some of the issues surrounding the introduction of the national tests of numeracy and literacy, and some of the implications for New South Wales. Several provisional recommendations have emerged from the analysis.

Firstly, there is much in common between the NSW tests and the national assessment frameworks and, at this time, there is not a strong case in support of additional tests.

The national test specifications contain recommendations about the distribution of items and the number of linking items. It is essential that these recommendations are implemented in the trialing of items in 2006 and in subsequent test development.

Following the trialing of items during 2006 it is essential that levels of student
achievement be determined which would allow student achievement from Year 3 through to Year 9 to be reported on a common scale against standards. As a first step it is desirable to examine and describe the six levels used for the NSW Basic Skills, ELLA and SNAP tests and, using school-based data or School Certificate data together with the recently released NSW Foundation Statements of Learning, then develop trial achievement levels for Year 9.

The following should be regarded as non-negotiables during the trialing phase and subsequent test development and administration:

- teacher involvement in the development and panelling of items
- teacher involvement in marking
- state measurement officer involvement in analysis and quality assurance
- linkages to existing state and territory trend data
- timeliness for reporting to parents and schools
- further enhancements in the provision of feedback to schools

Developing the administrative structures necessary for a national testing program will take time and require delicate negotiations between the state and territory jurisdictions. It would therefore be unwise to proceed with the national tests until 2008.

In addition to the implications for all states arising from the general concerns discussed above, there are several issues that are specific to New South Wales. These include administrative structures for the processing of raw test data and the subsequent reporting to parents and schools, the future role for the School Certificate examination and the consistency of reporting student achievement K-12.

These, and other matters including the effect of transition on student performance and funding, will be discussed in the final report. My plan is to formulate, using data collected through the survey of schools, case studies and further conversations with relevant groups, a set of options that will then be discussed with relevant stakeholders before the final report is completed.
Acknowledgements

I acknowledge the great assistance given by senior staff in the NSW Department of Education and Training in the supply of relevant documents and other material.

I acknowledge also the willingness of members of the Reference Group who provided submissions from their organisations, arranged meetings with members from their associations and supported the survey of schools.

I acknowledge the way that staff from the state assessment agencies were willing to be interviewed about the impact of national tests on their own statewide assessment programs.

Finally, but not least, I am grateful to the school principals who generously participated in the school survey at such a busy time of the year, and to the teachers and parents who completed the questionnaires.
Abbreviations

ACER  Australian Council for Educational Research
BEMU  Benchmarking and Educational Measurement Unit
EMSAD Educational Measurement and School Accountability Directorate
DET   Department of Education and Training
DEST  Department of Education, Science and Training
KPM   Key Performance Measure
MCEETYA Ministerial Council for Education, Employment, Training and Youth Affairs
NMAG  National Measurement Assessment Group
## Contents

Executive Summary i
Acknowledgements viii
Abbreviations ix
Contents x

### Chapter 1 Overview of project

1.1 Introduction 1
1.2 Project aims

### Chapter 2 Methodology

2.1 Methodology 6
2.2 Stages of the review 6
2.3 Literature review 7
2.4 Interviews and submissions 7
2.5 Survey 8
2.6 Trend data 10

### Chapter 3 Large-scale assessment programs – a summary of the current research

3.1 Introduction 11
3.2 National testing programs in US, UK and Australia 12
3.3 Statewide and national programs in Australia 13
3.4 Current research on assessment practice 13
3.5 Are current statewide tests “good tests”? 16
3.6 Summary 17

### Chapter 4 National tests – are they “good tests?”

4.1 Introduction 19
4.2 Assessment frameworks and statements of learning 19
4.2.1 Implications for NSW 22
4.3 Reporting student achievement 27
4.4 Teacher Involvement 31
4.5 Timing 32
4.6 Reporting to Parents 32
4.7 Feedback to Schools 33
4.8 Summary 34
Chapter One  Overview of Project

1.1 Introduction

Since 1989 the New South Wales Department of Education and Training has developed a comprehensive program for assessing the performance of students in government schools: Basic Skills Test (BST) and Primary Writing Assessment Test (PWA) in years 3 and 5, English Language and Literacy Assessment (ELLA) and Secondary Numeracy Assessment Program (SNAP) in years 7 and 8. Two further tests have been proposed: Computing Skills (Year 6) and Science (Year 8). These tests are now also widely used in non-government schools, especially Catholic systemic schools.

These tests serve two purposes: the first is to improve the learning outcomes of students in numeracy and literacy and the second is to inform the Minister on the performance of students in public schools in these two areas. Parents are informed about how their children are performing against standards, schools are given useful diagnostic information about learning outcomes and changes in achievement levels can be monitored. Appropriate statistical methods are used to place the results from the three series of tests (years 3, 5 and 7) on a single scale, allowing the progress of students to be measured and the impact of schools to be assessed.

In March 1997 all State, Territory and Commonwealth Education Ministers agreed on a national goal which stated that “every child leaving primary school should be numerate and able to read, write and spell at an appropriate level” (National Report on Schooling in Australia, 2000, p2). A National Literacy and Numeracy Plan was implemented which included the development of national benchmarks for each of years 3, 5 and 7, the assessment of students against these benchmarks and the national reporting of benchmark data.

A nationally agreed procedure was developed to equate state and territory tests to ensure that the reporting of student achievement data against the literacy and
numeracy benchmarks was comparable. This procedure has been in place since 2001 and has been administered by the Performance Measurement and Reporting Taskforce’s (PMRT’s) Benchmarking and Educational Measurement Unit (BEMU).

This process has allowed states and territories to conduct assessments within their own jurisdictions to meet local curriculum requirements and school organisation as well as to meet Ministerial Council on Education, Employment and Youth Affairs (MCEETYA) and PMRT agreements regarding the assessing and reporting of literacy and numeracy standards. The current program of national assessments was endorsed by MCEETYA in 2002 when it approved the first version of the Measurement Framework, and supported in December 2003 by the PMRT. In December 2004 PMRT recommended that the practice of allowing states to retain their own annual full cohort tests of literacy and numeracy be continued. PMRT recommended that common items be administered to a random sample of 700 students each year to assist the benchmarking of state standards. A trial was proposed for early 2006.

The Australian Government’s Schools Assistance Act Regulations 2005, however, requires that school authorities will implement before 1 January 2008 common testing standards, including common national tests in numeracy and literacy. Under the Regulations each relevant authority must ensure that the school administers a common national test before 1 January 2008 to each child who:

i) attends the school; and

ii) undertakes at the school a standardised assessment in reading, writing, spelling and numeracy at years 3, 5, 7 or 9.

