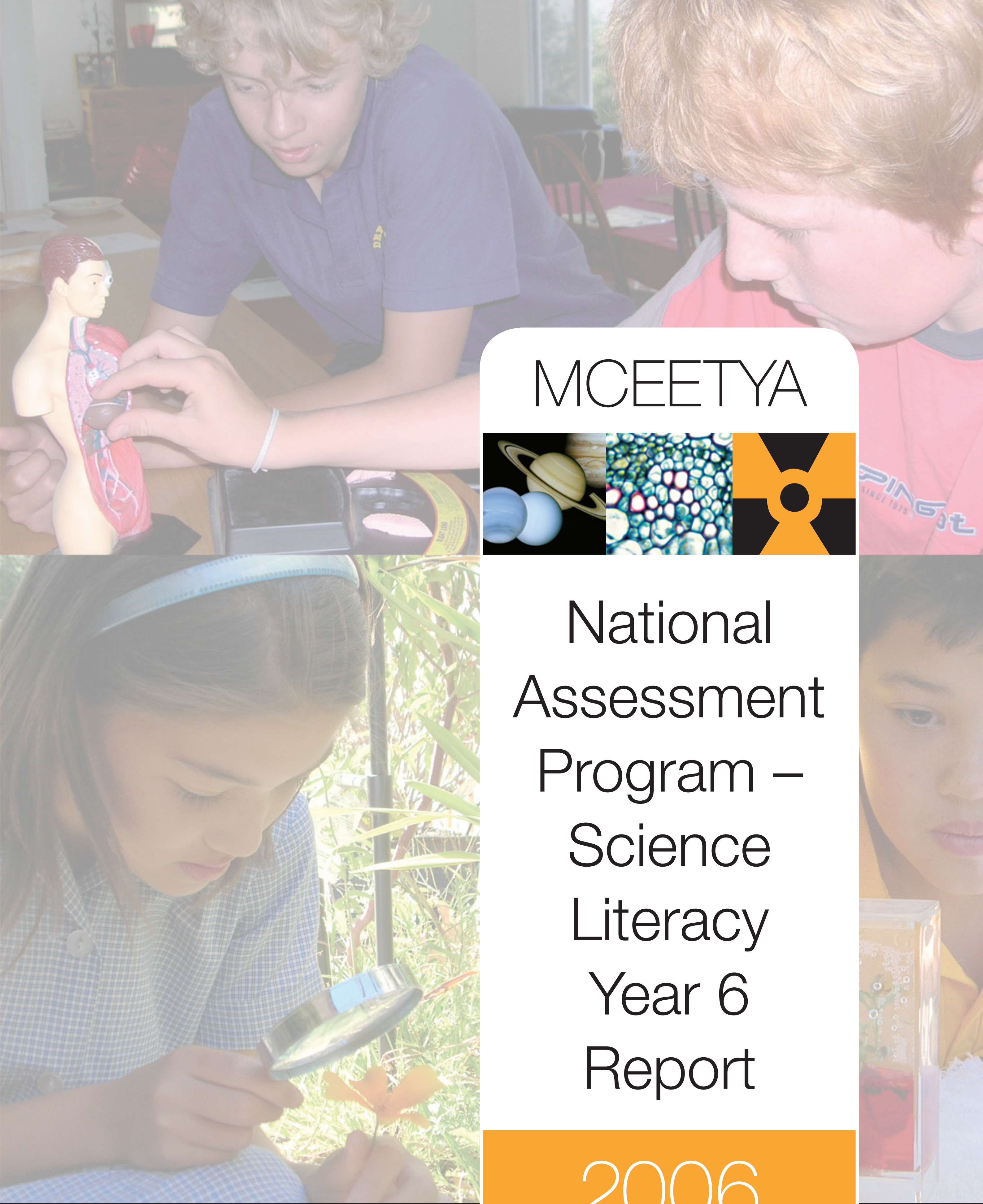


MCEETYA

National Assessment Program – Science Literacy Year 6 Report

2006

MINISTERIAL COUNCIL ON EDUCATION, EMPLOYMENT, TRAINING AND YOUTH AFFAIRS

NAP-SL 2006 Project Staff

Jenny Donovan from Educational Assessment Australia (EAA) was the Project Director of NAP-SL 2006. Melissa Lennon (EAA) was the Project Manager and Wendy Bodey from Curriculum Corporation (CC) was the Assessment Manager. The test development team was led by Gayl O'Connor (CC). This Public Report was written by Jenny Donovan, Melissa Lennon, and Gayl O'Connor and Noni Morrissey from CC.

The sampling and data analysis tasks were undertaken by Nathaniel Lewis and Goran Lazendic from EAA and Margaret Wu and Mark Dulhunty from Educational Measurement Solutions (EMS). The Technical Report was written by Margaret Wu (EMS), Jenny Donovan and Melissa Lennon (EAA).

