

# A Report for the National Education Performance Monitoring Taskforce

---

Options for the assessment and reporting of primary students in the key learning area of science to be used for the reporting of nationally comparable outcomes of schooling within the context of the *National Goals for Schooling in the Twenty-First Century*

## Presented by

Professor Samuel Ball  
*Principal Consultant*  
Professorial Fellow, University of Melbourne

Professor Ian Rae  
*Project Manager*  
Professorial Fellow, University of Melbourne

Professor Jim Tognolini  
*Consultant Science Testing*  
University of New South Wales

May 2000

## CONTENTS

---

<b>ACKNOWLEDGMENTS.....</b>	<b>3</b>
<b>PART I: BACKGROUND AND INTRODUCTION.....</b>	<b>10</b>
1. <i>Introduction</i> .....	10
2. <i>Methodology</i> .....	11
3. <i>Monitoring</i> .....	11
4. <i>Primary science</i> .....	13
5. <i>Science literacy</i> .....	14
<b>PART 2: INTERNATIONAL PERSPECTIVES – CURRICULUM AND ASSESSMENT .....</b>	<b>17</b>
1. <i>Introduction</i> .....	17
2. <i>New Zealand</i> .....	18
3. <i>Ontario</i> .....	21
4. <i>United States of America</i> .....	22
5. <i>England</i> .....	27
6. <i>International studies</i> .....	29
7. <i>Summary and conclusions</i> .....	32
<b>PART 3: AUSTRALIAN CONTEXT – CURRICULUM .....</b>	<b>35</b>
1. <i>Introduction</i> .....	35
2. <i>The role of teachers</i> .....	39
3. <i>Summary and conclusions</i> .....	40
<b>PART 4: AUSTRALIAN CONTEXT – ASSESSMENT AND REPORTING .....</b>	<b>41</b>
1. <i>Introduction</i> .....	41
2. <i>Assessment</i> .....	41
3. <i>Reporting</i> .....	46
3. <i>Summary and conclusions</i> .....	48
<b>PART 5: OPTIONS – COSTS AND BENEFITS.....</b>	<b>49</b>
1. <i>Introduction</i> .....	49
2. <i>The options</i> .....	49
<b>POSTSCRIPT .....</b>	<b>62</b>
1. <i>Commonalties in the curriculum</i> .....	62
2. <i>Developing standards</i> .....	63

<b>APPENDIX A: USES OF MONITORING .....</b>	<b>67</b>
<b>APPENDIX B: THE INTERVIEW SCHEDULE.....</b>	<b>69</b>
<b>APPENDIX C: THOSE INTERVIEWED .....</b>	<b>70</b>
<b>APPENDIX D: INTERNATIONAL DETAILS .....</b>	<b>72</b>
<i>D1: Assessment resource bank sample – New Zealand .....</i>	<i>72</i>
<i>D2: Extended activity assessment options – Ontario .....</i>	<i>72</i>
<i>D3: Science investigation assessment – Michigan .....</i>	<i>72</i>
<i>D4: Operational items and marking schemes – England .....</i>	<i>72</i>
<i>D5: Multiple-choice and free-response items – TIMMS.....</i>	<i>72</i>
<b>APPENDIX E: THE FISCAL COSTINGS.....</b>	<b>73</b>
<b>APPENDIX F: SAMPLE ITEMS.....</b>	<b>74</b>

## **ACKNOWLEDGMENTS**

The authors of this report would like to thank the NEPMT Secretariat and especially Ms Vivienne Roche and Mr Robert Souter. Their attention to detail, helpful ways and forgiving nature were greatly appreciated.

We would also like to thank all of those who gave generously of their time and advice by allowing us to interview them. Their names are provided in Appendix C to this report and we salute their enthusiasm for and knowledge of primary science.

Sam Ball

Ian Rae

Jim Tognolini

May 2000

## Executive summary

The National Education Performance Monitoring Taskforce (NEPMT) was established in April 1999 by the Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA). The taskforce was called upon *inter alia* to develop performance measures as the basis for national reporting. One such area was science.

The measures were to be able to be disaggregated for reporting purposes to allow attention to be paid to specified groupings of educationally disadvantaged students. The basis for specifying groupings included socioeconomic status (SES), geographic location and language and/or ethnic background.

Science in the primary school was seen as a crucial curriculum area. Ministers has previously decided to participate in the OECD Programme for International Assessment (PISA) which included ‘science literacy’ as one of the three ‘literacies’ it was assessing. MCEETYA subsequently endorsed the use the PISA data to monitor the science achievement of 15-year-old students but that would still leave the achievement of Australian primary students unmonitored.

This report advocates adoption of the PISA ‘science literacy’ definition for purposes of primary science monitoring in Australia. Operationally, this means that students would be assessed in relation to concepts chosen from major fields of science and a range of process skills (e.g. identifying evidence, developing hypotheses).

From a consideration of monitoring by overseas education authorities and international assessment practices, it was clear that monitoring needs to be implemented against some kind of standards or curriculum framework and that the assessment should include objective, recognition-type items (e.g. multiple choice), open-response, production-type items and practical assessments.

A consideration of the Australian context was based upon a perusal of states/territories’ curriculum and assessment documents and interviews with more than thirty experts across all jurisdictions. The conclusions reached on this basis fortified by the study of international practices can be divided into three groupings:

### Conclusions concerning curriculum

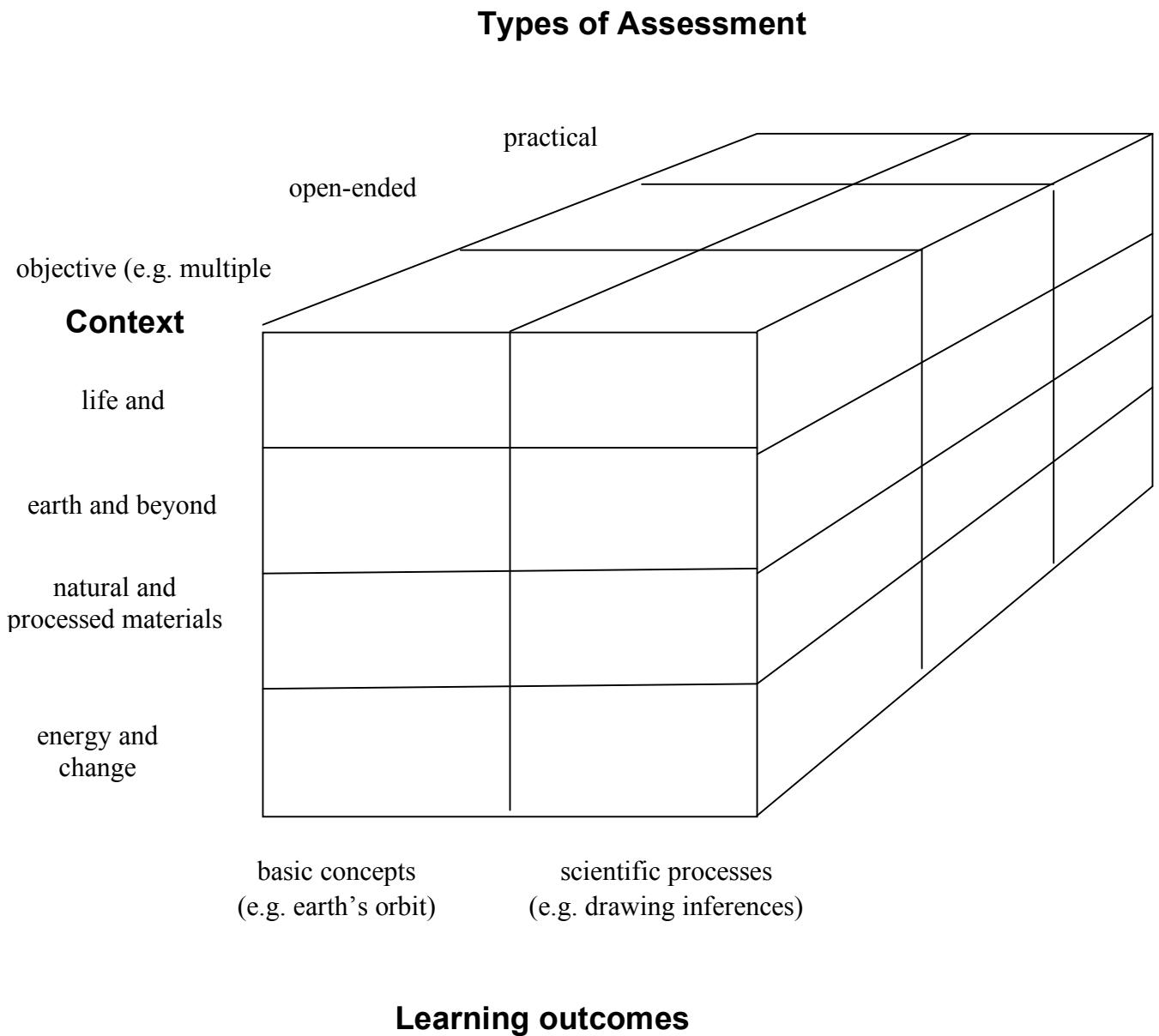
- Knowledge of science ‘facts’ are not important educational goals.
- Understanding of scientific concepts is essential.
- Skill in the use and understanding of scientific processes embedded in a context of scientific concepts is also seen as essential.

- There is an underlying commonality across Australian states/territories in terms of primary science concepts. This implicit commonality should be made explicit through a mapping exercise of state/territory curriculum frameworks.
- Differences in curriculums exist but they can be accommodated in a monitoring process developed for purposes of nationally comparable reporting.

### **Conclusions concerning assessment**

- Teachers should be involved in administering the tests, especially the practical assessments. This will mean about 1½ days of release time for training and for test and administration.
- The tests should include objective, open-ended and practical tasks.
- The assessments should take place when students are toward the end of Year 6.
- Assessment for monitoring purposes should occur at first, every two years, but a later move toward every three or so years should be contemplated.
- There is no need for census testing for the purposes of national monitoring. A scientific sample of up to 400 schools would be sufficient. Over-sampling might be given consideration if needed to ensure low sampling error in small population states/territories or in categories of disadvantage where numbers are small.
- Non-sampled schools should have the opportunity later to administer the assessments should they so wish.
- A three-dimensional specification model (Figure 1) is recommended as a framework for assessment and later reporting.

**Figure 1.** Three-dimensional projection of the assessment model indicating primary science context, learning outcomes and types of assessment



## Conclusions concerning reporting

- Both normative and criterion referenced reporting should occur.
- Normative referencing should allow states/territories to compare their student performance across a profile of data (see the marginals from Figure 1) with the data generated from all sampled Australian students.
- Criterion referencing should provide states/territories with information on percentages of students reaching specified bands of performance –
  - Band 1 (not yet competent with respect to some element in the profile)
  - Band 2 (within a reasonable range or about to achieve the element)
  - Band 3 (achieved the desired element)
  - Band 4 (moved beyond the desired element)
- Reporting will occur through the Australian National Report on Schooling (ANR). However, sampled schools should receive, in confidence, their results. Similarly, parents should receive, in confidence, through the sampled schools, information on their child's performance.
- Electronic reporting to each sampled school with some accompanying diagnostic analysis should be considered.

Five basic options for monitoring were presented (see Part 5 of the report). Each was analysed in terms of costs and benefits. The five were:

1. No national assessment and reporting.
2. Census testing of all Year 6 students.
3. Each state/territory develops its own assessment.
4. An item bank be developed and monitoring occur through schools or states choosing from the item bank.
5. An assessment involving objective recognition type items and open-ended items would be developed and administered to all targeted students in a scientific sample in each state/territory supplemented by other practical assessments for all those students. The practical assessments would be mediated and marked by the classroom teacher.

Option 5 is the preferred option because of the following benefits:



1. This option can provide comprehensive information about student performance in primary school science based upon objective and open-ended items, practical tasks and teacher judgments. It inherently ensures both reliability and validity in the data provided.
2. The fiscal cost of Option 5 is presented in Appendix E and it can be seen that this option is not markedly more expensive than Option 4 and is, overall, less expensive than Option 3.
3. The concept of sample testing and the use of teacher judgments is acceptable, in principle, to major teacher industrial groups so that Option 5 is unlikely to create industrial relations opposition if carefully implemented.
4. Sample testing as proposed in Option 5 allows monitoring information about student performance to be obtained with only the most minimal problems of sampling error in terms of state/territory and sector student performance. (Of course, this approach does not allow individual student performance results to be given to all parents. However, the parent is not necessarily the prime focus for the results of this exercise.)
5. This option ensures that the assessment will tend to mirror the curriculum and the pedagogy associated with primary school science. It builds on the experience of such an assessment approach in Western Australia. It provides a form of teacher professional development.
6. Option 5 carries with it the possibility not only of sampling schools (and sampling target students within schools – Option 5.2) but also of using the technique of item sampling. Thus, not all students need take all items as long as the various test forms have degrees of mutual item overlap.
7. Option 5 could also, at the discretion of Ministers, monitor every second or third calendar year. A yearly assessment is not a *sine qua non*. Indeed, two- or three-yearly assessments may provide sufficient monitoring especially after the dynamics of monitoring, research and program change are experienced. Thus, the recurrent fiscal cost could be lowered.
8. Option 5 would allow sampled schools not only to cooperate in the assessment process but also to obtain on a confidential basis information about their students' performance. The opportunity to have third-party information about strengths and weaknesses in their curriculum programs could be very beneficial.
9. Option 5 was overwhelmingly endorsed by those interviewed in this report.

The cost of Option 5 is calculated to be a little less than \$800,000 per assessment and reporting cycle.

As a final recommendation, we urge that the introduction of an assessment and reporting regime in primary science for monitoring purposes should occur in the context of a wider, positive package of reforms that would include:

1. Professional development programs relating to the monitoring framework of primary science.
2. The development for Australian teachers of an item or assessment task bank in primary science to be freely available for school use.
3. The development and use of a strategic communication plan which will clearly explain the purposes of national monitoring, the kinds of information to be published and the benefits that can accrue from the monitoring process.

## PART I: BACKGROUND AND INTRODUCTION

### 1. Introduction

The National Education Performance Monitoring Taskforce (NEPMT) was established in April 1999 by the Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA). The taskforce was called upon *inter alia* to develop performance measures as the basis for national reporting. One such area was science.

The measures were to be able to be disaggregated for reporting purposes to allow attention to be paid to specified groupings of educationally disadvantaged students. The basis for specifying groupings included socioeconomic status (SES), geographic location and language and/or ethnic background.

The taskforce was also charged with the function of identifying areas 'where it may be appropriate to establish national targets or benchmarks, in relation to the agreed key performance measures which assist state and school level planning and reporting for improvement'. (RFT Primary Science January, 2000)

Science in the primary school was seen as a crucial curriculum area. Science had been clearly targeted by the *1999 National Goals of Schooling in the Twenty-First Century*. The Federation of Australian Scientific and Technological Societies had been urging the state and federal governments to promote the study of science. According to the Australian Science, Technology and Engineering Council (ASTEC) report on Science and Technology in Primary Schools (May 1997), 'Australia's economic growth, employment opportunities for tomorrow's workers and the ability of citizens to make informal decisions about ... The foods they consume, the medical procedures they undergo, the machines, facilities and services they use and their impact on the environment ...' will increasingly require sound knowledge of science. In short, science is seen as an important key learning area much akin to the importance attached to literacy and numeracy which had already undergone the process of benchmarking, assessing and reporting.

In the view of the authors of this report, there are at least three reasons why primary school science should be a particular focus of national monitoring. First, it is often too late to arouse interest in science in secondary schools if it has been neglected in primary schools when children are at their most educationally curious. Second, results from some state testing programs have indicated problems in the teaching/learning of science in the primary school. Third, Australia has agreed to participate in the OECD Programme for International Student Assessment (PISA) which included 'science literacy' as one of the three 'literacies' it was assessing and reporting on. Indeed, Ministers have explicitly decided to use PISA data to monitor secondary students' science achievement. However, PISA was dealing with 15-year-old students. This left the primary school science area virtually unattended in terms of Australia-wide assessment and reporting. In any case, if Australia's education were to be scrutinised

by the PISA project assessing the performance in science of 15-year-olds, it would seem prudent to look at the basic primary school years to see if Australian students were getting a firm foundation in science.

In January 2000 the NEPMT invited tenders for an options paper. The authors of this report were informed in mid-February that they were the successful tenderers. They brought to the project expertise in curriculum, assessment and reporting (Ball), in science and science education (Rae), and in psychometrics, assessment and reporting (Tognolini).

## **2. Methodology**

The methodology underlying this report includes:

1. The collection through print and electronic technologies of curriculum and assessment materials in primary science from Australian educational and assessment agencies. This collection was analysed to see what the current situation is in terms of the intended primary science curriculum and subsequent assessment in Australian states and territories.
2. An analysis, including content analysis of overseas science curriculum and assessment in specified countries, to provide background to the focus on Australian curriculum and assessment.
3. Conducting semi-structured interviews of all Australian states and territories' relevant curriculum and assessment officers and selected office holders of professional associations concerned with primary school science. The topic areas and questions covered in the interviews are presented in Appendix A. Those interviewed are listed in Appendix B.
4. Conducting a fiscal cost analysis of those various options for assessment and reporting of primary science arrived at through the methodologies referred to above.

In the subsequent sections of Part 1 of this report, the major terminology and concepts will be examined to ensure a common basis for understanding the later presentation (Parts 2–6).

The terms and concepts to be examined are: 'monitoring', 'primary science' and 'science literacy'.