There are, in addition, regulations which specify how student and school performance should be reported.

These regulations, which overturn previous decisions of MCCEETYA as implemented by PMRT, obviously impact on the assessment regimens of the states and territories. The major issue is that of accommodating the requirements as specified in the Schools Assistance Act Regulations 2005 with the local needs of
the states and territories, and how the Regulations can be satisfied without losing the integrity of the local assessment and reporting programs.

The NSW Education Minister, Ms Carmel Tebbutt, asked for a review of statewide assessment programs in the light of developments around national assessment. The following section details the terms of reference and the aims of the review.

1.2 Project Aims

The overall aim is to examine the implications of the *Schools Assistance Act Regulations 2005* for the NSW assessment program, in particular to determine:

- which components operate most effectively
- whether reforms are needed
- how best to shape the national assessment regimen so that it reflects the twin aims of achieving accountability while maintaining the richness of assessments for teaching and learning enhancement in a such a way that avoids unduly burdening teachers and schools.

The terms of reference are:

1. to identify the essential criteria for effective student performance assessment in order to ensure quality evaluation of student performance, coherence with the curriculum, capacity to tailor teaching and learning strategies and opportunities for improving student outcomes

2. to examine the purpose, development, method, setting, timing, cost, reporting and feedback (provided or planned) for each NSW assessment in the growing national requirements governing timing, content, scope and reporting of assessments at years 3, 5, 7, 9 and 12.

3. to examine trends in performance data and reporting, including national consistency issues and NSW performance in national benchmarks over the period of statewide assessment practice.
4. to evaluate the effectiveness and efficiency of the current NSW assessment arrangements including value for money and coordination between the NSW Department of Education and Training and the NSW Board of Studies, and compare with existing and proposed national assessments, international approaches and those of other states and territories

5. to recommend improvements that will assist alignment with national requirements and, more generally, enhance high expectations for student educational outcomes

6. propose improvements aimed at producing better numeracy outcomes

7. examine the impact of the transition of students from primary to secondary schools on their numeracy performance, and identify the steps that schools can take to improve support for students over this period.

The project is structured around the following four issues:

1. the role of statewide/national tests in enhancing teaching and learning: how information is used by schools and systems for diagnostic and accountability purposes (R1)

2. the NSW assessment program and its relationship with the national framework: an analysis of the current NSW assessment processes, the assessment processes in other states and territories, what can be tested nationally and what must be maintained in NSW tests, and implications for reporting to students and schools (R2, R4)

3. trends in performance data and reporting: national consistency issues and NSW performance in national benchmarks over the period of statewide assessment practice (R3, R5)
4. numeracy outcomes in year 7 and the impact of transition (R6, R7)

The following chapter outlines the research methodology and timelines.
Chapter Two  

Methodology

A mix of methodologies has been employed: a review of previous and current research, document analysis, interviews, survey and case studies.

2.1 Methodology

The first step was a review of previous and current research on large-scale testing programs in Australia and overseas, both national and international. This focused on the roles of such testing programs, the tension between feedback and accountability, and how information from these programs is used by schools, educational authorities and governments.

The second step involved interviews with relevant officers in the NSW Department of Education and Training (DET), PMRT and other educational agencies in the states and territories. This had two aims: to access documents and information that may not be readily available in the public domain and to explore issues that arise.

The third step involved a survey of schools to determine the views of principals and teachers on the way information from statewide testing programs is used in schools. This survey will be followed by a case study of 12 primary schools and 12 high schools to some explore issues in depth. Three country centres will be selected, with two high schools and two primary schools chosen in each centre to give a mix of size and remoteness. The city schools will be chosen to obtain a mix of socio-economic/non-english backgrounds (NESB) combinations, and in the case of high schools, selective and comprehensive.

2.2 Stages of the review

The stages for the research are as follows:

Stage 1: Analysis and review of relevant literature and available documentation and some local
interviews (DET, ACER)

Stage 2: Preparation of survey material for schools
   Interstate interviews
   Mail out of surveys to schools

Stage 3: Progress report

Stage 4: Analysis of survey data

Stage 5: Case study interviews

Stage 6: Report completed

Stages 1, 2 and 3 have been completed as described in the following sections.

2.3 Literature Review

An extensive review of previous and current research on national and international large-scale testing programs in Australia and overseas has been completed and is being written up to form the early chapters of the report. While focusing on the purposes of large-scale testing programs and their potential for improving student learning outcomes, the review has also examined other current issues in curriculum and assessment to provide a broad background to the report.

In addition to professional and research journals, documents from the NSW DET, NSW Board of Studies, DEST and from other jurisdictions have been examined.

2.4 Interviews and Submissions

Meetings have been held with:

- Staff from the following jurisdictions:
  - New South Wales Department of Education and Training
  - Victorian Curriculum and Assessment Authority
  - Queensland Department of Education and the Queensland Studies Authority
  - Tasmanian Department of Education
  - West Australian Department of Education
  - New South Wales Board of Studies
• Associations related to members of the Project Reference Group:
  o Secondary Principals Council
  o Primary Principals Association Measurement Group
  o Catholic Education Commission Assessment Group

• Members of PMRT including Dr Gillian Shadwick and Dr Christine Ewan
• Dr Peter Titmanis BEMU, Perth
• Professor Jim Tognolini ACER

These meetings focused on national initiatives and their impact on schools and local assessment programs.

Written submissions have been received from:
• Secondary Principals Council
• Primary Principals Association Measurement Group
• Catholic Education Commission Assessment Group
• NSW Teachers Federation
• Association of Independent Schools
• Federation of Parents and Citizens Associations NSW
• NSW Board of Studies

Interviews with South Australian and Northern Territory educational authorities have not been carried out at this time but will be completed in February.

2.5 Survey

The survey of NSW schools commenced in mid November with questionnaires being distributed to samples of primary and secondary schools. A systematic random sampling procedure was used to select the target sample comprising:

• a 20% sample of primary schools
• a 50% sample of secondary schools
• a 50% sample of central, community and K-12 schools.

The (rounded) numbers of schools are given in the table below:
Support for the administration of the survey was given by the NSW DET, the Catholic Education Commission, the NSW Teachers Federation, the Federation of Parents and Citizens Association, Secondary Principals Council and the Primary Principals Association. As part of the development phase of the questionnaire, approximately 20 principals were interviewed to identify their perceptions of the issues surrounding statewide tests of literacy and numeracy. Ethics approval was granted by the Ethics Review Committee (Human Subjects) of Macquarie University.