MINISTERIAL COUNCIL ON EDUCATION,
EMPLOYMENT, TRAINING AND YOUTH AFFAIRS

Contents

List of Tables

List of Figures

Acknowledgements

National Assessment Program – Science Literacy Review Committee Members

Listed below are those individuals who, representing their school sectors, participated in the Review Committee during the development and implementation of the National Assessment Program – Science Literacy. This group made a valuable contribution to the success of the project.

Chair	Peter Titmanis
State and Territory Representatives	
Australian Capital Territory	Simon Tiller
New South Wales	Harry Vassila
Northern Territory	Lisa Scarfe
Queensland	Robyn Rosengrave
South Australia	Peter Russo
Tasmania	Meriloy Horsham
Victoria	Susan Dennett
Western Australia	Fred Deshon
Educational measurement specialist	Barry Schmitt
Language background other than English specialist	Hanya Stefaniuk
Australian Government representative	Wendy Whitham
Catholic school sector representative	Michael Doyle
Independent school sector representative	Estelle Lewis
BEMU	Perette Price Peter Bruce
PMRT Secretariat	Rhonda Horne
Indigenous education specialist	Jeanette Morris
Educational Assessment Australia	Jenny Donovan
Curriculum Corporation	Wendy Bodey
Educational Measurement Solutions	Margaret Wu

Foreword

In 2003 the first nationally-comparable science assessment was designed, developed and carried out under the auspices of the national council of education ministers, the Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA). In 2006 a second science assessment was conducted and, for the first time nationally, the achievement of students has been compared over time and publicly reported.

The National Assessment Program – Science Literacy is one of a suite of national assessments (with ICT and Civics & Citizenship) which are conducted with a random sample of students in three-yearly cycles.

Development of these nationally-comparable assessments stands as one of the most notable recent achievements of MCEETYA. The assessments are a key result of a collaborative venture between the States and Territories and the Australian Government to measure and report on how our students are progressing towards the achievement of the National Goals of Schooling in the Twenty-First Century.

This report on the 2006 National Year 6 Science Assessment provides the results from the national sample assessment. It provides a snapshot of student results against the national scientific literacy scale, and an analysis of various findings across States and Territories and student sub-groups. It also makes comparisons with the findings from 2003.

The report sets out information that will be useful for continual reflection and improvement in the teaching of science across the nation. In addition MCEETYA will be releasing Scientific Literacy School Release Materials. These will be a valuable resource for teacher professional development and the enhancement of student learning.

A separate Technical Report on the processes underlying the results of this assessment, as well as more detailed data, will be available to researchers and others via the MCEETYA website.

The next science literacy assessment will be conducted in 2009. The results of that cycle will allow us to identify trends based on three comparable datasets. The information derived from these national assessments enables us to better understand and improve our children's skills and knowledge.

I acknowledge the work of the Benchmarking and Educational Measurement Unit (BEMU) and the leadership of the Performance Measurement and Reporting Taskforce (PMRT) in the development and implementation of the National Assessment Program. I commend this report to teachers and educators and to all those with an interest in the education of our children.

Rachel Hunter
Chair
Performance Measurement and Reporting Taskforce

Executive Summary

In July 2001, the Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) agreed to the development of assessment instruments and key performance measures for reporting on student skills, knowledge and understandings in primary science. It directed the newly established Performance Measurement and Reporting Taskforce (PMRT), a nationally representative body, to undertake the national assessment program.

The PMRT established a number of national committees to advise it on critical aspects of the study and ensure that the assessments and results were valid across the States and Territories. The main function of these committees was to ensure that the scientific literacy assessment domain was inclusive of the different State and Territory curricula and that the items comprising the assessments were fair for all students, irrespective of where they attended school.

The National Assessment Program – Science Literacy measures scientific literacy. This is the application of broad conceptual understandings of science to make sense of the world, understand natural phenomena and interpret media reports about scientific issues. It also includes asking investigable questions, conducting investigations, collecting and interpreting data and making decisions. The construct evolved from the definition of scientific literacy used by the Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment (PISA):

> *… the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity.*
>
> (OECD 1999, p. 60)

The first national assessment was conducted in 2003. The Primary Science Assessment Program (PSAP) – as it was then known – tested a sample of Year 6 students. PSAP results were reported in 2005.

In 2006, a consortium of Educational Assessment Australia and Curriculum Corporation conducted the second national science assessment. The National Assessment Program – Science Literacy tested a sample of Year 6 students. The findings describe the scientific literacy of Year 6 Australian students.

Assessment domain

The assessment domain and instruments were developed in consultation with curriculum experts from each State and Territory and representatives from the Catholic and independent school sectors.

The domain outlined the development of scientific literacy across three main areas:

Strand A: formulating or identifying investigable questions and hypotheses, planning investigations and collecting evidence.

Strand B: interpreting evidence and drawing conclusions from their own or others' data, critiquing the trustworthiness of evidence and claims made by others, and communicating findings.