## **3. Monitoring**

This report is directly named in the title of the National Education Performance Monitoring Taskforce. The term 'monitoring' needs to be clearly understood because there are related synonymous terms that are not identical in meaning but they open up a potential confusion of purpose and process.

Monitoring may use some of the same processes as ‘evaluation’<sup>1</sup> and ‘applied research’<sup>2</sup> but the purposes vary. This options paper targets the processes of assessment and reporting for the purpose of monitoring. (Some further thoughts on monitoring are presented in Appendix A.)

These issues concerning monitoring will be taken up in more depth as we move from the Part 1 introduction of the report into the later substantive parts of the report. However, two matters call for some preliminary emphasis:

1. Monitoring implies some notion of standards that can have a descriptive or a prescriptive meaning. In some countries, the monitoring is mainly against a descriptive backdrop so that the monitoring simply describes the levels of achievement of the students. Increasingly, there is a move toward a more prescriptive regime in which the standards are established to indicate where students ought to be and the monitoring indicates levels of success against those standards. It does seem essential that any monitoring implies some notion of standards and areas of concern because on what other foundation could a monitoring assessment system be based? At least an agreed framework needs to be put into place if useable monitoring is to occur.
2. Monitoring implies a readiness to react if the assessment reports so indicate. In the metaphorical use of monitoring (be it in air conditioning systems, intensive care wards or internal combustion motors) there is a practical set of decision rules established to ensure that the monitoring is not just an ‘academic’ exercise. It may mean that if the monitored temperature rises above a given value the cooling system is activated or if the pulse rate falls below a given value an alarm calling for urgent medical intervention is sounded.

Of course in educational monitoring it is not necessarily wise to take literally what is, after all, a metaphorical analysis. Nonetheless, the implicit message for this report is that assessment and reporting of primary science are worthwhile only if there is an expectation of action if and where the reports suggest problems exist. Education is a profession beset with cynicism sometimes purposely engineered and sometimes induced by experience. A major message from those interviewed in the course of this report was either (in a few instances) that assessment in primary science should not be endorsed because it would lead nowhere, or (in most cases) that some form of assessment could prove fruitful if authorities noted the reports and acted upon them. This potential action should become a major part of the overall package and strategic communication plan established if Ministers decide to proceed with the monitoring of primary science.

---

<sup>1</sup> Educational evaluation has many purposes but it is usually based on the presence of a specific program which is being developed (formative evaluation) or is being assessed for some kind of accountability (summative evaluation).

<sup>2</sup> Applied research has an element of being curiosity-driven and its primary purpose is the creation of knowledge in an applied context. Applied research may lead to program initiation and development.

#### 4. Primary science

Twenty-five years ago, Helgeson, Blosser and Howe (1977) conducted an intensive study of science teaching in the USA. They concluded that *when* science is taught in the primary (elementary) school it is characterised by:

1. Frameworks generated by school districts detailing what should be taught at what Grade level.
2. Teachers who feel incompetent in their own knowledge of science.
3. Classrooms that teachers believe lack the essential resources for teaching science.
4. Schedules that do not permit enough time for ‘meaningful’ science instruction.
5. Teaching strategies often based upon a single textbook.

There are two noteworthy differences that may make this analysis of primary science less than valid for contemporary Australian policy makers. One is that the study occurred in the USA and the other that the data are now outdated.

The interviews carried out for this report included discussions with senior science educators and academics across Australia (see Appendices B and C). From those discussions it would seem that in Australia, as we balance on the cusp of the two centuries, there remain similarities with but also differences from the USA of 25 years ago.

From generalisations garnered from the interviews conducted across Australia, and from such reports as *Foundations for Australia’s Future Science and Technology in Primary Schools* (ASTECC 1977) it is noted that primary science in Australia is marked by:

1. Frameworks generated by states/territories that are often prescriptive in terms of skill outcomes and knowledge strands but much less so in terms of factual information.
2. Many primary school teachers who feel incompetent in their knowledge of science.
3. Classrooms that lack resources for certain kinds of science teaching (but this is not a major concern because the current thought is that primary science can be taught without the use of the resources expected and available in secondary science).

4. Schedules that make science teaching a marginal activity compared to literacy and numeracy and usually involving about an hour a week.
5. Teaching strategies that fortunately are now far removed from the dependence on a single textbook.

It is fair to argue that primary science in Australia has progressed in the past few decades but it is also well accepted that it still has a long way to go. Many primary teachers are still 'scared' of science and it operationally is marginalised by lack of time commitment in many primary classrooms.

It is well recognised that, especially in the primary school, teachers frequently and properly blur the demarcation lines of the academic disciplines. Thus, fortunately for the young students, topics that might otherwise be seen as belonging to science might appear in health, geography or technology lessons. Nonetheless, from both our interviewees and the reports on science education, there is a consensus that, in general, there is less than desirable commitment to science in the primary school.

Because there has been movement in primary science in Australia and because there is seen to be much change still needed it is, in principle, an important area for some kind of monitoring. As shall be pointed out in Part 3 there is much change occurring in primary science curriculum across Australian states/territories including Western Australia, South Australia, Victoria and, probably, New South Wales. Again, such change calls for a degree of monitoring of student performance.

It is the view of the writers of this report that there is a need to consider, not merely the assessment and reporting options for a monitoring program in primary science, but rather a larger contextual package. This primary science package (including presumably one of the monitoring options presented in this report in Part 5) would indicate that governments are dealing holistically with the key learning area. It would alleviate many of the concerns about assessment and reporting that are held by some educators as it would indicate clearly that assessment and reporting is only one aspect of the educational process. This matter will be taken up in detail in Part 6 of this report.

## **5. Science literacy**

This is a topic central to this report because explicit questions were raised in the request for proposals (RFP) about whether the PISA science literacy framework is an appropriate one if NEPMT decides to recommend to Ministers that primary science monitoring be undertaken. This is a topic dealt with in some detail in Parts 2 and 3 of this report.

As background, it is noted that there are many definitions of science literacy. Both Branscomb (1987) and Thomas and Durant (1987) provided eight different uses of the term. Jenkins (1994) concluded that 'science literacy is a problematic notion'.

The term ‘science literacy’ is given comprehensive treatment in the publication *Measuring Student Knowledge and Skills* (OECD 1999) and is the basis for the PISA project. As operationally defined in this publication, science literacy is assessed in relation to:

1. ‘the *content* or *structure* of knowledge that students need to acquire ...’
2. ‘a range of *processes* that need to be performed which require *various* cognitive skills’ and
3. ‘the *situation* or *context* in which knowledge and skills are applied ...’

The specific definition (*ibid.* p12) tells us that scientific literacy is: ‘Combining scientific knowledge with the drawing of evidence – based conclusions and developing hypotheses in order to understand and help make decisions about the natural world and the changes made to it through human activity.’

The components or dimensions are reiterated as:

1. *Scientific concepts*, e.g. energy conservation adoption ... chosen from the major fields of physics, biology, chemistry, etc.
2. *Process skills*, e.g. identifying evidence, drawing, evaluating and communicating conclusions. These ... cannot be applied in the absence of scientific content.
3. *Using science* in different situations.

We would emphasise the report’s strong statement that ‘no scientific process can be content free’ and that ‘the OECD/PISA science questions will always require an understanding of science concepts.’ (*ibid.* p14).

In general, the case is reasonable that in primary science, schools are not so much preparing students for a vocation as scientists but rather, by giving primary students some understanding of the more important scientific concepts and processes, ensuring they learn how science works and learning about functionally important aspects of science in our lives.

Attempts to measure aspects of scientific literacy in the adult population have indicated that at least, in the USA and the UK, scientific *illiteracy* abounds. In the USA fewer than 20 per cent of adults understood how a telephone worked and in the UK a large majority of adults described their level of understanding of the gene as ‘having little sense’ or ‘no idea’ (Lucas 1987).



There are unsubstantiated claims that in similar studies in New Zealand and in Australia somewhere between 25 per cent and 50 per cent of the adult population believe the sun circles the earth every 24 hours.

It would not be hard to argue that just as every citizen of Australia should have competence in the skills of literacy and numeracy and the concepts and skills of citizenship and computer literacy, so too, in a modern world they should be scientifically literate. Just what this means in terms of the primary school curriculum overseas and in Australia will be taken up in Parts 2 and 3 of the report. However, the writers of this report are in agreement with the crux of the PISA definition of science literacy and see it as the basic framework upon which to build the following Parts of this report.

## **PART 2: INTERNATIONAL PERSPECTIVES – CURRICULUM AND ASSESSMENT**

### **1. Introduction**

In recent years a number of overseas jurisdictions have adopted frameworks for science teaching and learning over the full range of school years. These frameworks have clear implications for assessment and reporting, though not all of them are explicitly linked to such practices. This approach is familiar to Australians because of the publication of the National Statements and Curriculum Profiles in eight areas of learning (Curriculum Corporation 1994) and subsequent adoption by most states and territories of detailed frameworks.

The first focus of development of science frameworks in a number of countries was specification of science curriculum content, presented as a series of stages or levels appropriate to stages of intellectual development as represented by years of schooling. As well as these structural considerations, questions were raised about the connections between science and technology, and the balance to be struck between science content and science processes. Jurisdictions have addressed these questions in their various ways. These questions will be familiar to anyone who has been involved in science curriculum development during the last decade.

Philosophical questions about the nature of science literacy which were discussed in the previous section are closely linked to the practical considerations surrounding any testing regime. Test developers need to decide whether school science learning should merely aim to equip students with repertoires of factual and procedural knowledge, or should attempt to go further in developing true science literacy. The answer to this question underlies one of the aims of the OECD Program for International Student Assessment (PISA), which has been endorsed by the Australian National Advisory Committee for the PISA project to monitor knowledge and skills of students in reading, mathematics and science. The question derives from the feeling that factual learning and regurgitation in response to test items is insufficient outcome of schooling, even if students are able to use the facts in defined situations and thus demonstrate some skill in the methods of science. Application of the knowledge and skills outside the school or vocational ‘science’ situation is a better test of science literacy, but this is difficult to assess in the primary years and often only brought into achievement testing of students completing their secondary schooling. Given the PISA definition of science literacy as ‘the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help to make decisions about the natural world and the changes made to it through human activity’, it could only be a very minor component of what was expected of primary school students or it would need to be interpreted in terms of elementary, simple applications. (See also, Part 3 of this report.)

Although the major emphasis of international testing has rested with students in secondary school, there has been a flow-on effect into learning across the school

curriculum and in all years of schooling. In many countries, including Australia, the response to international performance has been partly based on perceived need to strengthen the mathematics and science capabilities of students going through the education systems, with a view to enhancing national creativity or productivity, and partly based on national pride. Whatever the impetus, the ensuing impact of international testing has coincided with the independent development of frameworks which admit the impossibility of teaching exactly the same material in all schools/states/ countries, but rely instead on consensus over the elements of good curriculum. There are many examples of national responses and discussion forums, a typical one being the Australian Council for Educational Research Conference 1997 (Australian Council for Educational Research 1999).

In this survey of international practice we have selected jurisdictions not merely on the grounds of availability of data at comparatively short notice, but also to provide information from a small number of countries in which the education system is broadly similar to those of Australian states and territories. No fiscal cost data are available. Direct cost comparisons are probably of little use but we have included pertinent discussion in several of the succeeding sections and some speculations in the concluding section of this part of the report.

A note about references is in order. Printed versions are available for most of the material described in this report, but it is often not held in Australian libraries although copies may be in the hands of educational researchers such as members of staff of ACER. Most material is also available on the World Wide Web, although there are exceptions where payment is required and where websites are under construction or reconstruction, or the material is long out of date. Consequently, the diligent searcher after information needs to sample both print and electronic media to get maximum coverage.

## **2. New Zealand**

### **2.1 Science curriculum**

The New Zealand curriculum is presented as a three-dimensional framework, with year progression along one dimension, contextual strands along a second, and integrating strands along the third. The contextual strands use the typical 1990s nomenclature rather than traditional discipline-based names. Thus:

- Biology becomes ‘Making Sense of the Living World’;
- Physics becomes ‘Making Sense of the Physical World’;
- Chemistry becomes ‘Making Sense of the Material World’; and
- Earth Sciences and Astronomy, becomes ‘Making Sense of Planet Earth and Beyond’.

For each of the contextual strands at each Level, the New Zealand Science Curriculum provides:

- Achievement Objectives (investigate..., explore ..., describe ...);

- a dozen or so Sample Learning Contexts; and
- extensive lists of Possible Learning Experiences and Assessment Examples for teachers and students.

The two integrating strands which equate to the ‘process’ strands discussed above, are Making Sense of the Nature of Science and its Relationship to Technology, and Developing Scientific Skills and Attitudes. The first of these contains, at each Year level, sets of Achievement Objectives (understand ..., describe ..., investigate ...), and lists of Possible Learning Experiences and Assessment Examples. The second, however, features lists of suggested activities at each Year level for each of its sub-strands: Focussing and Planning, Information Gathering, Processing and Interpreting, and Reporting.

## 2.2 Testing

The New Zealand government produced a Green Paper, *Assessment for Success in Primary Schools*, in May 1998 dealing with national testing in primary schools, and there were expectations that legislation to give effect to its recommendations would be introduced in 1999 following release of a White Paper. The proposals in the Green Paper, which included mandatory testing at three Year levels, were subject to extensive consultation. There was opposition to some of its proposals by teachers and by the School Trustees Association, and so the government announced in August of that year that the timeframe for introduction of national tests had been extended to 2000. Following a report from an Assessment Working Group, the government opted to renew the contract for development of the Assessment Resource Banks (see below) for two further years, until June 2001, during which time there would be a trial of national testing. An election followed soon after, however, in which the government was defeated and new administration did not proceed with the trial. The assessment domain has not been rejected but it is under review by the new government.

The 1998 Green Paper recognised three groups of stakeholders for whom test information would be useful:

- the education authority, who sought a measure of accountability;
- schools, who would be able to use results to target teaching improvements, although in their objections the above groups stated that sufficient information for that purpose was already being obtained in schools; and
- parents, who are increasingly interested in obtaining such information from a variety of sources.

Reference was made in the Green Paper to the model provided by the erstwhile LAP testing in Victoria.

A fresh Green Paper on national testing is expected to be released by mid-2000, but seems to have been pre-empted by the statement *Information for Better Learning* released in February 2000 (available at [www.minedu.govt.nz/schools/assessment](http://www.minedu.govt.nz/schools/assessment)). The stated aim of national assessment is the enhancement of learning and the provision of information on which pupils, teachers, principals, boards, parents and the Government make decisions.

Although New Zealand designates seven essential areas of curriculum

- Language and Languages
- Science
- Mathematics
- Technology
- Social Studies
- the Arts
- Health and Physical Well-being

the Government proposal is to trial diagnostic tests in literacy and numeracy for Year 5 (approximate age 9 years) and Year 7 (11 years) students in a 10 per cent sample of schools. This represents an enhancement of New Zealand's National Educational Monitoring Project, introduced in 1993 which is based on in-depth studies of a 3 per cent representative sample of students.

A number of schools participate on a voluntary basis in a testing program developed by the New Zealand Council for Educational Research. Users must register with NZCER, but may then draw on items from Assessment Resource Banks (ARBs) which are accessible through the Council's website ([www.nzcer.org.nz](http://www.nzcer.org.nz)). This project began in 1993 with a feasibility study, and entered a second phase late in 1994 when the ARBs were implemented on a trial basis in 27 schools. Successive improvements to the search engine and accumulation of more mathematics and science resources culminated in the availability of the ARB on the NZCER website. These Resource Banks contain items which are ranked for difficulty and cover Levels 2–5 (students up to 14 years of age) of the New Zealand national curriculum, which was introduced in 1993. Included are some 200 English items, 800 Mathematics, and 800 Science, including some 'practical performance' questions that go beyond pencil-and-paper responses. At present these items are trialed in a small number of volunteer schools. Complete test papers are not provided but teachers may select from the Assessment Resource Banks to compile their own tests. The development of the ARBs has been marked by a staged approach over six years, with extensive feedback from participating schools.

Some of the difficulties experienced during the development of these assessment resources involved matters of comprehension or familiarity with words, although these were not always the expected ones. For example, in posing the question 'what does the term "predator" mean?' it was expected that some students would not be familiar with the word 'predator'. In practice 'predator' was not a problem, but many primary school students were unfamiliar with use of 'term' in this sense! Any language-mediated testing regime is likely to encounter such examples despite the

best efforts of test developers.

Two examples of items from the Science ARB are included in Appendix D.1 and we have chosen ‘practical’ items because these represent the class of greatest interest to test developers who are exploring ways to go beyond multiple-choice and open-answer tests. In each case, the test item is preceded by a copy of the database list that relates the item to the formal curriculum framework. The item on human body parts comes from level 2 of the living world strand, and that on fingerprints from level four.

### 3. Ontario

Canada has no national science framework (Council of Ministers of Education, Canada 1977) but each province proceeds in its own way although international instruments such as TIMSS have been used to provide a national perspective. We have thus selected one province for which information was available to us and present here some information of relevance to the present report. The province of Ontario, following an accelerated program of curriculum development, released a new science curriculum for schools in September 1998, following poor results by students of the province in international tests. Compared to the Australian and New Zealand frameworks, the Ontario course is much more prescriptive, but it is also arranged in strands and levels and is outcome-based and so could lend itself to province-wide testing. In Ontario, science and technology are included in the one field of study.

The *Ontario Curriculum, Grades 1–8: Science and Technology, 1998* arranges material in five strands – Life Systems, Matter and Materials, Energy and Control, Structures and Mechanisms, and Earth and Space Systems. There is no separate ‘process’ strand, but ‘habits of mind’ – incorporating the various science-based skills we have discussed earlier in this report – are incorporated into specific expectations (outcome statements) which form part of the curriculum.