In each school the principal was asked to complete a written questionnaire and to distribute questionnaires to:

- a random sample of eight teachers involved with the statewide numeracy and literacy tests
- ten parents (Primary: five from each of year 3 and year 5, Secondary: five from each of year 7 and year 8)

The questionnaires to principals, teachers and parents have a common section which asks for their perceptions of how informative they find the information from statewide numeracy and literacy tests in relation to individual students and how they use the information. For teachers and principals further sections focus on how the information can be and is used to improve student learning.

Data from the survey will inform the case studies. Distance Education Centres were
omitted from the survey. These centres will be contacted in the course of the review.

2.5 Trend Data

Trend data from the Basic Skills Tests, ELLA and SNAP have been provided by the DET Educational Measurement Directorate and are being analysed.
3.1 Introduction

Linn and Gronlund (2000), after citing from a 1992 report prepared for Congress by the Office of Technology Assessment, which states

“from the earliest days of the public school movement, American educators, parents, policy makers, and taxpayers have turned to [externally mandated] tests as multipurpose tools: yardsticks of individual progress in classrooms, agent of school reform, filter of educational opportunity, and barometer of the national condition” (p3),

argue that in the past 20 years at least governments in the US and elsewhere have used assessment as a lever for educational reform. Much of this reform has focused on improving student performance by making schools accountable, providing parents with choice, changing curriculum and modifying the way in which schools are resourced. Although the emphasis is on improving standards, the focus of large-scale assessment programs has been the school or the system.

The growth in advocacy for assessment for learning (formative assessment) rather than assessment of learning (summative assessment) can be seen as a reaction to this emphasis. Black and Wiliam, for example, argue that educational reforms based on large-scale assessment programs have had little or no effect on student performance and conclude that “assessment which is explicitly designed to promote learning, formative assessment, is the single most powerful tool we have for both raising standards and empowering lifelong learning” (ARG, 1999). Other researchers, such as Shepard (2000), argue that these large-scale programs are not merely ineffective but trivialise education and lead to lower levels of achievement in the higher order skills.

Despite the increase in criticism of large-scale testing programs, governments in many countries are continuing to rely on these tests to inform the educational debate and the formation of educational policy.
3.2 National testing programs in US, UK and Australia

Large-scale testing programs introduced in the US and the UK have an accountability focus but with different emphases. Parental choice and school accountability were two of the main incentives for the introduction of a national curriculum and testing program in England. The result has been the publishing of league tables of schools, with a school’s position determined by the performance of the students in that school on the national tests. School accountability was also an important factor in the No Child Left Behind initiative in the US but the program is also linked with the allocation of funding to individual schools. Additional Federal funding for schools is determined by the performance of schools on the statewide tests. In both countries these assessment programs have been praised and criticized; praised for improving student performance and criticised for narrowing the curriculum and leading to superficial learning.

National benchmark tests proposed for Australia had their genesis in March 1997 when all State, Territory and Commonwealth Education Ministers agreed on a national goal which stated that “every child leaving primary school should be numerate and able to read, write and spell at an appropriate level” (National Report on Schooling in Australia, 2000, p.2). In his Keynote address to the Curriculum Corporation 6th National Conference 6-9 May, 1999, the Hon Dr David Kemp, the then Minister for Education, Training and Youth Affairs outlined his vision for outcomes reporting and accountable schooling. He saw school reporting and accountability as important elements in improving student outcomes, making schools more effective and helping to make the government’s programs more effective. His argument was that nationally comparable reporting was “vital in improving the effectiveness of all Australian schools” (Kemp, 1999, p.2). He argued for reporting against benchmarks which were seen as the minimum standards that the community expects from Australian schools, and for the development of a national literacy and numeracy plan. This plan, which included the development of national benchmarks for each of years 3, 5 and 7, the assessment of students against these benchmarks
and the national reporting of benchmark data, was subsequently implemented.

Recent national initiatives now include national literacy and numeracy tests at each of years 3, 5, 7 and 9 and the development of learning statements and targets in all Key Learning Areas. An Australian Certificate of Education is currently the subject of a review undertaken by ACER and tenders have been called for reviews to examine national standards in English, Literature, Mathematics, Physics and Chemistry. All these initiatives have implications for state assessment programs which will be addressed in the final report.

3.3 Statewide and national testing programs in Australia

Over the past decade all states and territories have either developed their own numeracy and literacy tests or have used tests from other agencies, tests which all state jurisdictions argue are different from benchmark tests. Statewide tests report student achievement against a range of standards rather than against a single benchmark, they are aligned to local curricula and provide valuable information to parents and schools. Parents are informed of the standard their child has reached and schools are provided with information that helps them adjust their curriculum and pedagogical practices to better meet the needs of their students.

The New South Wales tests illustrate what the jurisdictions are arguing. The original Basic Skills Tests were designed to provide both performance and diagnostic data on individual students to parents, and still do. What has changed is the extent to which schools make better use of the information from the tests through better support material detailing the way the tests are marked and how data from the tests can be used in schools, and a computer program that allows schools to analyse results from their schools in different ways.

3.4 Current research on assessment practice

Current research shows clearly that a requirement of high quality assessment is that it is an integral part of the teaching and learning program (Gardner, 2005), and that
appropriate and timely feedback is provided to students. This is evident in recent work in Queensland with their “rich assessment tasks”, research by the Assessment Reform Group in the UK, by Hattie in New Zealand, by the Curriculum Corporation in Australia and the NSW Department of Education and Training with its Quality Teaching Program. Much of the current research has lead to an advocacy for assessment for learning rather than assessment of learning. Assessment for learning is assessment which promotes learning in students and thus leads to improved learning outcomes: it is “the process of seeking and interpreting evidence for use by learners and their teachers, to identify where the learners are in their learning, where they need to go and how best to get there” (ARG, 2002).

Many researchers (e.g. Shepard, 2000; Hattie, 2005; Black and Wiliam, 2005) see assessment for learning as consistent with current models of pedagogy based on constructivist principles and thereby enhancing student outcomes. In contrast, formal tests are seen as assessment of learning and perceived by many researchers (e.g. Shepard, 2000) as having undesirable curriculum and pedagogical consequences which do not lead to improved learning outcomes.

To assist teachers to access “the complex weave of classroom activities involving pedagogic style, student-teacher interaction, self-reflection, motivation and a variety of assessment processes” (Gardner, 2005, p 2) the Assessment Reform Group has enunciated ten principles of assessment: “assessment for learning is part of effective planning, focuses on how students learn, is central to classroom practice, is a key professional skill, is sensitive and constructive, fosters motivation, promotes understanding of goals and criteria, helps learners to improve, develops the capacity for self-assessment and recognises all educational achievement” (ARG, 2002).

The following four questions, contained in support material for the NSW Quality Teaching Program (DET, 2004, p.10) capture the essence of the ten principles in a succinct way.

What do you want the students to learn?
Why does the learning matter?
What are you going to get the students to do (or to produce)?
In an effective assessment regime outcomes regarded as important by the class teacher or educational authority are identified, the assessment strategy is consistent with pedagogy and appropriate to the outcomes being assessed, and student achievement judged and reported against standards. Timely feedback is implied, and the focus is on student improvement.

The belief that formal tests in general and large-scale testing programs in particular cannot be regarded as anything more than summative assessment can, however, be challenged. In a seminal paper, *The assessments we need*, Masters and Forster (2000) argue that “large-scale assessment programs have an import role to play in providing dependable information for educational decision making by policy makers, system managers, school leaders, teachers and parents” (p.1).

Masters and Forster readily acknowledge the way that large-scale testing programs can influence what schools and teachers do, and the undesirable consequences for curriculum and pedagogy that can result. Consequently the purpose of their paper is to identify design principles for a “good test”, one that will support student learning and lead to improvement in student achievement. These principles focus on the “kinds of learning addressed, the range of assessment methods used, and the ways in which student achievements are summarised and reported” (p.5). The authors argue for “designing assessment procedures primarily to establish where all students are in their learning, incorporating assessments of higher-order skills and thinking; including a variety of assessment methods and procedures to provide information about a range of valued learning outcomes; and reporting results in ways that encourage high achievement” (p5). The four questions (DET, 2004) presented above echo these principles.

It is clear that the “good test” envisaged by Masters and Forster is more than a benchmark test whose primary purpose is to establish whether a student has or has not met a minimum standard: they see the purpose of the test is to establish where “all students are in their learning”. To achieve this purpose an achievement
continuum must first be defined in terms of content and skills, the test items must cover the content and skills and have a range of difficulty. There must be sufficient easy items to allow the test to be accessible to all students and enough difficult items to challenge and provide reliable information about the more able students; there should be the same number of items at each level of difficulty.

In contrast, a benchmark test whose primary purpose is to determine whether a child has achieved a specified standard will have the majority of its items set at or near the specified standard. The result is that the test can accurately determine whether a child has achieved the benchmark but provides little information about the standard the child can achieve.

Issues surrounding large-scale testing programs have been critiqued at recent Australian conferences, the 2005 ACER Research Conference, the 2005 Curriculum Corporation Conference, and the 2004 and 2005 National Roundtables. The views expressed by speakers including Hattie, Kingsbury, Linn, Lowe and Holmes-Smith are consistent with those of Masters and Forster.

To sum up, evidence from recent research, from my discussions with people in the various jurisdictions and examples from schools, indicate that large-scale assessment programs can result in improved student outcomes if they share the qualities of good classroom assessment tasks. These qualities include a close relationship with what is taught and how it is taught, high quality items that allow the achievement of all students to be accurately determined against standards, and adequate and timely feedback to students and schools that supports their teaching and learning strategies. These qualities will be discussed in more detail in the following chapter.

3.5 Are current statewide numeracy and literacy tests “good tests”?

From documents and meetings with staff in state educational agencies it is clear that all jurisdictions share the same strong view that their numeracy and literacy tests satisfy the conditions for a “good test”. They assert that their tests are based on a
local curriculum or learning statements, teachers are involved in test development, administration and marking. Student achievement is reported against standards and, through linking items, student progress can be monitored on a common scale. Schools receive detailed and timely feedback by various means.

It can be argued that because the school curriculum in New South Wales is operationally defined by a set of syllabus documents which contain specific outcomes, the link between curriculum and tests is closer than in some other states. Consequently the specifications of the New South Wales tests reflect more closely what is taught in schools and the standards the students are expected to reach. Test items are linked to specific syllabus outcomes and vary in difficulty so that the achievement levels of almost students can be accurately determined on a common scale. Because of the large range of ability and the need to make the test accessible to all students there is, however, some difficulty in having enough difficult items to accurately place high performing students. This is likely to be a major problem for tests carried out on a national scale.

NSW, Queensland, Victoria and Western Australia provide information to schools on disk or through a data warehouse. The amount of information varies and NSW appears to be regarded as “best practice” because of the range of analyses that can be performed by schools, the close links with syllabus outcomes and the provision of links to specific teaching strategies developed by the DET Curriculum Support Directorate. Future embellishments are likely to include on-line access to data and support material.

3.6 Summary

This chapter has presented a short account of the current research on assessment which will be presented in full in the final report. My conclusion drawn from the review of existing research is that statewide assessment programs can be used to inform teaching and learning strategies in schools and thereby lead to improved student outcomes provided certain conditions are met. These conditions, in brief, are that:
• the assessment frameworks are based on a well-defined learning and
achievement continuum and the tests are related to what is taught in schools,
what students learn, how they learn and the standards that they expected to
demonstrate

• the tests items will allow the achievement levels of all students to be
accurately determined on a common scale against standards

• parents and schools will receive timely and appropriate feedback that will
identify strengths and weaknesses and which will support the improvement of
teaching and learning

These conditions will be discussed in the context of the national tests in the following
chapter.
Chapter Four  National tests – are they “good tests?”

4.1 Introduction

An argument has been made in the previous chapter that the purpose of all assessment, whether classroom-based assessment or large-scale testing programs is to improve student achievement. All jurisdictions argue that their statewide literacy and numeracy tests can be used to improve student outcomes because of the nature of the tests, the information provided to schools and the ways their schools are using the data. They have expressed some concern that the national testing program with its emphasis on reporting against benchmarks, reporting to the Federal Minister of Education only the percentages of students in each jurisdiction who have achieved the benchmarks, might not be as successful as their own tests in leading to improved student outcomes. Specific concerns they raised include the relationship between the tests and what is taught in their schools, reporting against standards, teacher involvement and feedback to schools.

This chapter will discuss some of these issues in light of the qualities deemed necessary for a “good test”.