An interesting feature of the Ontario science and technology curriculum is that various topics covered by the strands are introduced in specific years rather than being spread across years. Thus:

- for Energy and Control, Grade 2 is focused on energy from wind and moving water;
- Grade 3 on Forces and Movement;
- Grade 4 on Light and Sound Energy;
- Grade 5 on Conservation of Energy;
- Grade 6 on Electricity;
- Grade 7 on Heat; and
- Grade 8 on Optics.

Testing already exists in Ontario for English and mathematics at Grades 3 and 6, and there has been speculation that this could extend to Grade 9 and also include science, although it is acknowledged that this would be expensive. Ontario is one of the most advanced provinces in science curriculum, and so it is notable that no specific assessment or monitoring program in this field has been implemented.

A group at York University, near Toronto, which also prepared the draft provincial curriculum, has been working to develop test items and currently has available some 500 on a CD-ROM, arranged for Levels 1 – 4 (Orpwood *et al.* 1999).

Two examples of test items for the Energy and Control strand developed by the York team are included in Appendix D2, and once again we have chosen items which are activity based. The material in Appendix D2 clearly shows the context, for Levels 4 and 6 respectively, and includes notes for the teacher supervising the activity. They are clearly more suitable for school use than for wide-scale testing, but do indicate directions that might be taken in the development of open response and practical items for such applications.

#### **4. United States of America**

##### **4.1 National standards and testing**

National standards and testing have been strongly emphasised in the United States over the last 20 years, although as Wynne Harlen (formerly Director of the Scottish Council for Research in Education) has observed, the drivers have been as often political as educational (Harlen 1991).

Following a national conference in 1989, the National Educational Goals Panel was established in 1990 and it adopted the conference's six Goals for Education, to be achieved by 2000. The Goals included the raising of the high school graduation rate to 90 per cent (Goal 2), freedom of schools from drugs (included in Goal 6), world primacy in science and mathematics achievement (Goal 4), and (in Goal 3) that 'American students will leave grades four, eight, and twelve having demonstrated competency in challenging subject matter, including English, mathematics, science, history and geography'. A report on the first testing program found that students were poorly prepared in science and expressed concern about 'their apparent disinclination to enrol in challenging science courses (Jones *et al.* 1992). The report includes detailed analysis of test results and also a selection of items employed in the testing. Apparently there was little improvement during the period preceding the next round of testing, since the 1993 report of the Panel concluded that 'the current rate of progress is wholly inadequate' (*The National Education Goals Report: Building a Nation of Learners* 1993). This concern, although based on national test results, was perhaps heightened by comparisons which showed the United States students falling behind those of Hungary, Korea and Taiwan in 1991 international science assessments, although ahead of students in France and Switzerland.

Relevant statistics are available from the National Assessment of Educational Progress (initially biennial but later administered less frequently) program which has operated since 1969 with samples of 9-, 13- and 17-year-old students in public and private schools, and was broadened in 1990–1992 to 4th, 8th and 12th grade students. A brief history of NAEP and its predecessors is given by Húsen and Postlethwaite (1994). Science testing was last conducted in 1996 and is scheduled again for 2000.

There is complementary work by the National Assessment Governing Board on reporting achievement levels in the tests under the headings *basic* (partial mastery), *proficient* (solid academic performance), and *advanced* (superior performance). A number of states have also introduced voluntary testing in parallel with the national scheme, and examples of these will be reported below.

#### 4.1.1 Curriculum

NAEP has established Frameworks for ‘what students should know and be able to do in geography, reading, writing, mathematics, science, US history, the arts, civics and other academic subjects’ and also produces The Nation’s Report Card, both of which may be found on their website ([nces.ed.gov/nationsreportcard/science/](http://nces.ed.gov/nationsreportcard/science/)). A new science framework was drawn up for the 1996 NAEP testing round, and may be found on the NCES website. There is a good deal of commonality between this framework and others we have encountered. It incorporates three strands roughly equivalent to the formal disciplines of Earth Science, Physics/Chemistry and Biology (called Fields: Earth Science, Physical Science, and Life Science), a process strand (Knowing and Doing Science, consisting of Conceptual Understanding and Practical Reasoning) and a focus on the nature of science and technology. In addition, a number of themes are given in which these matrix elements may be developed.

#### 4.1.2 Testing

The results of the 1996 science tests were analysed by O’Sullivan and Weiss (1999). Approximately 131,000 students in Grades 4, 8, and 12 participated in the main and state science assessments. Each student three sections assigned by a sampling technique from 15 science sections, each of which included multiple-choice, constructed-response and hands-on tasks. Overall, there were 165, 219 and 59 respectively of these tasks, distributed across the three Grade levels. Sample questions are included on the website, but were not accessible at the time this report was compiled.

#### 4.2 Project 2061

This catchy title has been adopted by the American Association for the Advancement of Science (AAAS) for its *Benchmarks for Science Literacy* 1993, the year 2061 being that in which Halley’s Comet is expected to make its next appearance in our skies. The *Benchmarks* represent not a standard curriculum but a powerful tool for educators to use in fashioning their own curriculums, and it is recommended that it be used in conjunction with a 1989 publication of AAAS entitled *Science for All Americans*.

The *Benchmarks* consist of a series of threshold statements, and they are organised into 12 sections representing traditional ‘content’ areas – Mathematics, Physical Science, the Human Organism, the Nature of Technology – together with historical perspectives, the designed world and habits of mind. In each strand, the threshold statements are listed by broad groups of school Grades: K–2, 3–5, 6–8 and 9–12.



Each of the major strands is subdivided. For example, the Physical Setting breaks down into the universe, the earth, processes that shape the earth, structure of matter, energy transformation, motion and forces of nature. The expectation for Grades 3–5 in the ‘Forces of Nature’ sub-section are that students experiment with and understand the concept of force at a distance, as exemplified by the gravitational pull of the earth and the forces existing between magnets and between electrically charged objects.

### 4.3 Examples of state testing programs

Before presenting examples of state testing programs, we note that the National Center for Fair Testing (FairTest) maintains a website ([www.fairtest.org](http://www.fairtest.org)) in which it sets out its principles and provides critiques of state testing in three categories: K–12, university admission, and employment (including employment of teachers). Its most recent report, *Testing Our Children: A Report Card on State Assessment Systems* which is also found on the website.

Assessment practices in two states are examined here. Many state programs could have been selected but we have limited coverage to two jurisdictions that exhibit features of interest and for which adequate information is available.

#### 4.3.1 Michigan

Michigan state conducts an extensive testing in mathematics, reading, science, and writing, under the auspices of the Michigan Educational Assessment Program (MEAP), which has recently been transferred to the Department of Treasury following its initiation in the Department of Education. Science testing for Grades 5 and 8 was introduced in 1994, and Grade 11 in 1996, following publication of the state’s Essential Goals and Objectives for Science Education in 1991. Test items and tests are prepared by staff of the MEAP in conjunction with the science Content Advisory Committee, a 60-member body comprising mostly active teachers. At present there is testing in Grades 4 and 7 in mathematics and reading, and in Grades 5 and 8 in writing and science. Parents may exempt their children from the tests but only 1–2 per cent take this option.

Science tests are based 72 per cent on multiple-choice items (compared with 89 per cent in reading and 65 per cent in mathematics, the writing test being of quite different type), the remainder being open-response items, but there is also a hands-on science investigation. The test is not limited in time.

Commenting on validity of test scores, the MEAP newsletter noted recently that ‘like all published achievement tests, the MEAP assessments have a blueprint that indicates the objectives to be tested in each content area. There is an infinite number of ways to write test items to measure each objective, and multiple forms are composed for each test. Not all objectives are tested in any given form of a test. Both easy and hard items (difficulties assigned by the states’ educational advisory committees) are used in every form to balance the difficulty level of the items and to equate the different forms to one another. The sample of items chosen for a form of a test represents the domain of all possible test items that fit the blueprint. For a student to do well on a

test, he or she must have mastered the entire domain – not just bits and pieces. ... Content Advisory Committees ... verify that each test question meets the objective it is supposed to measure and fits the blueprint or framework. A Bias Review Committee then verifies that the items are not disadvantaging any particular group.’

This process ensures content validity, but criterion validity and construct validity, two other matters of concern to psychometricians, are also examined, particularly for tests administered to higher-year students. The results of MEAP tests are based on total test scores, not the scores on individual strands. Reliability measures – Cronbach’s coefficient alpha – for MEAP tests are publicly available on the MEAP website. For example, the reliability value for science in Grade 5, 8 and 11 tests in 1998–99 were respectively 0.886, 0.892, and 0.878; tests in other fields generally showed lower reliability but ranged from 0.610 to 0.962. The results of tests are reported school by school, including results for public (state) schools, in which testing is mandatory, and non-public schools for which participation is voluntary. Evaluative comments on overall results for the state are also provided. On an arbitrary scale, scores below 350 are described as not-yet-novice, 350–399 as novice, and 400 and above as proficient. In winter 1999 testing at Grade 5, 37.5 per cent of student results were graded as ‘proficient’, down from 40.4 per cent in the preceding year. In Grade 8, there was an increase from 22 per cent to 23 per cent in students reaching ‘proficient’ standard.

Examples of the ‘Size and Distance’ items from Winter 2000 tests are included in Appendix D3.

The FairTest commentary on Michigan’s testing begins with the judgment that significant improvement is needed. Recommended improvements include significant expansion of constructed-response tasks that at present are outnumbered by simple multiple-choice items.

### **4.3.2 California**

#### **4.3.2.1 Science content**

Stemming from a 1995 state decision, California has introduced *The Challenge*, a standards-based school district reform initiative in which schools are challenged to ‘state clearly and publicly what each student should know and be able to do at the end of each year in each subject area’. Standards have been developed in science and ten other areas, including academic disciplines and others such as career preparation, and exemplars and sample demonstration tasks are included. Relevant text is available on the Department of Education website ([www.cde.ca.gov/challenge/Contents.html](http://www.cde.ca.gov/challenge/Contents.html))

but the science file was not accessible at the time this report was compiled. However, the California State Board of Education has a web page devoted to science which includes curriculum information together with resource material for teachers and students –Teacher’s Place, Kid’s Corner, Science Search, and Ask a Scientist ([www.cde.ca.gov/board/science.html](http://www.cde.ca.gov/board/science.html)). Under the general heading SCORE the Board publishes the California Science Framework (not yet available on the website), and the Science Content Standards which are presented as follows:

- Kindergarten
- 1st through 5th Grades
- 6th Grade (focus on earth science)
- 7th Grade (focus on life science)
- 8th Grade (focus on physical science)
- for Grades 9–12, separate strands in Physics, Chemistry, Biology/Life Sciences, Earth Sciences, and Investigation and Experimentation.

Frameworks are commonly adopted for the presentation of school curriculums, and the strands are those used in most frameworks, although we note the use of traditional names for them. A ‘process’ strand is included, as it is in most frameworks. The material in the Contents Standards documents for the primary school years is divided, for each Grade level, into four sections – Physical Sciences; Life Sciences; Earth Sciences; Investigation and Experimentation – each of which contains one or more ‘concept’ sub-headings followed by lists of things that ‘students know’. For example, 4th Grade content consists of the following (only the third concept entry is expanded to show typical outcome statements):

#### **Physical sciences**

- Electricity and magnetism are related effects that have many useful applications in everyday life.

#### **Life sciences**

- All organisms need energy and matter to live and grow
- Living organisms depend on one another and on their environment for survival

As a basis for understanding this concept, students know:

1. Ecosystems can be characterised in terms of their living and nonliving components.
2. For any particular environment, some kinds of plants and animals survive well, some survive less well, and some cannot survive at all.
3. Many plants depend on animals for pollination and seed dispersal, while animals depend on plants for food and shelter.
4. Most microorganisms do not cause disease and many are beneficial.

#### **Earth sciences**

- The properties of rocks and minerals reflect the processes that formed them.
- Waves, wind, water, and ice shape and reshape the Earth’s land surface

## Investigation and experimentation

- Scientific progress is made by asking meaningful questions and conducting careful investigations.

We may compare this Californian framework with Australia's national curriculum profile (Curriculum Corporation 1994), equating our Level 2 or 3 to California's Grade 4. Although there is some superficial similarity of topic titles – in Life Science, for example, The Interdependence of Organisms – the Californian framework entries are narrower and rather more advanced. For example, the positive electricity and magnetism in this level, whereas the Australian framework introduces it at Level 4, the early secondary level.

## Golden State Examinations and STAR

The state has developed tests (Golden State Examinations) offered on a voluntary basis to students in Grades 9–12 in a variety of subjects and employing various combinations of criterion referenced multiple-choice and constructed-response items. This replaced the California Learning Assessment System which FairTest regards as 'perhaps the most controversial state exam in the nation' but which was abandoned in 1994 after conservative opposition to it. Its replacement, California's Standardized Testing and Reporting (STAR) program, is based on commercially prepared tests.

Taking tests is mandatory for all students in Grades 2–11, but parents are able to request in writing that their children be exempted from the tests. No information is available on how many do so.

The tests mainly employ multiple-choice items covering reading, written expression (language), spelling and mathematics for Grades 2–8, and reading, writing (language), mathematics, science and social science for Grades 9–11. There is no mandatory science testing at primary school level, although some school districts administer a science test under a parallel program known as Stanford. Details may be found on the websites <http://star.cde.ca.gov/> and <http://207.87.27.181/star/star99/>. While aggregate data are reported on the websites, no individual student results are provided since these are confidential and can be reviewed only by the teacher, the student, and the student's parent or guardian.

## 5. England

### 5.1.1 Curriculum

The UK National Curriculum, introduced in the early 1990s, following the passing of the Education Reform Act 1988, has been revised following the Education Act 1997 and changes will be introduced progressively in 2000–2002 although most change will occur in 2000. Science is one of 13 areas of study, and was little affected by the recent changes, since the Key Stages and the Statutory Assessment arrangements remained unchanged, the major changes to the National Curriculum were to:

- make more explicit the requirements for primary school teaching;
- add extra subjects such as Citizenship;
- improve inclusiveness; and
- revise the English requirements for, and extend other subjects to, Key Stage 4.

The National Curriculum takes a ‘framework’ approach, with full details available in published booklets and on the web site ([www.nc.uk.net/](http://www.nc.uk.net/)). The Science curriculum prizes knowledge, skills and understanding, using the traditional vertical subdivision of science into three content areas – Life Processes and Living Things, Materials and Their Properties, and Physical Processes. These are accompanied by a Scientific Enquiry strand which is expected to be taught in the contexts of the other three. The horizontal subdivision is more complex, with Key Stages, Levels (at which most students are expected to work), and expected Attainment Levels at prescribed student ages, as shown below.

Key Stage	Levels	Expected Attainment	Student Age
1	1-3	2	7
2	2-5	4	11
3	3-7	5/6	14

An interesting feature of the National Curriculum is its requirement that ‘during the Key Stage, pupils should be taught to, for example, use appropriate scientific language’ and to communicate, recognise hazards, and so on. In Australian curriculum frameworks, the emphasis is on outcomes - ‘the student is able to ... (for example) ... recognise the similarities between similar elements’ and appropriate contexts are given in which this recognition may be demonstrated.

Key Stage 2, at which students would be approximately 11 years of age and completing primary school, seemed to be an appropriate level at which to compare English and Australian frameworks. Whereas at Level 1, students are expected to ‘observe, explore and ask questions’, at Level 2 they ‘learn about a wider range of .. (subjects) .. and begin to make links between ideas and to explain things using simple models and theories’. In the Life Processes and Living Things strand, for example, there are separate sub-sections for life processes, human and other animals, green plants, variation and classification, and living things and their environment. Although presented in more traditional language and in considerably more detail (it is, after all, a curriculum, not a curriculum framework), the English version closely resembles that of Australia’s national framework or the Victorian CSF II, both taken at Level 3.

### 5.1.2 Testing

The statutory assessment instruments to accompany the National Curriculum are provided by the Qualifications and Curriculum Authority (QCA), which came into being in October 1997 to bring together the work of the National Council for

Vocational Qualifications and the School Curriculum and Assessment Authority. The QCA has drawn up standards for English and Mathematics at key Stage 1, and for English, Mathematics and Science at key Stages 2 and 3, together with the corresponding statutory tests. Students are expected to take tests at 7, 11 and 14 years of age although a minority may be slightly older or younger. While the standards and also QCA reports on each year's testing are available on their website ([www.qca.org.uk/ages5-14/standards.htm](http://www.qca.org.uk/ages5-14/standards.htm); see [www.qca.org.uk/news/press/00-january13.htm](http://www.qca.org.uk/news/press/00-january13.htm) for the 1999 results), the test themselves are available for purchase from QCA at moderate cost.

In the area of interest to us, that of primary school science, the results of 1999 tests for 11-year-olds (there is no science testing at the earlier stage) showed that while 70 per cent of students reached Level 4 in English (81 per cent in reading, 56 per cent in writing) and 69 per cent in Mathematics, the proportion achieving this Level in Science was 78% – a substantial improvement on the 1998 test result of 69 per cent. Detailed comments covered:

- Secure understanding of Life Processes, notably the function of the human skeleton and the requirements for plant growth, although only a third of students could locate germination within the life cycle of a plant.
- Improved understanding of Materials and Their Properties, including thermal conductivity, electrical insulation and the process of condensation, although only half recognised that copper is not magnetic.
- Electrical circuits depicted by symbols, a task completed successfully by two thirds of students (up from half the previous year), two-thirds of children understanding the formation of shadows (double the proportion in 1996 and 1997 when this topic was tested), but less than one-fifth demonstrating an understanding of gravity.