4.2 Assessment frameworks and statements of learning

As mentioned in the previous chapter, existing state numeracy and literacy tests are based on local syllabus documents, outcomes or learning statements. In New South Wales, for example, each test item is linked to a specific syllabus outcome so that the tests are closely related to what is taught in schools. Some states such as Victoria and Tasmania do not have such a tight syllabus structure but test items are linked to their curriculum outcomes or learning statements that determine what is taught in schools. In all cases, however, it is the local curriculum which determines the content and range of skills assessed in their tests and the expected standards to be achieved, which is consistent with best practice in test design.
In the absence of a national curriculum, the national assessment domains for numeracy and literacy are based on topics and outcomes that are common across the state and territory current curriculum frameworks and syllabuses, and which can be reasonably assessed in national tests. Areas which are restricted to one or several jurisdictions were excluded, since all students across Australia do not have the opportunity to learn that particular content or skill. Consequently there are some aspects of curriculum which are currently delivered in some states and territories but which are not represented within the assessment domains for the national tests. A summary of the draft frameworks (PMRT, 2005c) is shown in Table 1.

Table 1: Summary of draft frameworks for the national assessment of Literacy and Numeracy

<table>
<thead>
<tr>
<th></th>
<th>Reading</th>
<th>Writing</th>
<th>Spelling</th>
<th>Numeracy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year 3</strong></td>
<td>30-35 items approx 45 minutes</td>
<td>9 criteria 1 task – narrative approx 45 minutes</td>
<td>Approx 25 items 25 minutes + spelling in writing</td>
<td>50-60 items 2 x 30 minutes sessions</td>
</tr>
<tr>
<td><strong>Year 5</strong></td>
<td>30-35 items approx 50 minutes</td>
<td>9 criteria 1 task – narrative approx 45 minutes</td>
<td>Approx 25 items 25 minutes + spelling in writing</td>
<td>50-60 items 2 x 30 minutes sessions</td>
</tr>
<tr>
<td><strong>Year 7</strong></td>
<td>30-35 items approx 55 minutes</td>
<td>9 criteria 2 tasks – narrative &amp; argument approx 2 x 45 minutes</td>
<td>Approx 25 items 25 minutes + spelling in writing</td>
<td>50-60 items 2 x 35 minutes sessions</td>
</tr>
<tr>
<td><strong>Year 9</strong></td>
<td>30-35 items approx 60 minutes</td>
<td>9 criteria 2 tasks – narrative &amp; argument approx 2 x 45 minutes</td>
<td>Approx 25 items 25 minutes + spelling in writing</td>
<td>50-60 items 2 x 40 minutes sessions</td>
</tr>
</tbody>
</table>

Within each domain is an organising framework (Table 2). Literacy will be reported as Reading, Writing and Spelling but numeracy will be reported as a single scale.
Table 2: Organising frameworks

<table>
<thead>
<tr>
<th>Substrands</th>
<th>Reading</th>
<th>Writing</th>
<th>Spelling</th>
<th>Numeracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contextual understandings</td>
<td></td>
<td>Narrative competence</td>
<td></td>
<td>Algebra, function &amp; pattern</td>
</tr>
<tr>
<td>Text location skills</td>
<td></td>
<td>Thematic competence</td>
<td></td>
<td>Measurement, chance and data</td>
</tr>
<tr>
<td>Interpretation skills</td>
<td></td>
<td>Structural components</td>
<td></td>
<td>Number</td>
</tr>
<tr>
<td>Text and language knowledge</td>
<td></td>
<td>Cohesive control</td>
<td></td>
<td>Space</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Narrative effect</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sentence structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vocabulary</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Paragraphing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Punctuation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The national literacy and numeracy benchmarks articulate nationally agreed minimum standards at Years 3, 5 and 7 and benchmarks for Year 9 are currently being developed. These benchmarks are intended to be diagnostic and identify students at risk, as well as to assess system performance. They do not attempt to describe the whole of literacy and numeracy, nor the full range of what students are taught.

The Australian Government is also committed to the development of statements of learning in several curriculum domains at years 3, 5, 7 and 9 to build more national consistency in curriculum outcomes. States and territories are required to implement these statements of learning either as part of their next curriculum review, if that occurs between 2006 and 2008, or before 1 January 2008. The first statement of learning, English, was approved in February 2005, and other statements of learning are under development.

MCEETYA’s decision was that statements of learning were not intended to be the basis for national standards and national tests. The focus of the statements of learning is on curriculum, what students should be given the opportunity to learn,
whereas the focus of the national tests is on standards, in particular the minimum standards expected to be achieved by all students at defined stages of their schooling. In relation to English, for example, the statements of learning describe “the essential knowledge, understanding, skills and capacities that students should have the opportunity to learn, ie those elements of the intended curriculum in English where national consistency in curriculum has been agreed. The literacy benchmarks, on the other hand, focus on literacy, which could be said to be a subset of English and an extension beyond English. They describe important and essential elements of student achievement ie those elements of attainment nationally agreed to be those without which a student would have difficulty in making progress at school” (AESOC, 2005, p5).

A consequence of this decision was that statements of learning and the assessment domains for benchmarks were developed in different contexts (AESOC, 2005). Despite their differences, the English statements of learning have been influenced by the literacy benchmarks and the assessment domains developed by Masters and Forster (1997). It can therefore be anticipated that the statements of learning for Mathematics, which are being developed, will also be similarly influenced by the numeracy benchmarks. It thus follows that the assessment domains for the national numeracy and literacy tests will have played a major part in determining the intended national curriculum in both English and Mathematics. The relationship between statements of learning and assessment frameworks is seen, however, as evolving so that in future years the statements will influence revisions of the assessment frameworks. This issue will be explored in depth in the final report.

4.2.1 Implications for New South Wales

There is much in common between the NSW literacy and numeracy tests and the draft assessment frameworks shown above, but there are substantial differences. In the national tests it is proposed that language is assessed through the writing task and not reported separately as in New South Wales. On the other hand, spelling will be assessed and reported separately in the national tests. For writing New South
Wales tests require two text types, narrative and argument for all years but the national tests require the Year 3 and Year 5 students to write a narrative. The national numeracy framework includes an additional strand, algebra, function and pattern, to New South Wales and the measurement strand has been expanded to include chance and data. In contrast to New South Wales, numeracy will be reported as a single strand.

In literacy it has been argued that language is a separate skill which deserves to be reported separately but this is not a universal view. Likewise with spelling – should it be assessed and reported separately or be seen as part of writing? The argument for having only one text type, narrative, for Years 3 and 5 is based mainly on testing time and cost. The counter-argument is that primary teachers may pay too much attention to narrative because it is in the test, to the detriment of other text types. This issue will be examined further below.

Justification for reporting numeracy as a single strand can be found in earlier research which showed that student marks on measurement, space and number in the NSW Basic Skills Tests were very highly correlated and the strands were unidimensional. In contrast, although the marks on comprehension and language were highly correlated, the two strands were not unidimensional.

The national numeracy framework for Years 3, 5 and 7 has a uniform spread of items across the four strands of measurement, chance and data, space, and number. Currently in New South Wales there is a greater emphasis on number in Year 3 and measurement in Year 7. Space, in particular, has less emphasis in Year 7 (18%) than in the national framework (33-34%). Given the percentage of items in the existing Year 3 tests devoted to number (ACT: 26%, NSW:46%, NT:40%, QLD:36%, SA:28%, Vic: 53%, WA/TAS:50%), the recommendation in the national framework is perhaps surprising. The justification (PMRT, 2005b) was as follows:

“The emphases on number in many of the state and territory Year 3 instruments, if intentional, may reflect instructional emphases. In classrooms there is often a concentrated effort to ensure that students in the earlier Year
levels have sound ‘number sense’ before they are expected to apply that knowledge to measuring and working with data in classroom situations they may encounter in their daily lives.

While this may be the case, each state and territory mathematics curriculum gives equal weight and importance to each of the strands. Teachers are encouraged to devote equal instructional time to all aspects of the mathematics curriculum. To reflect the intentions of the curricula, equal emphasis should be given to each of the strands in the assessment.” (p17)

The reduced emphasis on number in the early tests will, however, be balanced by the inclusion of items on chance and data in these tests.

Given the high intercorrelations between the separate numeracy strands it can be argued that little will be lost in reporting numeracy on a single scale. A counter-argument is that the strands have discrete content and that separate scores and, at least at school and system level, scores on the separate strand can highlight gaps in teaching. This was evident following the introduction of the NSW Basic Skills Tests in 1988 when space was reported as a separate strand. This was an area that had been given less emphasis than the other areas in primary schools and the mean for Space was lower than that for the other strands. The following year there was a distinct improvement in this strand.

I am not arguing that the national framework should be changed for reporting student achievement to individual parents as reporting separate strands would make the student report unnecessarily complicated. Schools, however, could benefit from knowing how their students performed on the separate numeracy strands (algebra, number, measurement, space, chance and data). This would allow them to identify areas where student performance was less than expected and perhaps adjust teaching programs accordingly. The “space” example cited above gives support for this assertion.
The loss of the language strand is regretted by some educationalists in New South Wales as is the reduction of the text types from two to one. Currently NSW is the only state to require two tasks, the others require only one task although the text types vary as shown in the following table.

<table>
<thead>
<tr>
<th>State/Territory</th>
<th>Text types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian Capital Territory</td>
<td>Argument</td>
</tr>
<tr>
<td>New South Wales</td>
<td>Narrative and argument</td>
</tr>
<tr>
<td>Northern Territory</td>
<td>Argument</td>
</tr>
<tr>
<td>Queensland</td>
<td>Argument/recount/description</td>
</tr>
<tr>
<td>Victoria</td>
<td>Narrative/description</td>
</tr>
<tr>
<td>Western Australia/Tasmania</td>
<td>Narrative</td>
</tr>
<tr>
<td>South Australia</td>
<td>narrative</td>
</tr>
</tbody>
</table>

The justification for one writing task is in terms of cost and demands on young students (PMRT, 2005c).

> “Because of the costs associated with marking and the demands on younger students of completing all the components of the literacy assessment, students in Years 3 and 5 [will] be assessed on one piece of extended writing. The recommended genre is the literary narrative, which is considered engaging for students in this age range and had been proved in testing programs to provide an effective means of showcasing their writing.” (p.21)

This argument has merit given that the total testing time for Year 3 is almost three hours. On the other hand, it can be argued that knowledge of which test type will be examined may influence the teaching program in the primary years with undue emphasis on narrative to the detriment of other genres. Again, the “space” example cited above fives some credence to this hypothesis.

At a national level it is recognised (PMRT, 2005a) that individual states and territories may decide to conduct assessments that are in addition to the elements
specified in the frameworks for the national assessment of literacy and numeracy. PMRT proposes, that if this is done that consideration be given to the printing, scanning and marking requirements to “ensure that the process of matching students’ results to the common national instruments can be achieved accurately and efficiently.” (p.1)

If this recommendation is accepted, New South Wales would be permitted to have a separate test for the language strand of literacy but whether this is necessary is unclear. A separate test would have merit if and only if the assessment of grammar and vocabulary as part of the writing task was deemed unsatisfactory, if teachers regarded the separate reporting of language as essential, and the benefits of a separate test outweighed the cost and demand on young students.

Underlying the desire for additional tests to assess strands not covered by the national frameworks is the assumption that external assessments are always more reliable, more valid and add value over and above classroom tests. Classroom assessment can be reliable and valid, and assess a range of outcomes which cannot be assessed through statewide tests. It is not necessary to assess everything through statewide or national tests, and this is recognised in the national frameworks.

My recommendation, at this stage, is that New South Wales not argue for additional test items in addition to the elements specified by the national frameworks.

What will be explored in the coming months are the options available for New South Wales. Already schools receive excellent support material on how to use data from the statewide literacy and numeracy tests but there are opportunities for further developments. These include the development of on-line school-based testing using the bank of items that have been used in earlier Basic Skills, ELLA and SNAP tests; an option that is already available in Victoria. The items are already calibrated so that student achievement could be reported against standards. Another possibility is building on the Quality Teaching Program material and advising schools on how to assess topics in all Key Learning Areas, focusing on the higher-order skills. These
possibilities will be discussed further in the final report.

4.3 Reporting student achievement

Currently states and territories report student achievement in their statewide tests across a range of standards with no comparability of standards across jurisdictions. It is clear from conversations with assessment staff in the various state jurisdictions that they wish to continue to report against their established standards in order to preserve continuity with existing trend data. Their primary concern is to receive data so that their existing reporting practices to continue. This is an issue that will be examined in depth in the final report but at this stage I am not convinced that it is necessary to maintain the exiting time series. Given the changes in curriculum, in the tests and in the reporting requirements there is the opportunity to re-calibrate our reporting standards so that they better meet the needs of parents and educators.