In a recent trial, scripts were returned to students after marking, and this has been generally well received ([www.qca.org.uk/news/press/99-october21.htm](http://www.qca.org.uk/news/press/99-october21.htm)) by students and by teachers, who found it useful to be able to discuss the results with parents. Most benefit to students seem to have been derived when scripts were discussed with teachers, and there was no evidence that requests for remarking were more frequent where scripts had been returned. The QCA noted that 'returning scripts to students made the examination system more transparent and examiners more accountable' but doubted whether the benefits outweighed the costs of returning 13.5 million scripts! An 'on request' model is under consideration.

A few test items are included in Appendix D4, together with the relevant mark schemes which show how labour-intensive the marking of such items must be, and consequently reveal the great investment made in testing in England.

## 6. International studies

## 6.1 Second International Science Study

The Second International Science Study (SISS), conducted in 1983–1984 under the auspices of the International Association for the Evaluation of Educational Achievement (IAEAA), collected data from 17 countries for students in the final year of secondary schooling who were taking science subjects which would permit them to undertake university study in those disciplines.

Although SISS itself lies outside the scope of the present report, we may glean from it some information which will be useful to those developing a national testing regime. The problem of administering broad-ranging tests to students who might have specialised in biology, chemistry or physics during the final years of their courses was addressed in SISS by having all students take a core test at about Year 9 standard – this being the last year in which they might all have been expected to share a common curriculum – and then having the three sub-groups sit more specialised tests in their chosen disciplines.

Performances for the sample of Australian students placed the group just into the upper half of countries ranked by performance, but the SISS is of little relevance to the present report except that discussion of the results places great stress on the sampling methodologies which must always be kept in mind when international comparisons are made. For example, performances by different country groups are negatively correlated with school participation rates, and correlate better with years of schooling than with age. This would not be a factor that would affect inter-jurisdictional comparisons of achievement in primary science in Australia. A succinct account with references to full reports is provided by Rosier (1996).

## 6.2 TIMSS

The Third International Mathematics and Science Study (TIMSS) was conducted in 1994–1995 and involved students from five Grade levels in 41 countries. Altogether this amounted to over half a million students. Contextual data were gathered by means of questionnaires to thousands of teachers and school principals so the overall result was the generation of enormous quantities of data, which were released in 1996 and 1997 in a series of reports (<http://timss.bc.edu/>).

### 6.2.1 Curriculum framework

Behind the program of testing lie the TIMSS Curriculum Frameworks, which in the Science field include:

- Content: material in earth sciences, life sciences, physical sciences, science/technology/mathematics, history of science and technology.
- Environmental issues, nature of science, and science and other disciplines.
- Performance expectations: understanding, theorising/analysing/solving

problems, using tools/ routine procedures, investigating the natural world, and communicating.

- Perspectives: attitudes, careers, participation, increasing interest, safety, and habits of mind.

This is a more extensive and sophisticated list than that in other framework documents we have seen – that is, it is more a curriculum than a curriculum framework – and it derives from consensus developed among working groups around the world, led by TIMSS National Research Coordinators. The Australian Council for Educational Research was closely involved, being responsible for scaling the data and providing data analysis. Extensive reports on data, technical matters and analyses have been produced and references to these may be found on the TIMSS website. In addition, TIMSS Monograph No.1 sets out the curriculum framework (Robitaille *et al.* 1993).

### 6.2.2 Testing

Testing was conducted at three levels: the first involved students enrolled in the two adjacent Grades that contained the largest proportion of 9-year-olds – Grades 3 and 4 in many countries – and the second, 13-year-olds (Grades 7 and 8). The third level took in students in the graduating cohort. The middle group is regarded as the core of TIMSS, and all 41 countries participated at this level, while only 26 and 25 respectively participated at the first and third levels. The achievement tests were limited to Mathematics and Science for the first two levels (3rd/4th, 7th/8th Grades), but in the final years' tests they covered Physics and Advanced Mathematics as well as the follow-on testing of Mathematics and Science Literacy. Detailed background to the last of these may be found in TIMSS Monograph No. 4 (1988).

While most questions in the TIMSS tests were of multiple-choice type, approximately one-quarter were in free-response format which required students to generate and write their own answers. About one-third of the items are kept secure by IEA for possible future use in measuring trends in achievement, but the rest are available in books for purchase, on the website <http://timss.bc.edu/TIMSS1/Items.html> and as PDF files on another site accessed from it:

<http://timss.bc.edu/TIMSS1/TIMSSPDF/ASitems.pdf>.

More information about items, and about aggregate results, may be found in IEA publications (M. Harmon, T.A. Smith, M.O. Martin, D.L. Kelly, A.E. Beaton, I.V.S. Mullis, E.J. Gonzalez, and G. Orpwood, *Performance Assessment in IEA's Third International Mathematics and Science Study* (TIMSS), TIMSS International Study Center, Boston College, 1997).

Examples of both kinds of item, and the listing of items under the various branches of science, are shown in Appendix D5. For each multiple choice item, the correct answer is provided at the foot of the page, together with information about the performance expectation and the percentage of correct results observed. For free-response questions the categories of responses and the scores these would attract are shown on the following page, together with the aggregate information as before.



An opportunity for participating countries to monitor progress was afforded by a repeat study, TIMSS-R, in late 1998 (southern hemisphere countries) and early 1999 (northern hemisphere). This involved 8th grade students in completing achievement tests in mathematics and science and questionnaires about classroom experiences, attitudes towards science and mathematics, and certain policy issues. Release of the results is planned for 2001.

## 7. Summary and conclusions

Comparisons among practices in different jurisdictions have revealed some significant differences but the underlying causes for these are not easy to discern. In what follows, we draw on international experience to make some general observations.

Broad scale assessment for monitoring purposes seems to have begun seriously only in the 1980s with some acceleration in the 1990s, so it is a comparatively new phenomenon. Its effects are therefore hard to assess, but some advances are evident, and these include the development of curriculum frameworks (almost a universal phenomenon), development of assessment instruments, the heightened awareness of governments, and the bringing together of educational researchers and practitioners to build the emerging systems. Whether there has been much improvement in student learning and achievement is harder to tell. Most educational literature is silent in this point, or relates to superseded practice. Overall, it might be simply that there has been insufficient time for any feedback to have had its effect on teaching and learning. Similarly, governments will not have had time to judge whether their financial support for education has been adequately and properly directed. A second confounding factor is that a number of jurisdictions have changed their assessment schemes during what we might regard as a decade of development, and so the necessary consistency in practice has simply not existed. The need for regular, consistent assessment has been specifically addressed by the OECD PISA project.

The development of *curriculum frameworks* has been widespread. These have included the customary separation of earth, physical and life sciences, under a variety of old and new names, but also in most cases the development of ideas about the nature and practice of science. Occasionally this 'process' strand is implicit, or incorporated in more traditional 'content' strands, but more often it is presented separately although it is made clear that it is not to be taught separately. The most sophisticated framework embodiment of the orthogonal relationship between the two is that of New Zealand, in which the two domains, together with year progression, make up a three-dimensional curriculum framework. While most jurisdictions present the framework lists as the basis for what needs to be learned, sometimes with appropriate descriptors of how this will be evidenced, in England the emphasis is on what should be 'taught'.

A few jurisdictions combine *science and technology*, especially in the junior years, while others present them separately but indicate links between them or (as in the United States) discuss technology along with other science 'process' matters. This offers a guide to Australia for resolution of the potential dichotomy, especially as

‘technology’ in the Australian context usually implies construction and/or design – stemming from the old Technical School system and closely allied to Technical and Further Education (Vocational Education and Training) – rather than machines and processes which the scientist would normally regard as examples of technology.

International experience strongly indicates that testing should be conducted with a range of test items, including *multiple-choice*, *open-response*, and *practical*. Of these, the first is the best-developed mode and lends itself to rapid, reliable marking. Open-response questions seem to be more favoured by educational curriculum experts, although this preference may be more based in faith than logic. It is widely agreed that science education should go well beyond rote learning and assessment well beyond regurgitation<sup>3</sup> but it is not evident that open-response questions have done more than lower the success rate of such strategies, while rewarding other learning strategies. The provision of exemplar responses, included with the Appended examples of open-response items, shows how the difficulties of marking open-response items may be eased, but also reveals preferred responses which might not come easily to the minds of students. Practical exercises are even more popular with science curriculum experts but the assessment regimes we have encountered usually still really involve pencils and paper. Practical tests with groups of materials to be classified (rocks or plants, for example), or quantities to be measured (length, mass or volume) are much more time consuming, especially of teacher time.

Assessment items, usually grouped in *tests*, are made available to schools and then after the testing period they are made publicly available in printed or Internet format, although some (as in TIMSS) may be retained as benchmark items for future testing. In New Zealand is there an Assessment Resource Bank available from which teachers may select items. This practice is especially suited to the use of tests for teaching and learning in schools but may not be acceptable for a mandatory testing regime unless some constraints were applied, as with Victoria’s VSAM (Victorian Student Achievement Monitor) tests.

International practice suggests that *reporting* serves a number of functions and needs to be tailored to these ends. Thus, individual results are generally made available only to students and their teachers and parents/guardians, while aggregate results for schools (in some cases), school districts, states or the national cohort are made more widely available in printed reports and on the Internet. An alternative to numerical scores is the adoption of achievement points, as in Michigan, so that a student’s result may be reported as falling into a particular category which has a verbal descriptor (satisfactory or proficient, for example) and aggregate results reported in terms of the proportions of students reaching particular standards. Most jurisdictions report an achievement level based on the whole science framework, but in subsequent analyses of the massive data sets accumulated in the testing process it is common to disaggregate so that performance on particular types of questions or in particular content areas can be scrutinised, and even to disaggregate on the basis of personal or

---

<sup>3</sup> There is an old-fashioned prejudice that multiple-choice items lead to rote learning. This would be so only if the items were badly designed. Multiple-choice items can be used to assess mostly levels of Bloom’s Taxonomy (but not assess creativity and ability to synthesise). (See Appendix D.)

locational attributes of students. Returning scripts to students, as is being trialed in England, seems unlikely to become widespread because of cost although it would have undoubted educational benefits for the teacher–student interaction if it could be done quickly enough, which seems unlikely in any but the smallest jurisdictions or with the smallest of samples.

Some jurisdictions test *all students in a cohort* (Michigan, California, England), while others use a *sample of the population* (United States, TIMSS). We accept that reliable samples of populations can be selected, although great care would need to be taken with this. There is a major difference in cost, depending of course on the proportion of the population sampled, but there are fixed costs, of developing the test instruments, for example, so the savings resulting from sampling may not be as great as simple considerations suggest. The use to which test results are put might influence the choice of a sample as against the whole cohort. The results from a carefully selected sample would be sufficient to service any interest in aggregate data, to monitor the ability of schools to deliver the curriculum and of students to master it. It might appear that this sampling option would deny most students and schools the opportunity to monitor their own achievement levels, and to compare them with system-wide achievement. However, this potential problem could be overcome if schools not in the sample had the opportunity to take the assessment at some date after the sampled schools had been assessed. A rolling sample might take in all schools over a period, but unless there is annual testing of a large sample this ‘period’ could be 20 years!

Science testing in primary schools internationally seems to centre on *Grade 4* students in most jurisdictions although there are some where testing is conducted at every year level and some where testing in primary school years is limited to literacy and numeracy (English and mathematics), with science introduced only in the secondary school years. When testing is conducted at two levels the progress of a cohort can be monitored, but local experience suggests that while the results of such linear studies are interesting to researchers there are no consequent actions that might follow from observed changes in achievement, and so the exercise may not yield sufficient value to make it worth the cost. It is true that having two points in time where assessments are made enables the use of value-added indices to show growth from time to time. This is not the usual practice.

Cost is also a factor in deciding whether to *test every year* or to administer tests on, say, a *three-yearly cycle*, as is proposed for 15-year-olds with PISA. ‘Cost’ is a surrogate term for effort too (industry teacher effort), since the preparation of test material, and the compilation and analysis of results are likely to prove a formidable task if undertaken on a national scale. One must also consider whether testing will be taking place in other fields of study, at least in English and mathematics and possibly in history, with the result that too much student and teacher time would be consumed in testing, at the expense of learning. Finally, the response times of educational systems facing demonstrated need for change makes it seem unlikely that there would be sufficient need for annual testing.

## **PART 3: AUSTRALIAN CONTEXT – CURRICULUM**

### **1. Introduction**

The process of educational assessment and reporting can only be adequately considered in the context of the curriculum. Assessment and reporting are tools that enable a better understanding of the adequacy of the curriculum. While assessment can have multiple purposes (monitoring, planning, improving teaching, accountability, diagnosis, motivation, to name a few) all such purposes ultimately refer to curriculum in some way.

For this reason, this report, having examined the international context of primary science, now looks at primary science curriculum within Australia. The evidentiary bases of Part 3 are twofold:

1. the documents presenting the stated curriculum across Australian states/territories; and
2. the interviews conducted with primary science experts mainly from government, catholic and independent authorities. There were more than 30 such interviews (see Appendices B and C).

In reporting the interviews, care has been taken not to assume anything other than a subjective sample of knowledgeable curriculum and assessment managers many of whom are focused on primary science. The objective here is to present an accurate picture of the primary science curriculum across Australia as it relates to the latter parts of this report dealing with assessment and reporting.

This will not be a state-by-state-by-territory microdetailing but a presentation of ideas or themes that do not know geopolitical boundaries. Where a theme is of particular importance to a state/territory/sector the locus will be mentioned. However, the intent is not a detailed description of each curriculum.

A cautionary note should be made at this point that these two evidentiary bases relate mainly to the stated or intended curriculums across Australia. They do not provide information on the 'received' curriculum – that which is actually occurring in the classroom. Outside the strict disciplinary convention of the science lesson (and we assume, perhaps falsely, that the science lessons conform to the stated curriculum) lie many scientific activities (processes) and concepts embedded in other lessons (see Part 1). As well, primary students have access to a range of stimuli for learning outside the classroom.

The achieved curriculum (that which students actually learn) is not a simple function of the intended classroom curriculum. From television, the Internet, independent reading, peers and older siblings, family conversations and other less obvious sources, students learn science and about science. Curriculum experts and classroom teachers

sometimes ignore this complexity when they argue that assessment tasks are too difficult for their students. A proper monitoring should be able to look beyond the parochial boundaries of a given state's intended curriculum when developing an assessment framework.

This cautionary note is presented in the Introduction to the focus on curriculum because it is important to realise that assessment scope and procedures (to be dealt with in the next Part of this report) should not be so narrow as to miss the larger picture of student learning.

### **The intended curriculums**

Two of the most controversial issues encountered across Australia both between states and within states (sector vs sector) are:

1. the question of the respective roles of facts and concepts (knowledge) and of processes and skills (thinking and working scientifically); and
2. the degree of commonality and difference to be noted among state/territory curriculums.

It would seem that the rhetoric and the reality on both these issues are sometimes at odds with each other.

### **The curriculum emphases in primary school science**

Arguments about curriculum emphases were to be heard in various forms throughout Australia. However, when challenged some of the sharper differences disappeared. Virtually everyone agreed that the scientific process strand was an essential ingredient. There was less agreement about content. Many who argued that there was no mandating of science content nonetheless agreed that:

*Science skills could not be taught in a vacuum and that some content was inevitable.*  
(This agrees exactly with the PISA definition of 'science literacy'.)

Even though specific content was not mandated, there would be culpability if students, at the end of their primary school years, had not mastered certain concepts and principles underlying science knowledge (for example, that our moon orbits Earth and that Earth orbits our Sun, that ecosystems illustrate the interdependence of living things.)

The four 'content' strands emanating from the 'National Statements and Profiles' (*viz.* under various semantic guises, Life and Living, Earth and Beyond, Natural and Processed Materials, Energy and Change) were virtually common across Australia just as they were found to be internationally.

The most extreme forms of argument occurred between two sectors in one state where one sector argued for the importance of science content (and indeed had produced teacher support materials to provide teachers with interpretations of core content but not 'all of the possible content'. (The State of Queensland 1999.) Thus, at Level 2 under the Earth and Beyond strand, outcome 2.2 the teacher is informed of such content areas as: 'The Earth is part of the solar system. The moon orbits the Earth.' and, 'The planets in the solar system change their position in relation to the sun.'

In contrast, a science education leader of one of the non-government sectors in the state argued that it was of no concern what content was taught but what was of concern was that students learn to think scientifically and gain expertise in the process skills such as observation, drawing inferences, developing fair tests and being able to describe accurately. This was an extreme view that seemed to be held by only a small minority.

It is noteworthy that the PISA framework definition of science literacy is a clear summary of what states/territories are indicating in their frameworks and curriculum documents. As we saw in Part 1 of this report, this PISA framework involved:

- A relatively dismissive attitude towards facts as important educational goals in themselves.
- A strong endorsement of scientific thinking and scientific processes as essential to primary science curriculum.
- A realisation that scientific thinking and scientific processes develop mainly in the context of scientific concepts and principles. The processes of science cannot be taught in a vacuum and primary school students should have a grasp of basic scientific concepts across the four content curriculum strands.

It is fair summary to suggest that despite varying emphases, there was general consensus that developing the process skills involving 'scientific thinking' was an essential and major part of primary school science. There was almost as much consensus that certain basic scientific concepts and principles should be understood by students at the end of their primary school years. However, there was less certainty as to what those concepts and principles were. There was also consensus that there should be no emphasis on the learning of facts per se.