Earlier policy documents (Kemp, 1999) indicated that reporting of achievement in numeracy and literacy was to be against benchmarks. While the intention was that reporting should be at individual and school level, for some time only results aggregated at state and territory level have been released, together with a limited amount of disaggregated data. Issues surrounding reporting at individual and school levels have become more important following the passing of the Schools Assistance Act Regulations 2005.

Recent documentation (PMRT, 2005a) and discussions with staff at BEMU indicate that, while the primary purpose of the national tests is to allow accurate reporting against benchmarks, the assessment specifications in the draft frameworks have been broadened in response to the reporting requirements of state and territory governments and the enhancements sought by MCEETYA. These enhancements include the “capacity to report on the performance of the most able to the least able, including benchmark achievement, and the creation of a single vertical scale with the capacity to map the growth in students’ achievements as they progress through school” (p.5).
To achieve this aim the national tests must contain items of varying difficulty with sufficient easy items that can be accessed by almost all students, a range of moderate items for the majority of students, and sufficient difficult items for the more able students. In theory, for accurate placement of all students, the item difficulties should be uniformly distributed across the full ability range.

The distribution of item difficulties and the number of link items are discussed in the assessment frameworks for numeracy (PMRTb, 2005) and literacy (PMRTc, 2005). The documents contain recommendations that there should be a range of item difficulties with items distributed uniformly across the quartiles of expected student achievement, and that there be at least 12 link items in adjacent tests to allow vertical equating of these tests. These recommendations are consistent with what is seen as best practice in test design but they will not be easy to implement.

In New South Wales and in most other states there is already some concern that, because of time constraints, their current tests do not contain sufficient difficult items to determine accurately the standards achieved by their high ability students. Despite having items of varying difficulty there tends to be a ceiling effect because of a lack of very difficult items. Extension modules are under consideration by the NSW Educational Measurement Directorate to address this issue.

One solution, used by the Board of Studies in reporting HSC achievement against standards and by DET in reporting achievement in the Basic Skills, ELLA and SNAP tests, is to leave the highest described standard as open-ended. There is then need only to define accurately the lower boundary of this top level; there is no upper boundary. However, except for students who gain full marks on the test, students in this top category can still be ordered if necessary. If this procedure is accepted, the problem becomes one of determining the number of levels or standards to be used for reporting, and then describing these standards. These are not easy tasks.

In relation to New South Wales, the standards used to report student performance in the Basic Skills Tests were set in 1989 and it can be argued that it is time to re-calibrate these standards. Perhaps the appropriate time would have been when the
results for the Year 3 and Year 5 Basic Skills tests are reported on a common scale. Introduction of national tests now provides a suitable opportunity.

The desire to determine accurately where all students are in their learning and to chart the progress of students through their time at school is far more difficult than describing the standard reached on a particular test. Complexity is increased in two aspects. We must first describe how student learning develops, what is termed the achievement continuum, and then to have assessment tasks at different points along this continuum. An example of what is required is contained in the 1996 National School English Literacy Survey (Masters and Forster, 1997), which describes in detail the construction of an achievement continuum for English literacy and the corresponding assessment tasks. This achievement continuum is essential if we wish to report against developmental standards rather than standards related to a specific task. Describing such an achievement continuum is difficult, few exist and fewer have been based on research evidence.

The second step is to equate the various assessment tasks so that we can report student performance on this achievement continuum. Equating in this context, “vertical equating”, is normally achieved by the use of common or link items in adjacent tests. In the New South Wales Basic Skills tests, for example, there are link items to equate Year 3 and Year 5 tests, and Year 5 and Year 7 tests so that it is possible to report student achievement in numeracy and literacy in years 3, 5 and 7 on the same scale. Currently, a common scale is used for reporting Year 3 and 5 results, but a different scale used for Year 7.

It has long been recognised that the choice of link items is critical in vertical equating, and for some time educational measurement researchers in the US considered that vertical equating was not possible. They argued that it is difficult to choose link items that provide a range of achievement across adjacent levels and that the results from equating are not robust, being too dependent on the actual link items chosen. This is evident in New South Wales where a two year gap between tests makes the choice of suitable link items very difficult. There is a case for some research on the robustness of vertical equating in New South Wales.
The above discussion highlights some of the difficulties associated with reporting against standards for state cohorts at specific years, and for reporting in such a way that parents and teachers can chart the progress of students through their time at school. To achieve the first there must be an adequate spread of item difficulties in each test; to achieve the second developmental standards must be adequately defined and link items carefully chosen.

Currently the assessment frameworks contain definitions of numeracy and literacy, describe test specifications and benchmark standards for Years 3, 5 and 7 but there are no descriptions of the expected range of student achievement or achievement continuum. No benchmark standards have been described for Year 9: the closest to a description of the Year 9 benchmark is a comment in a PMRT (2005d) document that the Year 9 benchmark could be set equal to the “proficiency standard” for year 7.

These shortcomings are a result of the assessment frameworks being developed in parallel with the statements of learning rather than after. Statements of learning can provide some insight into the achievement continuum and the expected range of student achievement, at least in the first instance. It would be desirable to carry out a survey of Year 9 achievement in literacy and numeracy, using the statements of learning as frameworks and a methodology similar to the 1996 National English Literacy Survey. Given the timetable for the implementation of the tests, such a survey is unlikely.

The methodology to be employed for the trialing of test items in 2006 is to take items from existing state and territory tests for Years 3, 5 and 7 and to develop additional items for the Year 9 tests, estimating the required levels of item difficulty. In the construction of the Year 3, 5 and 7 tests it is essential that the recommendation about the range of item difficulties be implemented. In the absence of a Year 9 survey it is necessary for a further Year 9 trial to be conducted in 2007 to establish the range of item difficulties required to allow the accurate positioning of Year 9 students.
The discussion thus far has focused on item difficulty. A second and no less important issue is the question of linking items. As stated above it is well established that choice of the linking items is all important in defining a common reporting scale so research needs to be carried out to determine the effect of different choices of linking items on the overlap between Years 3, 5, 7 and 9. Research on the New South Wales tests could inform the process.

To summarise, what is essential in the trialing of items in 2006 is that:

- the items are distributed uniformly across the quartiles
- there are adequate link items between adjacent tests

Additionally, further research is required to develop an achievement continuum for each of the numeracy and literacy strands to describe the expected levels of achievement from Year 3 through to Year 9. This could form the basis for reporting against standards in New South Wales and other states.

### 4.4 Teacher Involvement

Strong views have been expressed about teacher involvement in the testing program. In all states some teachers were involved in the development of test items and in the marking process. Staff in several jurisdictions have stated that teacher involvement gives the schools a sense of ownership of the tests, which increases their value. They expressed their concern that if teacher involvement decreased there was the likelihood that schools would see the tests as less relevant to their teaching and learning program.

The importance of having teachers involved in the marking process cannot be underestimated, and reflects the perception about marking the School and Higher School Certificate examinations. Marking is seen as an important professional development activity which allows teachers to really understand what the standards mean. This is especially true for teachers in rural areas; the country marking centres are highly valued. When online marking was raised with members of the Catholic
Education Commission’s Measurement Group, they reiterated their view that the interaction of markers in corporate marking centres was the key to good professional development. They argued that their teachers should be rotated through the centres to maximise the benefit to their schools.

4.5 Timing

The timing of the tests has been raised as a concern if the tests were to be seen as more than benchmark tests with an accountability focus. A statement made by staff in one jurisdiction was that “if the tests are just benchmark tests, it does not matter when they are given. Schools will receive little information that can be used to change pedagogy”. It can be argued that if the tests are just to provide information on the level of achievement across Australian schools, they should be given towards the end of the year.

As current statewide tests are regarded as more than benchmark tests most jurisdictions argue that administration of the tests should be early in the year so that schools can receive their reports in time to make changes and rectify deficiencies found in the performance of their students. Most states support May as a suitable time.

4.6 Reporting to Parents

Currently, student achievement is reported against standards with no consistency across states. Student reports are distributed by the state agency responsible for development and administration of the tests. In New South Wales the Educational Measurement Directorate of DET has this responsibility for all schools, government and non-government, who participate in the testing program.

Examination of the documentation and a discussion with staff from BEMU indicate that the national agency responsible for test development and analysis will provide the Federal Minister of Education with information limited to the percentage of students in each state who achieve the national benchmarks with some amount of disaggregation of the data. Raw data comprising individual item scores for each
student and the item difficulties will be provided to the relevant state jurisdictions who could then report to parents and schools against their own standards. This procedure raises a question of ownership of the data and which state agency will be responsible for reporting to parents and to schools. When this issue was explored with state jurisdictions their answers varied. In some states the Department of Education was the relevant body, in others it was the state Curriculum and Assessment agency.

New South Wales has a different structure to most states, with the K – 12 curriculum and the School Certificate and Higher School Certificate examinations being the responsibility of the NSW Board of Studies. The remaining statewide tests are the responsibility of the NSW Department of Education and Training. This differentiation of responsibility is not an issue at present as those non-government schools that participate in the statewide testing program agree to the DET processing their results and reporting to their parents and schools. An issue arises when all schools are required to participate in the national tests.

In New South Wales the Minister of Education and Training has the statutory responsibility for all schools, government and non-government, in the state. The question to be addressed, and which will be addressed in the final report, is how the raw data might be processed, the reporting framework and the mechanism by which parents and schools receive student results.

4.7 Feedback to Schools

A further issue that was raised by state jurisdictions and in submissions received was the issue of feedback. Currently schools in each state receive extensive feedback on individual student and school performance, with information given at the item level. Given that the test items, especially in New South Wales, are tightly linked to syllabus or curriculum outcomes schools can identify areas of the curriculum in which their students are performing well or badly. Curriculum support material is generally available for schools to use to improve the teaching and learning in the areas where student performance is lower that what would be expected.
Implementation of national numeracy and literacy tests will mean that, although there is much in common between the local test specifications and the assessment frameworks for the national tests, the tests NSW students will undertake in year 3, 5, 7 and 9 will not be as closely related to what they are taught as at present. It will, however, still be possible to link test items to local syllabus outcomes for the aspects of numeracy and literacy covered in the national tests but it will be conceptually a different process. Rather than test design and item selection commencing from local syllabus documents, the starting point will be the national frameworks so the linking of test items to local outcomes will occur after rather than before item selection. This is not necessarily best practice in test design but may allow schools to receive the type of feedback they receive at present.

The importance of providing to schools appropriate feedback and support that are related to curriculum and pedagogy cannot be underestimated. In all states the amount of feedback and the ease with which schools can access the information have improved, so that schools see more value in the tests for their teaching and learning programs.

4.8 Summary

This chapter has canvassed some of the issues surrounding the introduction of the national tests of numeracy and literacy, and some of the implications for New South Wales. A number of provisional recommendations have emerged from the analysis.

Firstly, there is much in common between the national assessment frameworks and, at this stage, there is not a strong case for additional tests to cover the language strand in literacy. There is the opportunity to develop on-line tests for schools from existing items, and to provide advice to schools on how to develop school-based assessment tasks for assessing other aspects of Literacy and other Key Learning Areas.

The test specifications contain recommendations about the distribution of items and
the number of linking items. It is essential that these recommendations are implemented in the trialing of items in 2006 and subsequent test development.

Following the trialing of items during 2006 it is essential that levels of student achievement be determined which will allow student achievement from Year 3 through to Year 9 to be reported on a common scale against standards. As a first step it is desirable to examine and describe the six levels used for the NSW Basic Skills, ELLA and SNAP tests and, using school-based data or School Certificate data together with the recently released NSW Foundation Statements of Learning, then develop trial achievement levels for Year 9.

From the information to hand it is clear that the following items should be regarded as non-negotiable during the trialing phase and subsequent test development and administration:

- teacher involvement in the development and panelling of items
- teacher involvement in marking
- state measurement officer involvement in analysis and quality assurance
- linkages to existing state and territory trend data
- timeliness for reporting to parents and schools
- further enhancements in the provision of feedback to schools

Developing the administrative structures necessary for a national testing program will take time and require delicate negotiations between the state and territory jurisdictions. It would therefore be unwise to proceed with the national tests until 2008.

These issues will be researched further and a more detailed analysis presented in the final report.
References


DET (2004a) Quality teaching in NSW public schools: An assessment practice guide

DET (2000b) Quality teaching in NSW public schools: Continuing the discussion about assessment practice


Masters G. and Forster, M. (2000) *The assessments we need*. ACER, Hawthorne,

PMRT (2005a) Assessment frameworks for the common national instruments in literacy and numeracy. Item 6.3. 20\textsuperscript{th} October 2005.


PMRT (2005d) Unresolved elements of the frameworks for the national assessment of literacy and numeracy. Item 5.1.2. 28 September, 2005.

*Proceedings of 9\textsuperscript{th} Round Table Conference*. Sydney, 2004.

*Proceedings of 10\textsuperscript{th} Round Table Conference*. Melbourne, 2005.