Inevitably, we will return to the topic of curriculum emphases later in this report.

### **The degree of commonality and difference to be noted among state / territory curriculums**

Again, the rhetoric and reality were not identical. At one extreme the curriculum policy was to divest responsibility to schools; and indeed in some states with respect to independent schools there was, as the sector name states clearly, independence. However, even here there was evidence of a considerable centripetal force under the auspices of a mandated curriculum framework. Virtually all states/territories paid

allegiance to a common source, the national statements and profiles of the early 1990s. While many had made their own adaptations, a core of commonality seems to remain especially in terms of strands and levels. Even in systems such as the ACT where schools are given considerable independence, this has not led to fragmentation and curriculum idiosyncrasy in science.

The work being undertaken by the Curriculum Corporation (Science Online) is still in its early stages. Nonetheless, the work carried out to date confirms the point that 'online' science can be a useful aid to schools because there is commonality across the states/territories in science curriculum.

While some detail might lead to argument and there is no certainty that a given set of facts or even topics are addressed in any specific primary school, it is also true that there would be expectation that most well-taught students would have learned a range of scientific concepts. Even more certainly it would be agreed that students should have learned certain process skills such as observation, developing fair tests, objective viewing of evidence and logical drawing of inferences and conclusions.

The science concepts and principles that seem to be accepted commonly across Australian states/territories include energy resources, energy transfer or transformation in simple devices, alternative energy sources, light and shadow, sound vibrations, earth's structure, earthquakes and volcanoes, soil erosion, the solar system, ecosystems, living things have features that suit their natural environment, animal body systems and endangered species.

This listing, adapted from an in-house Curriculum Corporation document, is not intended to be a comprehensive listing, nor is it the result of an in-depth study by the authors of this report using content analyses and in-class observations. However, it is presented to foster confidence that such a listing can be generated.

In short, one could argue that across Australia there is a corpus of scientific skills and concepts worth monitoring. It may well be that some of the concepts are more accessible in some schools than in others, but obtaining information about what has been learned and what has not across this corpus of concepts and skills could be useful for curriculum planning, resource allocation and professional development.

In a short postscript to this report (see Postscript) we shall suggest means by which this issue of curriculum commonality with regard to concepts and principles might be addressed.

Despite the arguments presented to this point, there is no denying that individual states and territories have some singularities. For example, the Northern Territory emphasises biological and environmental content in its science curriculum and New South Wales joins science and technology together rather than presenting them as separate studies. The possibility of encouraging states/territories to indicate the likelihood of the students encountering particular concepts and skills in their classroom learning should be advocated. Reporting could not only indicate performance in particular areas but also whether this area was emphasised or not.

From a monitoring viewpoint the presentation of this more sophisticated performance would lessen anxieties.

## **2. The role of teachers**

Curriculum is mediated by teachers and teachers in primary schools are typically lacking in confidence when teaching science. Their own secondary education often lacked science beyond Year 10. If science were taken in Years 11 and 12 it was usually biology. Primary teachers are usually particularly concerned about having to cope with topics emanating from physics and chemistry.

Interviewees also noted that in pre-service teacher education courses it was not uncommon for the education student to take only a one-semester, two hours per week science course in total.

With respect to professional development, while primary science programs were available many of the least prepared primary teachers neglected such programs. The word 'scared' cropped up even in the naming of professional development science programs.

It is probably true that the expertise of the teacher accounts for only about 15 per cent to 25 per cent of the variability in assessed performance of the students. Nonetheless, 'only' is an emotive term and the role of teachers in terms of the achieved curriculum is deemed important. More crucially it is an issue about which something can be done.

In passing, some science educators noted that the area where primary teachers often indicate they feel most vulnerable is their lack of science knowledge. However, many professional development courses emphasise the pedagogy of science teaching. This irony is understandable but bears further examination.

In Part 1 of this report, reference was made to the idea of ensuring that if Ministers were to implement a monitoring process (assessment and reporting) in the area of primary science, it should occur within the context of a positive program package. Foremost within that package should be a professional development program that addresses the elements underlying the assessment and reporting process.

It is not in the province of this report to suggest the details of this package nor of the professional development component. Nonetheless, it could include the products and services offered as a result of the Curriculum Corporation's Science Online project and it could include professional development programs dovetailed into existing state/territory programs. Those developed as part of the positive package associated with the monitoring process could be offered online (using Science Online methodologies) or in more conventional ways.

In any case, professional development will continue to be a need in primary science. There is considerable change occurring in the curriculum documents produced by



states/territories. Most of the change is evolutionary rather than radical but the need is obvious for professional development to help teachers understand the changes.

In turn, monitoring becomes relevant as curriculum changes occur in order to be informed of the effect of the change. It is in this context of the intimate relationship between curriculum, curriculum change and monitoring that we advocate the need to enhance teacher professional development in primary science.

### **3. Summary and conclusions**

Curriculum underlies assessment and reporting. The two most important issues to be solved in order to develop a satisfactory monitoring of primary science are:

1. the emphases to be placed on concepts and processes and skills; and
2. the degree of curriculum commonality across the states and territories.

The conclusions reached were that there was commonality across the states/territories and that the PISA definitional framework operationalised the emphases placed by states/territories on the relative importance of facts, concepts and processes. Facts are not an end in themselves but scientific processes and concepts are essential.

The importance of introducing the monitoring of primary science in the context of a positive program package for schools and teachers is emphasised. Specifically, promoting enhanced professional development both in traditional mode and online is advocated.

## **PART 4: AUSTRALIAN CONTEXT – ASSESSMENT AND REPORTING**

### **1. Introduction**

Good assessment is valid and reliable. For the purpose of the monitoring referred to in this report, it means the information obtained from the assessment is clearly related to student performance with respect to primary science and this information is accurate and reliable and useful for its intended audiences.

Good assessment potentially has positive impact on a range of related educational services. It can be used to target resources more effectively and efficiently, to inform pre-service and in-service teacher education programs and to establish a more scientific basis for education decisions made both in the classroom and in head office.

In this part of the report we shall first present the assessment processes we recommend. We will then turn to the other vital topic of reporting.

### **2. Assessment**

#### **The kind of assessment**

There is a strong preference for ensuring that the assessment, if it were to occur, include practical tests of process skills such as ‘creative problem solving’ and ‘open-ended investigating’. These tests would be embedded in scientific issues. Teachers should be involved in such testing. This would be costly and should include teacher-release time for training but the cost would be worthwhile in ensuring the curriculum or face validity of the assessment and the professional development of the teachers. A majority of interviewees indicated that this kind of assessment is an essential ingredient of a performance monitoring exercise. The context of the practical assessment must be scientific and not some problem-solving task out of an old IQ test.

We would also strongly argue that the assessment should also include a variety of other item types including reliable, well-constructed, multiple-choice items and open-ended short-answer items that could be readily scored on a scale that allowed for partial credit for responses that were on track but incomplete (see Appendix F). This kind of wide-ranging item types is present in the Western Australian science tests but is not likely to be present in such international tests as TIMSS.

In short, there would be three kinds of assessment tasks: practical assessments, open-ended assessments and objective items. However, there is no support for items that assess factual information such as the names of certain parts of the flower or the name of the planet that is fifth from the sun. Memory-type, rote-learning items are emphatically rejected. As some of our interviewees said, ‘there are tens of thousands of facts. Which would you ask?’ ‘We don’t want to pummel students with facts.’ However, the need for students to have learned in Gagne’s terminology, (Gagne 1965)

scientific concepts and principles was warmly supported. In similar manner the interviewees were aware that primary students should be able to measure (for example, what is the temperature?) but the essential scientific skill really worthy of assessment is to *know when* to use a thermometer and *why* it should be used. In short, the simple measurement skill while necessary would not be a sufficient outcome.

Only some members of one sector in one state seemed antagonistic to the idea of a national performance assessment. This government sector argued that their state was so different that it would be unfair to lump them into a national assessment. Their counter-proposal was that each state generate its own primary science test. However, there was no support for this idea elsewhere.

The monitoring of primary science is distinct in two major ways from the literacy/numeracy monitoring as currently practised in Australia. First, each state/territory had in existence its own literacy/numeracy assessment. Understandably, most seemed unwilling to abandon that assessment in favour of a common one, although some modifications have taken place to the assessments to create less heterogeneity. As a consequence, there is an annual process of equating the various sets of results. With only two exceptions, states/territories do not have their own science literacy assessment regimes. Thus, the historical context is different.

A second major difference is that the current literacy/numeracy assessments were established by individual states/territories for a range of purposes including, for example, providing information to parents on their child's progress. At least some of those purposes are not major mandates for national monitoring of primary science. Thus, it makes sense for primary science monitoring to be developed in terms of the needs of nationally comparable reporting.

### **The timing of the assessments**

There was considerable thought given to the question of when primary science assessment for purposes of national reporting should take place. At least in the start-up to primary science monitoring, the assessment of students should take place at or towards the end of Year 6. At this stage the student typically has yet to undertake six full years of education before finishing secondary schooling. This is true of Year 6 students in all states / territories. There is clearly no problem to this approach in states such as New South Wales and Victoria, which have a primary – secondary break at the end of Year 6. However, in states such as Western Australia there was some discussion of Year 7 as being a better time to carry out the monitoring. (However, by this year level, most Australian students would have experienced a year of specialist, secondary science teaching.)

The notion of a second testing at another Year level (for example, Year 2 or Year 3) was rejected at this stage. The difficulty of defining the concepts and skills to be tested early in the primary years was seen as one difficulty. The cost of actually testing very young primary students using individual or small group assessment techniques was also seen as a problem.

This leaves a decision on whether to test at year 6 or years 4 or 5. It makes some sense to maximise the opportunity for schools/states/territories with different curriculum routes to teach primary science outcomes (year 6) than to intrude the assessment into an earlier year. The decision will vary depending upon curriculum area. (It is of great importance to check literacy and numeracy by year 3 for example.) However, in Science with the purpose of monitoring, delay until toward the end of primary schooling has the advantages of being able to assess a more mature learner who has had greater opportunity to develop scientific skills and processes and develop a better understanding of basic scientific principles.

### **The frequency of assessments**

There was a consensus that there was no necessity, for monitoring purposes, to assess every year. Every two or three years was seen as reasonable.

The writers of this report recommend that, at least initially, a two-yearly assessment should be undertaken with the option of reducing it to once every third year once primary science as a curriculum area is perceived to have reached a satisfactory standard and once there has arrived a degree of stability among the states/territories with respect to their stated (intended) curriculum documents.

It is worth noting again that while this is a less frequent assessment regimen than the annual literacy/numeracy assessments, the historical context and the purposes are different (see above).

### **Census vs sampling**

This is a question that arises whenever large-scale education monitoring programs are being developed. The value of census (population) testing is that it provides results that are free of sampling error (in that all students are assessed) and that it can be used to report on all targeted students and thus provide information for all parents. The drawback is that it is clearly a relatively costly alternative to sample testing, especially if such costs as teacher release time for training in administering the tests and for marking the tests is factored in.

If a scientific sampling plan – and, as needed, weighting of results – is used, the sampling error can be kept so low as to be negligible. If there is no imperative to inform parents of their children's results then census testing loses that benefit over sampling.

Thus, the authors of this report strongly recommend that a sampling regimen be adopted. The sampling unit would be the school rather than the student. The reason is that to sample students would mean almost every school would participate to some degree and this would be uneconomic in cost terms as well as in terms of teacher time.

With the use of schools as the sampling unit, it would be possible to call for a sample size of about 400 schools. Sampling is a highly specialised skill that can be applied as needed. To carry out the sample planning, it is necessary to specify the levels of error deemed reasonable by the client in relation to the escalating cost as the sample size rises. A consideration will be whether to over-sample in the smallest states/territories (Tasmania and Northern Territory) and in the disadvantaged groups that are reported on separately.

It should be noted that the 400-school sample size is considerably larger than the sample sizes called for in TIMSS and in PISA but the purposes and major unit of reporting differ between those international studies and this monitoring program.

One other aspect of sampling that requires consideration concerns the tasks in the assessment. Item sampling can be used to extend the curriculum coverage of the assessment (students receive overlapping sets of items). It can also be used to reduce the time of testing with respect to the practical tasks if it deemed that these tasks will be inordinately time consuming.

### **Sampled schools and non-sampled schools**

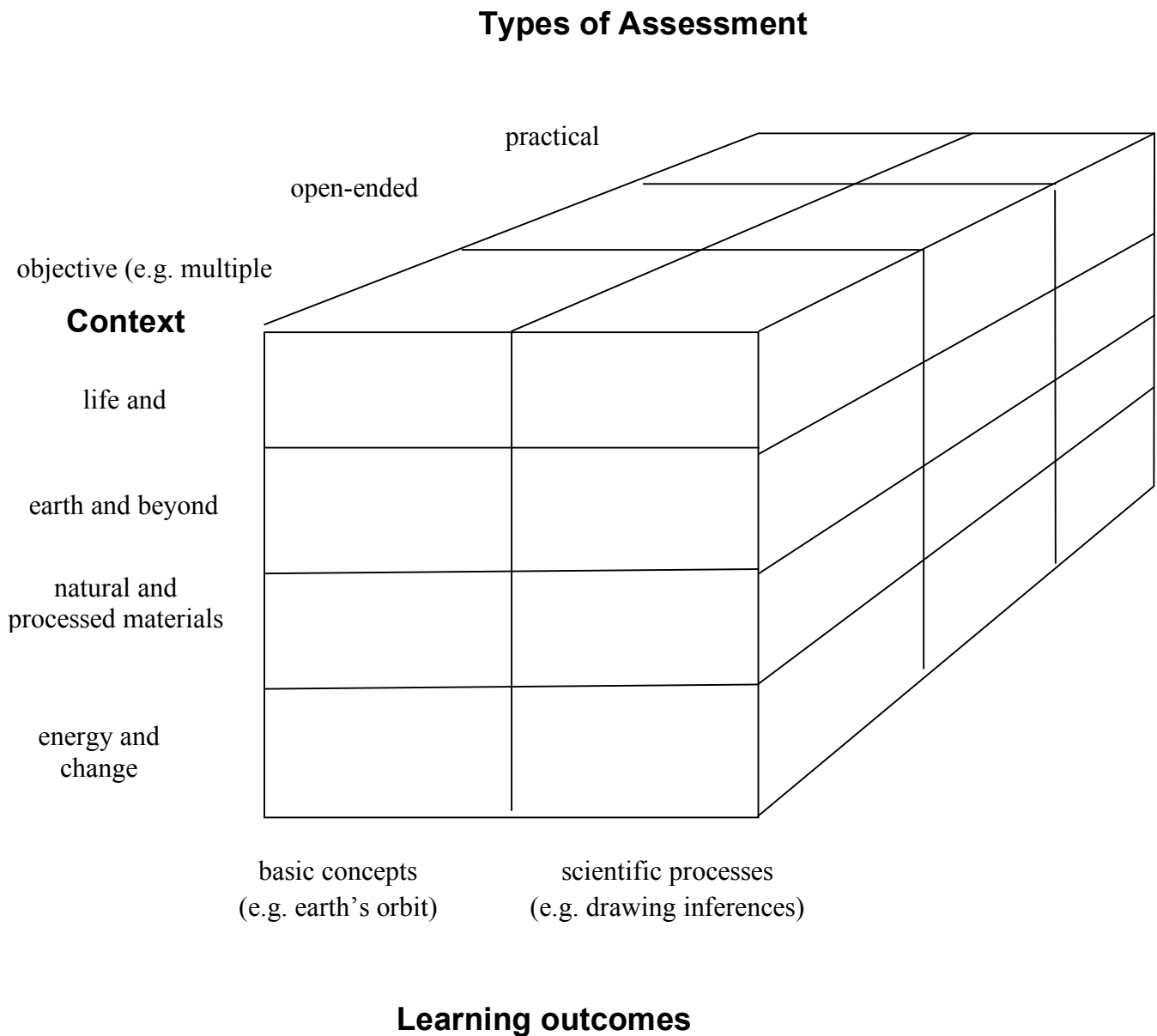
There are benefits to being a school chosen as part of the sample. At least one teacher in the school will be given professional development that can be rippled through the rest of the teaching staff. The professional development should make the teaching of primary science more effective and provide an excellent model for assessing student performance. As well, if the school wishes to do so, the parents of the Year 6 students can be given useful information about the performance of their children (each student's results confidentially to the respective parents).

The question then is what to do with the non-sampled schools. In our view, strong consideration should be given to providing those schools with the opportunity, at a slightly later time, to test their own students. They should be able to use the same comparative data as the sampled schools receive. (See below, Reporting.) That is, the tests (or at least a major part of the tests) should be made available across Australian schools.

**The assessment specification model**

We present now a visual representation of our assessment specification model based upon the foregoing text which, in part, is based upon the views we formed as a result of the interviews.

**Figure 1.** Three-dimensional projection of the assessment model indicating primary science content, learning outcomes and types of assessment



It can be seen from Figure 1 that we are advocating an assessment model that embodies both important basic scientific concepts and scientific processes skills. They would be assessed in the context of the four content strands. The assessment would include three different types of item. The use of this model as a specification for test development would ensure that many of the lessons learned by states in their own assessment programs would be constructively used here. For example, in terms of the Western Australian experience (WA unpublished internal paper 1999) the Figure 1 model specifications also advocates a wide range of item types including practical tasks and a comprehensive coverage of science concepts across the four content strands.

Figure 1 provides not only the specifications for assessment, but also a potential framework for reporting. Assessment without reporting is a barren activity. We now turn to a discussion of reporting of primary science monitoring assessments.

### 3. Reporting

#### Kinds of report referencing

There are two basic kinds of report referencing and this applies irrespective of the audience or of the focus of reporting. One kind is usually termed 'normative'. It expresses how well a focus unit (student/class/schools/sector/state) performs in comparison with some other agreed unit (for example, how well NSW performed in comparison with Australia-wide results; how well a school performed in comparison to like-schools with similar socioeconomic profiles; how well a student performed in relation to others in that state.) Despite criticisms, normative-referenced reporting can be quite informative for a relevant audience (for example, Education Department, school principal, parent.)

A different kind of report referencing which provides complementary and supplementary information is criterion-referenced reporting. Here there have to be standards or levels (criteria) against which the achievement (performance) can be reported.

A student could do well against a standard, but normatively be in the bottom half of his or her cohort. (Are the standards set too low?) Or, a student may be in the top 20 per cent of her cohort but barely reach the standard set for his or her year level. (Is there ineffective instruction or too little instructional time in this key learning area?)

We recommend that both kinds of reporting occur with respect to primary science. In order for this to happen, there will need to be an agreed set of performance levels for, say, Year 6 students. These would presumably be in four bands of performance.

- **Band 1** would indicate a student ending Year 6 was not yet competent in the skill levels and conceptual understandings specified.

- **Band 2** would indicate the student was within a reasonable range, or was about to achieve, the actual skill and concept levels specified.
- **Band 3** would indicate the student had achieved the desired skills and concepts for a student completing Year 6.
- **Band 4** would indicate the student had moved beyond the desired levels.

The assessment instruments and tasks would be classified against these Bands by competent judges who would take into account curriculum documents and related data in arriving at the standards.

In a postscript to this report (see Postscript) we shall indicate a possible means of developing the criteria (standards) to be used in the application of the four Bands.

### **Audiences**

The primary audience would be MCEETYA, its constituent states and territories and the federal government. There would also be other audiences including schools that participate as part of the sample. However, each school would be given mainly information in confidence about that school and, as a reference, the aggregated results of the state from which the school comes. Similarly, parents whose students attend a sampled school should have the right to be provided, in confidence, a report on that child again using the state as a reference.

### **Analyses of data**

Ministers should expect to receive a full reporting of results and sophisticated analyses that would include item level (item response theory) data. For summary purposes, the reporting would be descriptive, indicating the percentage distributions falling within the four specified bands by state and territory and by particular categories of students (e.g. boys-girls or socioeconomic status levels).

Reporting should not be encapsulated in terms of a single number such as overall percentage of questions correctly answered. Figure 1 provides a solid basis for developing a results profile. This  $4 \times 3 \times 2$  category representation would hardly be stretched to provide results for the 24 separate cubes, but most of the nine marginals could be analysed for purposes of nationally comparable reporting. For example, each of the four content strands could be tallied to provide information on the hypothesis that students are better taught the concepts in the life and living strand than in the energy and change strand. Such an analysis could provide diagnostic information in terms of the content for professional development.

If it were decided to provide participating schools with reports (as recommended above) then those should be provided electronically and in the same sophisticated fashion as already occurs in many states already. Thus, schools could use the reports to see, for example, what areas are being relatively neglected in the school



curriculum, whether some students need remedial attention and whether other students need to be given special challenges.

### **3. Summary and conclusions**

The best single summary of this part of the report is graphically portrayed in Figure 1 (see above). Part 4 argues for three kinds of assessment: objective items, open-ended responses and practical tasks. Assessment of concepts and processes should take place in the second half of Year 6 and should occur, initially, every two years but a three-year regimen could be brought in subsequently. All four context strands would be assessed.

Sampling of schools rather than census assessment was preferred because it was seen to meet the objectives of national reporting without the extra cost that census assessment would incur. We would encourage strong consideration of item sampling. We believe that non-sampled schools should be able, at a later date, to have their students undergo the assessment.

These assessment procedures are consistent with the PISA framework.

Reporting to Ministers should not only provide a full data set but more importantly, should provide both normative and criterion-referenced information. A set of standards based on four bands (not competent, within range, achieved the desired level, beyond the desired level) should be developed and criterion-referenced reporting would indicate the percentage of students in each band in terms of the marginals agreed by Ministers (e.g. each of the four content strands, conceptual understanding, processing skills).

Reporting to schools which are part of the sample was also considered very desirable. Their reference group would be the state in which they are situated.

## **PART 5: OPTIONS – COSTS AND BENEFITS**

### **1. Introduction**

The various options presented in this part of the report arise out of the research described in earlier parts of the report.

With each option benefits and costs analyses will be presented. Benefits from one viewpoint may be seen as a cost from another viewpoint. Indeed, in most educational innovations there are few unmixed blessings and few unmitigated disasters.

This report will indicate in what way and from what perspective a particular outcome might be seen as a benefit and in what way and from what perspective it might be seen as a cost. This approach will be operationalised as the report turns to a consideration of the options.

We have separately reported (see Appendix E) the fiscal costs of the various options and the ‘bottom line’ of those costings will be referred to in the following text.

There are two assumptions underlying the options. These assumptions are based upon the discussions of the primary school science curriculum (Part 3 above) and of assessment and reporting (Part 4 above). From the earlier parts of this report the assumptions brought into this presentation of options are:

- That the assessments to be undertaken would occur with students who are in the second half of Year 6.
- That the framework underlying the assessment would be at least quite similar to that adopted for the PISA Science Literacy Project.

### **2. The options**

There are two extreme options that need to be dealt with at the outset. A strong preference for either of these would marginalise interest in the more moderate options. However, we cannot vouchsafe that each one of these first two options will not subsequently be chosen; and they are therefore seriously presented for consideration.

**Option 1      Either**

**1.1      There should be no formal specific monitoring of national education performance of primary school science.**

**Or**

**1.2      The TIMSS assessment (mainly Year 4) should be used as a means of monitoring primary school science.**

*(The only major difference between Option 1.2 and 1.1 is that TIMSS primary testing is offered as a data source to soften Option 1.1.)*

**1.1a Benefits**

1. The questions concerning standards and expectations (either implicit or explicit) with regard to the curriculum of primary school science will not need to be addressed at the national level. States and territories will continue to handle curriculum in their own way and possibly contentious debate among states and territories on the topic will be avoided.
2. This option decision would not likely be marked by industrial disputation concerning assessment in primary school science because the status quo will be maintained.
3. For those who believe accountability stops at the teacher, principal or state/territory level, and that there should be no national monitoring, the acceptance of Option 1 will be seen positively.
4. There will be no extra financial cost associated with implementing Option 1.1.
5. The disquiet expressed by some educators at the potential misuse of comparative descriptive statistics among states/territories will be avoided at least as far as primary science is concerned.

**1b Costs**

1. The first benefit presented for Option 1.1 (see above) is also interpretable as a cost. It is in the best interests of primary science education that there be at least a national discussion of standards and expectations. Thus, failure to develop a national monitoring of primary school science would mean that a discussion we ought to have would not be stimulated.
2. If Option 1.1 were accepted, a possible stimulus for professional development programs for teachers would not evolve. If assessment were to take place (see later Options), it is inevitable that at least some teachers in each state/territory

would have to receive professional development, if only to ensure adequate assessment and marking.

3. Systems and sectors within systems would not be provided the capacity to monitor the status and subsequent changes in their system or sector performance with respect to student achievement in primary school science.
4. The pluralism we currently enjoy in Australian education by virtue of having three sectors within each of the eight systems of education is enhanced, not damaged by a sensible national monitoring and reporting. Failure to implement a monitoring program will mean states/territories will not be given extra information about the effectiveness of their primary science programs.

**Option 1.2 (As for Option 1.1 except TIMSS would be used as the means of monitoring primary school science.)**

**1.2a Benefits**

1. As for 1.1a.
2. Ministers would have international comparative data as part of the nationally comparable monitoring information.

**1.2b Costs**

1. As for 1.1b.
2. The framework for TIMSS is clearly not consistent with that for PISA (which is recommended for national secondary science monitoring). (Robitaille *et al.* 1993)
3. Australia will not have control of the TIMSS assessment regime, which could mean that less-than-appropriate data are supplied for monitoring purposes. An assessment regime established in Australia by Australian states/territories is more likely to be geared to Australian monitoring needs than one developed internationally.
4. TIMSS uses age (not year level) as its primary criterion for entry into the sample frame. The 10-year-old age level involves students who are in Years 4 and 5 on some states and years 3 and 4 in other states. In any case this is deemed to be too early for national monitoring of the achievement of performance outcomes of primary science. In years 1-3 the emphasis is on literacy and numeracy; science becomes more important in years 4-6 than it had been in years 1-3.
5. TIMSS has had a ceiling effect so that it fails to differentiate among the top 15 per cent (approximately) of students at the age level of relevance here.

### 1c Recommendation

Ministers have agreed that student performance in key subject areas important to the achievement of the *National Goals for the Twenty-First Century* should be monitored in some fashion if this were practicable.

As later options will show such monitoring does seem to be practicable. Besides the costs in lost opportunities of adopting Option 1 seem to outweigh the benefits in the view of the writers of this report. Therefore, Option 1 does not appear to be worthy of further consideration. This is not a reflection on TIMSS, which will continue to provide useful information. Indeed, only through international assessments can we obtain valid international comparisons. It simply indicates that TIMSS is not seen as the best means for national monitoring of primary science.

The rest of the options concern the kind and extent of assessment and reporting within a monitoring system.

**Option 2 All students in the target category (Year 6)<sup>4</sup> in all states and territories should be comprehensively assessed (census assessment) using some type of paper-and-pencil or electronic test.**

(Note that this option would not allow for practical kinds of student assessment because to do so would involve such a huge resource allocation including nation-wide teacher-release for professional development and training that it could scarcely be seen to be relatively cost beneficial. We assume that there could be a need for release time for teachers involved in administering and marking practical assessments.)

**2a Benefits**

1. Census testing would ensure that all parents of Year 6 students would be provided with information about their students' achievement level.
2. The cost of test development is the same for census as for sample testing so the per student cost would be heavily discounted under Option 2.
3. All schools and all teachers would have the stimulus of having the performance of the students in primary science assessed (and reported to the school). In a sampling situation only a relatively few would receive this stimulus.
4. While the Option 2 assessment would not include practical-type common tasks, it could cover the major science strands including 'working scientifically'. Thus, questions concerning science curriculum area coverage could be answered not only for purposes of national reporting but as a positive also for individual schools.

**2b Costs**

1. Teachers might be drawn into thinking that primary science is mainly 'verbal' and lacking in practical activities. This would be because Option 2 would not include these more practical activities in the assessment regime.
2. Teacher organisations would consider Option 2 in a less accepting manner since most have a policy that indicates a preference for sample testing.
3. States may see Option 2 as a move that is inconsistent with their current policies concerning primary science assessment or competitive with their current practices. In either case, it would be unlikely that Option 2 would be acceptable or preferred

---

<sup>4</sup> See above Part 4 for a discussion of the year level at which assessment for monitoring should take place.

4. While ‘authentic assessment’ may not be an essential ingredient in a monitoring program, its absence as in Option 2 would incur the criticism of some curriculum experts.
5. With census testing special problems can arise with respect to exclusions policies and the difficulties of including all schools.
6. The overall cost of about \$3 million is much higher than most other options (see Appendix E).

## **2c Discussion and recommendation**

The costs of this option seem to outweigh the benefits. At least some of the possible benefits are at most secondary considerations in terms of the national monitoring remit. Even though the unit cost (per student) would be the lowest of all options considered (see Appendix E) the total cost would not be as competitive. We note too that this option would be the most unsettling in terms of teacher cooperation. For these reasons Option 2 is not considered to be a preferred option. It would not be a useful model for Ministers in terms of monitoring other subject areas. It would require major changes to overall policy and current priorities for Option 2 to be worthy of further consideration.

**Option 3      Each state and territory develop its own assessment regime in primary science and a system for equating the results to provide national monitoring be established.**

(This was an option put forward by one sector in one state and was advocated particularly in terms of the first of the benefits presented below.)

**3a Benefits**

1. If a particular state sees itself as having a markedly different primary science curriculum, it may feel more comfortable designing its own assessment.
2. Pluralism in curriculum and assessment among Australian education authorities is further enhanced.
3. The content validity of each separate state assessment would conceivably be greater for that state/territory than a national assessment.
4. If the curriculum rather than student performance were the prime concern this option would be more desirable.

**3b Costs**

1. The likelihood of developing a system of national monitoring in other areas such as performance in civics or in information technology could be hindered if it were decided to leave the assessment of primary science to each state/territory.
2. Ministers would have considerable difficulty interpreting the results from potentially eight different assessment regimes. Any statements used on such monitoring would be open to considerable debate.
3. The direct assessment costs would be borne by the eight states/territories separately. However, the overall cost of up to eight separate test developments would be demonstrably greater than one national effort.
4. The overall cost estimated at about \$4 million is the largest of all the options considered.

**3c Discussion and recommendations**

The argument that a given state has a markedly different primary science curriculum from other states is based upon an analysis that magnifies specific differences and minimises obvious similarities. It is also based upon the expectation that the stated curriculum and the operationalised classroom curriculum virtually are identical.

As was noted in Part 3 above, it is a reasonable contention that there is much in common that characterises what is happening in the various states/territories. Of



course, there are differences too and a reliable and valid monitoring system should be able to tease out where specific differences in student achievement are occurring.

In any case, the national reporting program is not *primarily* concerned with the summative evaluation of each state's or territory's individual curriculum in primary science but rather in monitoring student achievements. While assessment procedures might overlap (evaluation vs monitoring) the purposes are not identical.

The idea of each state/territory developing its own assessment for national monitoring purposes was not generally approved among those interviewed. However, it was emphasised that, whatever the assessment regime that was recommended to Ministers, it should take full account of current state/territory assessment programs. This is a point the writers of this report would readily agree with.

It is only when a common comprehensive assessment regime is instituted that each state/territory can see not only where its strengths are but also what gaps exist. Gaps in skills and concepts cannot be found if the assessment is severely tailored to the individual curriculum.

If this recommendation against Option 3 is difficult for some states/territories, a possible amelioration could be provided by encouraging states/territories to indicate, as has happened in the International Association for the Evaluation of Educational Achievement (IEA) studies, what assessment items or tasks are less appropriate for that state/territory in terms of its current curriculum.

If students do poorly in those items/tasks deemed less appropriate, the reason is then obvious as is the remedy if the state/territory decides to change the situation. If students do well despite expectations it will provide evidence that the stated curriculum and the experienced curriculum are not necessarily the same. That in itself can be important information for curriculum planners.

- Option 4**                    **Ministers agree to develop an item bank of primary-level science assessment tasks each with its difficulty level and concept/skill focus clearly delineated. Assessment would then involve allowing schools, sectors within states or state/territories to choose from within the item bank. Reporting would be mainly in terms of overall difficulty level achieved.**

*(This is an option initially put forward by a measurement expert and it is in harmony with the New Zealand assessment resource – the Assessment Resource Bank (ARB) developed by the New Zealand Council for Educational Research (NZCER)).*

#### **4a Benefits**

1. This approach is potentially less anxiety inducing on teachers (if choices are school-based) or on sectors (if choices are sector-based).
2. Such an item bank has great potential as a resource for primary school since it would operationalise the objectives of the science key learning area and it would enable teachers to develop their own class tests to a far more professional level than if left unassisted.
3. An item-bank-based assessment regime would enable states/territories and sectors to ensure items used in the assessments were curriculum valid.
4. The approach under Option 4 would enable a more reliable equating in terms of difficulty level of state/territory and sector results.
5. The basis of an item bank currently exists electronically in such programs as VSAM, Science Online, ACER resources (e.g. TIMSS) and the NZCER.
6. The financial outlay (see Appendix F) would be relatively low compared with most other options, but maintenance would be higher.

#### **4b Costs**

1. The cost of developing a sophisticated comprehensive primary science item bank with psychometrically defensible properties would be substantial in the first instance and would require maintenance (see Appendix F).
2. The time to develop an item bank with appropriate properties would put off a national assessment of primary science by at least a year and perhaps more. The experience of NZCER and other organisations which have attempted this kind of enterprise is that it takes much longer than developing a more conventional assessment approach. While it is true that embryo item banks now exist, the time would still be considerable.
3. As a monitoring device it has the unfortunate property (shared with Option 3) that students in a particular state/territory will most likely have a quite different set of assessment items than in another. While this is not necessarily a problem in

evaluating the success of individual state/territory primary science programs, it is a potential problem in monitoring primary science from a national perspective.

#### **4c Recommendations**

In the medium term the development of a primary science item bank, with the assessment tasks tagged by strand and difficulty level, is an attractive proposition. The attractiveness stems from its potential to upgrade primary science teaching in a number of ways (for example, professional development of teachers, improved classroom assessment). However, these tend to be positive side effects rather than the main target of a monitoring program. While Option 4 would benefit teachers of science across the middle and later years of primary school, it would likely delay and perhaps obfuscate the targeted purpose of monitoring student performance in primary science.

The recommendation by the project team is that Option 4 *not* be considered the preferred option; however, as a gratuitous (in the better sense) recommendation we would consider it worthwhile if Ministers were to approve the development of an item bank project either *de novo* or based on that developed by NZCER or being developed by VSAM (Victorian Student Achievement Monitor). This item bank, along with such other programs as professional development and training, would help to enhance the positive package context into which national monitoring would be placed.

**Option 5      Either**

**5.1      An optically scanned test would be developed and administered to all targeted students in a scientific sample of schools in each state/territory supplemented by other practical assessments for all those students. The practical assessments would be mediated and marked by the classroom teacher or, depending on the nature of the task, also marked optically**

**Or**

**5.2      As for 5.1 except that the supplementation of teacher-mediated and marked practical assessments would be administered only to a scientific sample of the targeted students in the sampled schools.**

(The only difference between Option 5.1 and 5.2 is a reduction in the total assessment time. Practical assessments and teacher marking are time-consuming. The potential difference in costs is indicated in Appendix E. The overall impact between the alternatives is otherwise inconsequential. The costs and benefits (apart from time costs and related release time consequences) are identical. Therefore the following analysis will refer to Option 5.1 and 5.2 as an entity.)

**5a Benefits**

1. Option 5 can provide Ministers with comprehensive information about student performance in primary school science based upon objective and open-ended items, practical tasks and teacher judgments. It inherently ensures both reliability and validity in the data provided.
2. The fiscal cost of both 5.1 and 5.2 are presented in Appendix E and it can be seen that this option, in either form, is less expensive than Option 2 and Option 3 and is, overall, not markedly more expensive than Option 4.
3. The concept of sample testing and the use of teacher judgments is acceptable, in principle, to major teacher industrial groups so that Option 5 is unlikely to create industrial relations opposition if carefully implemented.
4. Sample testing as proposed in Option 5 allows Ministers to obtain monitoring information about student performance with only the most minimal problems of sampling error in terms of state/territory and sector student performance. (Of course, this approach does not allow individual student performance results to be given to all parents. However, the parent is not the major focus for the results of the national monitoring exercise.)
5. This option ensures that the assessment will tend to mirror the curriculum and the pedagogy associated with primary school science because it provides a range of types of assessment. It builds on the experience of such an assessment approach in Western Australia. It provides a form of teacher professional development.

6. Option 5 carries with it the possibility not only of sampling schools (and sampling target students within schools – Option 5.2) but also of using the technique of item sampling. Thus not all students need take all items as long as the various test forms have degrees of mutual item overlap.
7. Option 5 could also, at the discretion of Ministers, monitor every second or third calendar year. A yearly assessment is not a *sine qua non*. Indeed, every two or three years assessments may provide sufficient monitoring especially after the dynamics of monitoring, research and program change are experienced. Thus the recurrent fiscal cost could be lowered and other subject areas dovetailed into the overall national monitoring program in the off years.
8. Option 5 would allow sampled schools not only to cooperate in the assessment process but also to obtain on a confidential basis information about its students' performance. The opportunity to have third-party information about strengths and weaknesses in its curriculum program could be very beneficial to the school.
9. Option 5 was overwhelmingly endorsed by those interviewed in this report.

#### **5b Costs**

1. The fiscal cost (see Appendix E) is not inconsequential. (However, also refer to 5a.2 above.)
2. There would be a necessity to provide teachers of sampled classes with up to a day of release time for professional development including explaining the underlying purposes of the assessment, how to administer the tests and how to mark open-ended and practical assessments. There would also be the need for release time to assess all (Option 5.1) or some (Option 5.2) of their students at individual or group tasks. Thus the practical and open-ended augmentations in Option 5 would increase costs. Nonetheless, they are deemed to be highly desirable augmentations.
3. While the judicious use of sampling provides sufficient information for generalisations that meet Ministers' purposes, the procedure also creates schools that benefit from the attendant teacher professional development and schools that do not have this benefit. This problem could be overcome by allowing non-sampled schools to use the assessment after the sampled schools have been assessed, and also to obtain school level information. Ultimately, it involves an increase in the cost of educational assessment because some extra marking and reporting would be involved.
4. It is possible that some marginal groups about which Ministers wish to have monitoring information might only be reliably reported on by using over-sampling. This could incur further costs.

### 5c Recommendations

The preferred position of the writers of this report is that NEPMT recommend to Ministers either Option 5.1 or 5.2 with a preference for 5.1 over 5.2. The preference is dependent upon the supplementary practical assessments being capable of being administered by a specifically trained teacher to a class in a day. It is assumed the teacher would be given release time of up to one day for this purpose.

If the practical assessments cannot be so configured, then the cost of assessing all children in a class using this procedure would outweigh the benefit. For example, a 50 per cent scientific sample from each sampled class in each sampled school would provide little sampling error to the generalisations sought for purposes of nationally comparable reporting.

The overall benefit of Option 5.1/5.2 is that it achieves the objectives for monitoring in a reasonably cost-effective manner at the same time as it provides professional development in science directly to a sizeable group of primary school teachers. It would also provide a 'ripple effect' to other teachers who could observe the kinds of questions that operationalise primary school science.

## POSTSCRIPT

Obtaining closure on a topic is a luxury rarely experienced in education. This report on primary science assessment and reporting is no exception. Two clear examples of further work needed to further the monitoring process in primary science are briefly presented in this Appendix:

- Commonalties in the curriculum
- Developing standards

Both topics are separately referred to in the text of the report and both require further development if monitoring as recommended is to proceed.

### 1. Commonalties in the curriculum

In Part 3 of the text of this report it was argued that there are commonalties across the primary science curriculums of the states/territories. These emanate from the national statements and profiles documents and they include:

- a firm consensus on the importance of process (e.g. developing fair tests, describing phenomena accurately, drawing valid conclusions);
- a general agreement on the strand structure including four strands of life and living, earth and beyond, natural and processed materials, and energy and change;
- an operationally manifest but less explicit agreement that certain scientific concepts and principles are the legitimate province of primary science.

If monitoring is to proceed in primary science there needs to be some more explicit determination of what these concepts and principles include. As indicated in Part 3 the Curriculum Corporation has made initial attempt to be explicit in this area. This is important if their Science Online Project is to be useful at the national level.

There is expected to be a mapping of the primary science curriculum across Australia available from the Curriculum Corporation in August 2000. Thus, there is no good reason why a start could not be made to the development of the first monitoring assessment by, say, September 2000 if curriculum considerations were the only problem.

In current statewide testing programs it is normal to have panels that verify the relationship between curriculum and assessment items. This would also be part of the normal development of a primary science monitoring assessment that could be ready to be administered by the second half of 2001 and certainly by 2002.

Similarly, the Australian Council for Educational Research (ACER) has carried out its own investigations in order to develop science tests that are relevant across states and territories.

ACER has developed such tests and they have been administered across Australia with no manifest problems related to curriculum content.

This must not be construed as an attempt to destroy the opportunities of states to exercise their individual creative curriculum talents in primary science. However, it must be seen as appropriate to have some reasonable expectations of what our students should learn to understand during their primary school years.

## **2. Developing standards**

In Part 4 (section 3) of the report, the topic of criterion referencing was taken up in the context of the reporting of results. If four Band levels are to be used to specify student performance then the question of what is competent performance and the various levels surrounding competence (not yet competent, about to achieve it, well beyond it) need to be operationally defined.

This iterative process is to be achieved over time. At first it would be necessary to have a taskforce of competent judges consider relevant curriculum documents and, where available, data from primary science testing. Such a group could develop a sound initial operational definition of the bands based on the first assessment development exercise in this monitoring exercise.

The competent judges would then be informed by the actual data generated and would be in a position then to 'fine-tune' the standards they had originally set. By the end of the first two assessment exercises stability would be expected. This would not denigrate the first two assessments. For monitoring purposes standardised adjustments might be made retrospectively.

The current lack of a clear set of criteria or standards should not be used as an excuse not to proceed with monitoring. Rather like the egg and chicken paradox it is less relevant to ask what should come first than it is to assert that they are dependent on each other.

Whether the first monitoring assessment is administered in the second half of 2001 or, more comfortably, in 2002, the standards could be drafted at the time the tests are developed and fine-tuned and published as part of the reporting process.



## References

- ABS 4221 (13 April 2000), *Schools of Australia*, Canberra, ACT.
- Adams, R.J., Doig, B.A. and Rosier, M. 1991, *Science Learning in Victorian Schools: 1990*, ACER, Melbourne.
- American Association for the Advancement of Science 1993, *Benchmarks for Science Literacy*, Oxford University Press, New York and Oxford.
- ASTEC (May 1997), *Foundations for Australia's Future: Science and Technology in Primary Schools*. Australian Government Printing Office, Canberra.
- Australian Council for Education Research 1999, *Raising Australian Standards in Mathematics and Science: Insights from TIMMS*, ACER, Melbourne.
- Branscomb, A.W. 1987, 'Knowing How to Know', *Science, Technology and Human Values*, 6(36), 5-9.
- Council of Minister of Education, Canada 1997, *Common Framework of Science Learning Outcomes, K to 12*, Council of Ministers of Education, Canada.
- Curriculum Corporation 1994, *A Statement on Science for Australian Schools and Science: A Curriculum Profile for Australian Schools*, Curriculum Corporation, Melbourne.
- Gagne, R.M. 1965, *Conditions of Learning*, Holt, New York.
- Harlen, W. 1991, 'Performance Testing and Science Education in England and Wales' in G. Kulm and S.M. Malcolm (eds), *Science Assessment in the Service of Reform*, pp. 163-188, American Association for the Advancement of Science, Washington.
- Helgeaon, S., Blosser, P.E. and Howe, R.W. 1977, *Science Education*, Ohio State University Centre for Science and Mathematics Education, Columbus, Ohio.
- Húsen, T. and Postlethwaite, T.N. (eds) 1994, *The International Encyclopaedia of Education*, 2nd edn, Pergamon, Oxford.
- Jenkins, E.W. 1994, 'Scientific Literacy' in *The International Encyclopaedia of Education* (edited by T. Húsen and T.N. Postlethwaite), 2nd edn, pp. 5345-5350, Pergamon, U.K.

- Jones, L.R., Mullis, I.V.S., Raizen, S.A., Weiss, I.R. and Weston, E.A. 1992, *The 1990 Science Report Card: NAEP's Assessment of Fourth, Eighth and Twelfth Graders*, US Department of Education, Washington, D.C.
- Le Guen, M. 1991, *Measuring to Succeed: National Evaluation in France*, OECD, Paris.
- Lucas, A. 1987, 'Public Knowledge of Biology', *Journal of Biology Education*, 21, 1, 41-45.
- Nuttall, D.L. 1994, 'Monitoring National Standards: United Kingdom' in *The International Encyclopaedia of Education* (edited by T. Húsen and T.N. Postlethwaite), 2nd edn, pp. 3906-2912, Pergamon, U.K.
- OECD 1999, *Measuring Student Knowledge and Skills: A New Framework for Assessment*, OECD, Paris.
- Orpwood, G., Block, M., Wright, B., Bartley, A., Herridge, D. and Marks, M. *Classroom Assessment in Science and Technology: A CD-ROM Resource Handbook for Teachers*, York University Faculty of Education.
- Orpwood, G. and Garden, R.A. 1998, *Assessing Mathematics and Science Literacy*, Pacific Educational Press, Vancouver.
- O'Sullivan, C.Y. and Weiss, A.R. 1999, *Student Work and Teacher Practices in Science. A Report on What Students Know and Can Do*, U.S. Department of Education, Washington, D.C. (also on the website <http://nces.ed.gov/naep>).
- Robitaille, D.F., McKnight, C., Schmidt, W.H., Britton, E., Raizen, S. and Nicol, C. 1993, *Curriculum Frameworks for Mathematics and Science*, Pacific Educational Press, Vancouver.
- Rosier, M. 1996, 'Science Achievement in Post-Compulsory Schooling in Australia from an International Perspective' in P.J. Fensham (ed.), *Science and Technology Education in the Post-Compulsory Years*, pp. 59-76, Australian Council for Educational Research, Melbourne.
- The State of Queensland (Department of Education) 1999, *Statements of Content for Education Queensland Schools*, Brisbane.
- Thomas, G. and Durant, J. 1987, 'Why Should We Promote the Public Understanding of Science?' in Shortland, M. (ed.) *1987 Science Literacy Papers*, University of Oxford, Oxford.

Western Australia Department of Education 1999, *Issues in System Level Measurement of Student Performance: Lessons from Monitoring Standards in Education in WA* (unpublished document).

## APPENDIX A: USES OF MONITORING

Monitoring uses some of the tools of research and evaluation (for example, measurement) but its purpose is to provide information about how the education system *as a whole* is functioning (Nuttall 1994). The purpose of monitoring in education is usually to chart changes in the level of achievement over time. As Le Guen (1991) points out, in some countries national monitoring informs the accountability of the national system of education.

There is, of course, the metaphorical use of the term ‘monitoring’. Thermometers monitor the temperature in various parts of a building to determine where the air conditioning system should direct cooling air flows. A patient in intensive care is monitored in terms of blood pressure, body temperature and blood chemistry. Some of the monitoring is continuing (pulse) and some may be discontinuous but regular (the daily use of a sphygmomanometer to measure blood pressure). Monitoring indicates whether a system is in a state of homeostasis or is changing in a particular direction.

Economic indicators used for monitoring purposes, such as the inflation rate, GDP or balance of payments, exemplify the principle that monitoring is not necessarily, of itself, diagnostic. However, by indicating the status and the kinds of change that are occurring in a system, monitoring provides the basis for subsequent, in-depth diagnosis and evaluation.

National monitoring of performance in primary school science will not necessarily provide in-depth reporting on specific problems in the provision of primary school science. However, it may well point to problems in various aspects of performance and it should indicate whether the problems are being ameliorated or are worsening. Monitoring can therefore point to the kinds of planning, research and evaluation that should follow on from the monitoring process.

Monitoring can also have useful side effects. Gipps (1986), for example, sees an important contribution of the Assessment of Performance Unit (APU) in England and Wales the generation of high quality assessment instruments and the provision of a detailed listing of common misconceptions of students (for example, correlation equals causation, or, the sun circles the earth) that can be used to improve teaching.

Nuttall (1994) has reviewed monitoring across OECD countries and points out that the most member countries do have some form of national monitoring of their education programs. Some specific examples are presented in some detail in Part 2 of this report. Nuttall’s overall summary suggests the following generalisations:

1. Once begun, monitoring systems continue to be developed or at least changed to meet changing needs of the monitoring agencies. (Strangely, Nuttall argues from an article by Wood and Power (1984) that Australia is the only exception where monitoring systems were begun and then discontinued. Obviously, this was a premature conclusion.)

2. Despite major educational, cultural and historical differences among the countries analysed, there were striking similarities with respect to their monitoring system which typically involve a focus on a few major subjects (reading, mathematics, science) and which are tested regularly, but not necessarily yearly, using a range of assessment methods.
3. Almost all monitoring involves sampling and this sampling is not only sampling of students or schools, but includes item sampling to ensure coverage without oppressive amounts of testing.

## **APPENDIX B: THE INTERVIEW SCHEDULE**

The areas covered in the semi-structured interviews:

### **A. Curriculum**

1. What is the structure of primary science in your state?
2. Is there (explicit or implicit) essential knowledge/facts/concepts?
3. Are there (explicit or implicit) essential skills (for example, sorting, classifying, problem-solving, experimental methods)?
4. May we have access to your curriculum documents?
5. Do you classify Item 2 and 3 above by Year level? Age level?

### **B. Assessment/reporting**

1. Who do you see as the audiences for assessment / reporting of primary science across Australia for NEPMT?
2. Could such assessment be incorporated into current state testing programs? Should it be?
3. Within the context of primary science assessment would you favour sample or census testing?
4. What is the current situation in your state re primary school science assessment and reporting?
5. Should/could primary science assessment be incorporated into literacy testing?
6. Would teachers in your state help out in testing of practical skills?

### **C. General questions**

1. Can you provide your views on PISA science structure?
2. Do you have any advice or thoughts on this topic?
3. Can you suggest other bodies/organisations we should talk to?

## APPENDIX C: THOSE INTERVIEWED

The following is an alphabetical listing of those who were interviewed as part of this project. Most of the interviews were at the individual level, a few were small group interviews and two were by telephone.

- **Damien Brennan** Assistant Director, Catholic Schools Religious Education and Curriculum, Queensland
- **Terri Burnet** Senior Education Officer – Mathematics, Science & Technology, Education Services Directorate, Queensland
- **Neil Champion** Manager – Science Key Learning Area, Board of Studies, Victoria
- **Terry Chapman** Executive Director, Association of Independent Schools, New South Wales
- **Di Charles** Project Officer Learning Areas – Maths/Science Research & Development, DEET, South Australia
- **Jocelyn Cook** Senior Education Measurement Officer, Education Department of Western Australia
- **Mary Colvill** President, Australian Science Teachers Association, Norwood Primary School, ACT
- **Susan Dennett** Acting General Manager, Curriculum Initiatives Branch, DEET Victoria
- **Fred Deshon** Senior Curriculum Officer – Science, Education Department, Western Australia
- **Peter Fensham** Science Education, Monash University, Victoria
- **Shaun Fitzpatrick** Teaching & Learning Consultant, Curriculum & Education Services, Catholic Education, South Australia
- **Lyndall Foster** Chief Education Officer – Technology, Department of Education & Training, New South Wales
- **Denis Goodrum** Associate Professor, Edith Cowan University, Western Australia
- **Rosemary Hafner** Inspector – Science, Board of Studies, New South Wales
- **Audrey Jackson** Executive Director, Association of Independent Schools, Western Australia
- **Sue Kidd** Education Officer – Science, Catholic Education Office, Victoria
- **Barbara Kroll** Education Officer – Science, Archdiocese of Brisbane, Catholic Education, Queensland
- **Philip Lambert** Inspector – Primary Education, Office of the Board of Studies, New South Wales
- **Jan Lokan** Deputy Associate Director – Measurement, ACER, Victoria
- **Lorrie Maher** Executive Officer – Education Services, Association of Independent Schools of Queensland
- **Susan Mann** Director – Curriculum Program, Curriculum Corporation, Victoria
- **Tony McArthur** Professional Assistant to the Executive Director, Catholic Education Commission, New South Wales

- **John McArthur** Secretary, Ministerial Council on Education, Training and Youth Affairs, Victoria
- **Jim McMorrow** Department of Education and Training, New South Wales
- **Bob Nield** Manager, Curriculum Initiatives Section, Education & Community Services, ACT
- **Catherine Nikkerud** District Coordinator Schools, Department of Education Training and Employment, South Australia
- **Chris Payne** Senior Education Officer – Research & Evaluation, Department of Education & Training, New South Wales
- **Andrew Perry** Education Services Officer, Association of Independent Schools of Victoria
- **David Robertson** Executive Director – Operations, Association of Independent Schools of Victoria
- **Christine Rodgers** Science Consultant – Secondary Curriculum & Teaching Team, Catholic Education Office, Western Australia
- **Peter Russo** Curriculum Officer, Learning Area – Science, Department of Education Training and Employment, South Australia
- **Jim Scott** Chief Education Officer – Science, Curriculum Support Directorate, Department of Education & Training, New South Wales
- **Pauline Sharma** Education Officer – Science, Catholic Education Office, Victoria
- **Mark Snartt** Principal Project Officer – Curriculum Development, Queensland School Curriculum Council, Queensland
- **Julie Thompson** Director – Professional Development, Association of Independent Schools, New South Wales
- **Fergus Thomson** Executive Director, National Council of Independent Schools' Associations, ACT
- **Bevis Yaxley** Principal Education Officer, Tasmanian Education Department, Hobart, Tasmania
- **Yvonne Zeegers** Lecturer – Primary Science & Technology Education, Faculty of Education, University of South Australia



## **APPENDIX D: INTERNATIONAL DETAILS**

- D1: Assessment resource bank sample – New Zealand**
- D2: Extended activity assessment options – Ontario**
- D3: Science investigation assessment – Michigan**
- D4: Operational items and marking schemes – England**
- D5: Multiple-choice and free-response items – TIMMS**

This Appendix is included as separate Acrobat PDF files. Double click on the appropriate file to open it:

Appendix D1: File APPD1

Appendix D2: File APPD2

Appendix D3: File APPD3

Appendix D4: Files APPD4a, 4b , 4c and 4d

Appendix D5: File APPD5

**(Please note these materials are no longer available).**

## **APPENDIX E: THE FISCAL COSTINGS**

Please note that this appendix has been removed as it contains confidential costing information for the MCEETYA National Education Performance Monitoring Taskforce.

## APPENDIX F: SAMPLE ITEMS

**Assessment Task:** *Practical Task*

**Syllabus:** *Science and Technology K–6 (New South Wales)*

**Level:** *Stage Three (Years 5–6)*

**Outcome assessed:**

*Students will describe the process of investigation which can involve **exploring** and **discovering phenomena** and events, **proposing explanations**, **initiating investigations**, **predicting outcomes**, **testing**, **modifying** and **applying understanding**.*

(This task will address the areas printed in bold)

The tasks all assess the process of ‘finding out’ in an accurate, objective manner some new information. Consequently the marks are awarded for the processes highlighted in the boxes given in the first task.

It should also be noted that the tasks of finding out ‘how much/ how fast/ how hot’ etc. are very simple to set up and for the students to understand and recognise the endpoint of their task.

Investigations to answer a ‘How’ question (e.g. How does a spider breathe?) are more complex while ‘Why’ questions arguably cannot be answered since the final answer may well be ‘Well that is how our Universe works’.

However, observation of students carrying out these ‘How much’ style tasks still gives us an indication of their skills in designing and carrying out scientific tasks.

### Task 1

Description: *The students will be shown a number of common balls involved in common sports or games. They will be asked to respond to two questions:*

- Which ball bounces best?
- How can you find out which ball bounces best?

The former question is of no real essence except to challenge the students to take a stance, make a prediction for them to test at a later stage.

The second question sets the tone and purpose of the investigation and may be used to collect individual responses to determine student skills in drafting a first plan for an investigation. If used as such, the student responses should be collected at this stage.

For the practical component, the students work in groups so the teacher needs to be much more attentive to group and individual performances.

Stage	Description	Awards
Planning	<p>Students:</p> <ul style="list-style-type: none"> <li>• agree on a method which describes the ‘best bounce’ in terms of height after one bounce or number of bounces before coming to rest</li> <li>• devise a method to compare the balls fairly ie dropped, from same height onto same surface with ball at normal playing inflation</li> </ul>	<p>Mark for:</p> <ul style="list-style-type: none"> <li>• agreeing on appropriate definition of ‘best bounce’ (1 mark)</li> <li>• recognising need for fair test (1 mark)</li> <li>• instituting a fair test by eliminating appropriate variables as much as possible (2 marks max.)</li> </ul>
Calibration	<p>Students:</p> <ul style="list-style-type: none"> <li>• devise a sensible method to control the height from which ball is dropped, noting that a ‘standard’ height (e.g. 1 metre) is superior to a subjective one (e.g. waist height) as it allows for more meaningful comparison.</li> </ul>	<p>Mark for:</p> <ul style="list-style-type: none"> <li>• Agreeing on a set height in metres (2 marks) or body comparison (1 mark)</li> </ul>
Measuring	<p>Students:</p> <ul style="list-style-type: none"> <li>• determine an appropriate method of measuring and recording the end point, i.e. height after one bounce or number of bounces</li> <li>• apply the method rigorously</li> </ul>	<p>Marks:</p> <ul style="list-style-type: none"> <li>• As for calibration</li> <li>• For consistent implementation</li> </ul>
Recording	<p>Students:</p> <ul style="list-style-type: none"> <li>• record their measurements in a neat orderly manner</li> </ul>	<p>Marks:</p> <ul style="list-style-type: none"> <li>• correctly matched recording of results</li> </ul>

Analysis	Students: <ul style="list-style-type: none"> <li>• compare the results for the various balls in terms of their stated criteria</li> <li>• correctly match their results with their plan ie match their description of 'best bounce' with experimental results</li> </ul>	Marks: <ul style="list-style-type: none"> <li>• for correctly interpreting results</li> </ul>
Conclusion	Students: <ul style="list-style-type: none"> <li>• correctly identify the ball which bounces best</li> </ul>	Marks for: <ul style="list-style-type: none"> <li>• applying the criteria rigorously in naming the 'best bouncing' ball even where this went against their original prediction</li> </ul>

**Follow-Up Activity:** Students can be asked to respond, in written form or by designing and carrying out further experiments, to questions such as:

- Does it matter what the floor is made of?
- Can you rank the balls in order of best to worst bounce?
- How would you work out *How much better* one ball bounces compared to another?
- Would the same ball bounce best on all different surfaces?

## Task 2

The students are asked to compare two shoes to see which on gives better grip on the floor. The equipment available to them includes ruler, spring balances, graph paper, elastic bands and standard laboratory equipment.

The students must **plan** their activity and agree on an objective method of comparing shoes. Just trying to scuff the floor while wearing the shoes is unsatisfactory. A simple method is to drag the shoes along the floor while attached to a spring balance and taking readings of the force required to just get the shoe moving. Alternatively, they might use elastic bands to pull the shoes and measure the extension of the bands just before the shoe moves.

They may make a **prediction** of which shoe they think will give better grip.

They carry out the task, using appropriate methods to **measure** and **record** the required force.

They use their results to **determine** which shoe required the greater force to start moving.

They **compare** their prediction with their results to see whether they predicted correctly.

In their **conclusion** summarise their findings.

### Task 3

The students are asked to determine whether or not it is true that your body length is equal to your hand span.

The students may need to **discuss** the term handspan and ways of measuring body length and hand span.

Working in pairs they **measure** each person's required dimensions and **record** their results. This may be done **numerically** or **physically**, by giving students two strips of paper each, one equal to the body length the other equal to the hand span.

Strips of paper can be **hung** on the wall for **comparison**. Numerical results can be **drawn** as a **scattergraph**.

Follow up discussion would include:

- How close do the two measurements have to be to say they are 'equal'?
- How many people in the group had their two measurements equal?
- Was the original statement proven true or false?
- What were some things that made the measurements inaccurate?
- How could you make the measurements more accurate?

## Open-ended paper and pencil tasks

Some of these items address assess knowledge and will therefore have a different structure. Secondly, some of these have been taken from Year 10 tasks and are included as illustrations of ideas. They will clearly be too difficult for Year 6.)

These items will generally require some stimulus taking the form of a question to be answered, problem to be solved or viewpoint to support.

Students are then rewarded for the quality of the answer. Note that the ‘quality’ of the answer is determined by the nature of the task set.

### Example: **Supporting a Viewpoint**

*‘Two students are talking about soft drinks. They argue about which soft drink is ‘best’.*

*They decide to settle their argument by asking their friends which soft drink they think is best.’*

1. What evidence do the boys intend to use to decide which soft drink is ‘best’?
2. Does this really prove which soft drink is best?
3. How would you improve their experiment?

Marks would be awarded for the students recognising that the boys will use the opinion of their peers to settle the argument. They would score extra marks for indicating that this does not prove which soft drink is best, since there is no criterion given for a soft drink being ‘best’. What they are measuring is opinion.

The stimulus can then be extending the experiment by doing a taste test, then a blindfolded taste test etc.

The students could then be asked to come up with suggestions for the sort of things would make one soft drink ‘better’ than another? Here they would score marks for coming up with valid criteria.

Note that a task could be easily designed where a product is tested on criteria that can be easily given numerical values of some sort, leading to an objective comparison. Examples include, comparing running shoes for the grip they give, dishcloths for the amount of water they soak up etc.

### Solving a problem

Students X and Y are talking about how ‘thick’ liquids are. They agree that some liquids are ‘thicker’ than others, but cannot decide how they can compare the ‘thickness’ of two liquids.

1. What words would you use to describe the ‘thickness’ of a liquid?
2. Explain how you would go about finding which of two liquids was ‘thicker’ than the other. Make sure you describe the equipment you would use, the steps you would carry out and the results you would expect.

(This task could then be extended to let the students actually carry out their test on, say, honey and water. Further they could research a number of properties of liquids and look for a link with their ‘thickness’.)

Marks would be awarded in:

1. For the clarity of description, noting that the definition here of the ‘thickness’ may be quite subjective. Some may refer to being runny or not where others may opt for a notion of density ie, it is heavy even when you don’t have much of it. Marks will be awarded for consistency ie if they decide that thick means ‘not runny’ then they should give honey as an example of thick and water as an example of thin.
2. For coming up with a technique to compare the thickness of two liquids that is consistent with their definition in 1.

### Interpreting information

The table below shows the properties of a number of alcohols.

Name of alcohol	Formula	Molecular Mass	Melting temp. (C)	Boiling temp. (C)	Energy released when burnt (kJ/g)
Methanol	CH <sub>4</sub> O	32	-98	65	23
Ethanol	C <sub>2</sub> H <sub>6</sub> O	46	-114	78	30
Propanol	C <sub>3</sub> H <sub>8</sub> O	60	-126	97	36
Butanol	C <sub>4</sub> H <sub>10</sub> O	74	-89	117	41
Pentanol	C <sub>5</sub> H <sub>12</sub> O	88	-78	138	?
Hexanol	C <sub>6</sub> H <sub>14</sub> O	102	-45	157	?



1. What pattern do you see in the formulae of the alcohols?
2. Describe how the molecular mass of the alcohols changes as you go down the table. Use the section of the Periodic Table given to explain the changes.
3. Describe the trend linking the molecular mass and the amount of energy released per gram when burnt.
4. Predict the energy released per gram when pentanol and hexanol are burned.
5. What products will be formed when alcohols burn with oxygen in the air?

### Analysing information

Louise and Jordin have made a force meter using elastic bands. As the force on the elastic bands is increased, they measure the new length of the bands. From this they can work out how much the bands have stretched. Their results are given below:

Force on bands (grams)	Length of bands (cm)	Total stretch of bands (cm)
0	22	0
50	26	
100	31	
150	34	
200	38	
250	41	
300	47	

1. In the table on your answer sheet, fill in all the missing values from the 'Total stretch of bands' column.
2. Use the graph paper provided to draw a line graph of the 'Total Stretch' values and the 'Force on bands' values. Make sure you draw in the line of best fit.
3. Jordin now hangs her pencil case from the end of the elastic bands and finds it stretches the elastic bands by 14 cm. Use the graph to find out how heavy the pencil case is.
4. Later, another student reports that her group used the forcemeter to hold up a retort stand. This made the elastic bands 59 cm long. How much stretch does this equal? What mass would the retort stand be according to this result?
5. On your graph draw a line to show how the graph might have turned out if they had used elastic bands which were exactly twice as stiff as the ones that

were used in the experiment. (In other words if they were twice as hard to stretch)

### Ernie's Eye Diagram

Ernie labelled a diagram of the eye. He made just **three** mistakes in **naming** the parts. (This task assesses student knowledge of the structure of the eye and spelling. Furthermore it assesses student understanding by asking them to reword their ideas to suit different audiences.)

The students are provided with a labelled diagram of the eye, supposedly labelled

1. Ernie labelled a diagram of the eye. He made just **three** mistakes in **naming** the parts.

*Put a cross next to his three mistakes and rename them correctly.*

2. Give two ways to help Ernie remember the names of the parts of the eye.

---

---

3. Ernie also made **three** spelling mistakes. *Correct his mistakes by writing the corrections next to the mistakes.*

4. The diagrams show what happens to rays (  $\longrightarrow$  ) of light as they pass through different lenses.

*Describe the pattern of the rays of light as if:*

*a. You were talking to your cousin who is in Year 4 in Primary School.*

---

---

---

*b. You were talking to your English teacher*

---

---

**Assessment Task:** *Practical Task*

**Syllabus:** *Science and Technology K–6 (New South Wales)*

**Level:** *Stage Three (Years 5–6)*

**Outcome assessed:**

*Students will describe the process of investigation which can involve **exploring** and **discovering phenomena** and events, **proposing explanations**, **initiating investigations**, **predicting outcomes**, **testing**, **modifying** and **applying understanding**.*

*(This task will address the areas printed in bold)*

### **Description**

The students will be given a set of tasks that they will complete individually. They take the form of ‘Station Tasks’. At each station, students perform a simple task to show their competence at a skill they have been taught recently.

The marks awarded for different stations will vary depending on the nature of the task but will generally lie in the range of 6–13.

### **Examples**

1. Students may be given a number of measuring containers and asked to indicate how much liquid or powder is in each. This could be varied to include measuring mass, time, or temperature in appropriate settings. **Marks** awarded on basis of accuracy, inclusion of units, use of correct units etc
2. Students are given samples of coloured cardboard and asked to match them with colour charts obtainable from hardware stores. (See TAPS – Techniques for the Assessment of Practical Skills in Foundation Science, Scotland) **Marks** awarded on basis of number of correct matches
3. Students are given small samples of chemicals and asked to make and record observations of the ensuing reaction. (e.g. vinegar and antacid tablets) **Marks** awarded on basis of correct observations, using more than one sense, making observations rather than drawing inferences (e.g. ‘I smell a bad smell’ versus ‘Poison gas came out’)
4. Indicators are chemicals which change colour when mixed with other chemicals. We use them to tell the difference between chemicals.

On the desk you will find labelled bottles of indicators (Universal Indicator and Phenolphthalein) and chemicals (labelled Acid, Water and Base)

- Test each chemical with both indicators.

- Observe the colour changes that take place. Record them in the table below.
- (All required equipment to be on the desk. Students given instructions regarding disposal of equipment and chemicals)
- (Marks awarded for correctness of observations as well as describing the change ie ‘from green to red’ as well as procedure

Substance	Universal Indicator	Phenolphthalein
Acid		
Water		
Base		

**Part 1:**

Collect a Triple beam Balance

Weigh a 250 mL beaker: Mass = .....grams

Add a teaspoon of salt to your beaker

Weigh the beaker and salt together Mass = .....grams

Therefore, Mass of salt = .....grams

**Part 2:**

Add 150 mL of water to your beaker containing the salt. Stir. Record all your observations in the space below:

---

---

---

(Marks for accuracy of measurement of mass and correct calculation. Marks for appropriate description of salt dissolving in water.)

### Task 6.

This task aims to assess the student understanding of the concept of Mechanical Advantage. It is taken from a Year 10 Assessment Task and included here only as an example of what sort of scope there is in designing tasks.

(For this task, four simple lever systems (physically built up) are required with the fulcrum, load, and effort labelled and the load and effort locked in place.)

*At this station there are four simple lever systems set up. Please do not touch them. All the levers are balanced.*

1. Calculate the mechanical advantage of each lever system. Show all your working.
2. How would you be able to double the mechanical advantage of lever system 3, if you could move the load along the lever while the position of the effort stayed fixed?

Marks are awarded in a range of ways:

- System 1: Magnitude of load and effort given. Students calculate  $MA = \text{Load}/\text{Effort}$
- System 2: Magnitude of Load only is given. Students measure load arm and effort arm. From this they calculate the effort and then the Mechanical Advantage.
- System 3: Only the load arm and effort arm are given. Students work out MA from  $\text{Mechanical Advantage} = \text{Effort Arm}/\text{Load Arm}$

### Further sample

Further samples are provided as a separate Acrobat PDF file. If you are reading this report as a Microsoft Word file on a computer, you can open this file by double clicking on the icon below.



6