Strand C: using science understandings for describing and explaining natural phenomena, and for interpreting reports about phenomena.

The assessment items drew on four concept areas:

- Life and Living
- Earth and Beyond
- Natural and Processed Materials
- Energy and Change.

These evolved from a review of the 'National Statements and Profiles' and were common across Australian curricula at the time of test development.

In August 2006 the Statements of Learning were endorsed by Ministers of Education in all states and territories. Future National Assessment Program – Science Literacy tests will draw on concepts and content of the Statements of Learning in Science.

Assessment instruments

The assessment instruments were administered to a random sample consisting of 4.8 per cent of the total Australian Year 6 student population. The students' regular classroom teachers administered the National Assessment Program – Science Literacy on the following dates:

- 18 October 2006 – Northern Territory, Queensland, Tasmania
- 25 October 2006 – Australian Capital Territory, New South Wales, South Australia, Victoria, Western Australia.

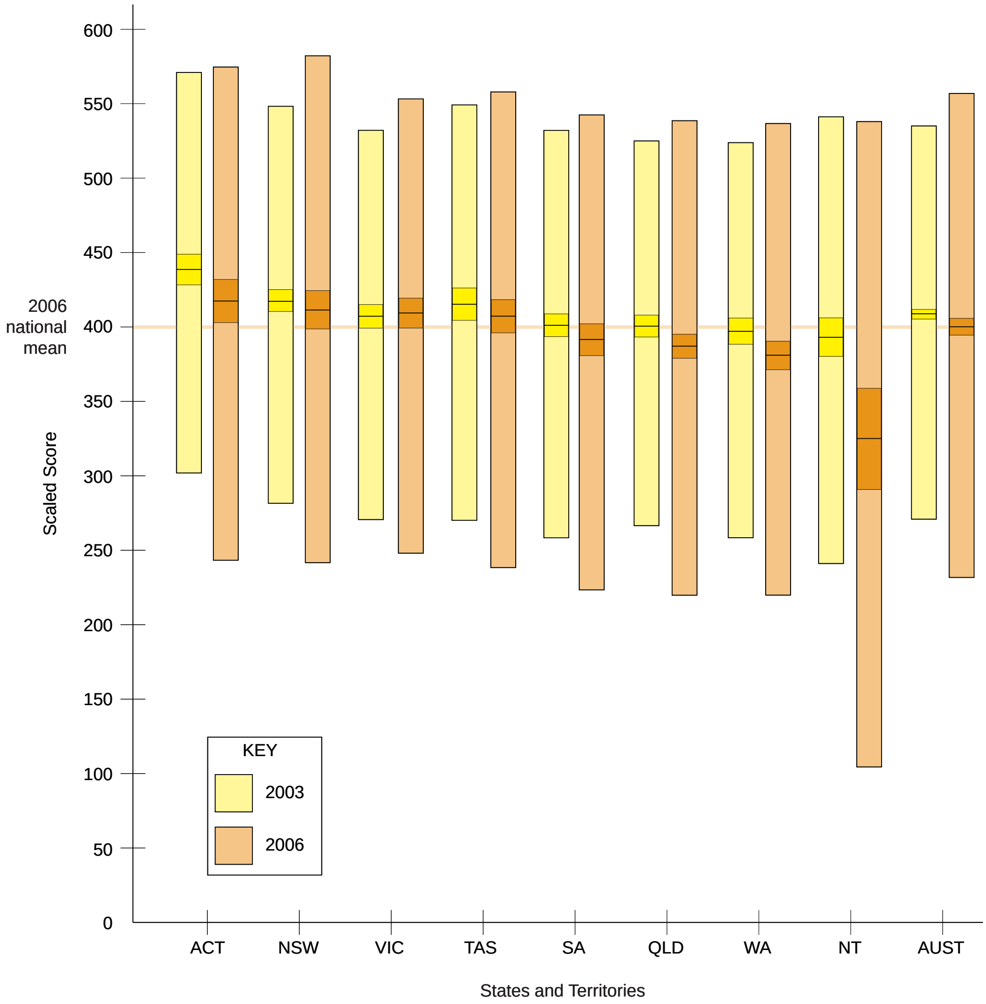
The assessment instruments consisted of seven pen-and-paper assessments, including multiple-choice and short-answer type items, and two practical tasks. Each student completed one of the pen-and-paper assessments and one of the practical tasks. Students were allowed 60 minutes for the pen-and-paper assessment and 45 minutes for the practical task. The practical tasks required the students to conduct an experiment in groups of three and then respond individually to a set of questions about the experiment.

Student performance in scientific literacy (2003 and 2006)

One of the main objectives of the National Assessment Program – Science Literacy is to monitor trends over time. One way of doing so is to compare mean achievement scores and the distribution of student scores on a scale. The science literacy scale was established in 2003. In 2006 the sample frame was designed to be more inclusive of remote schools and to provide better discrimination between students in the top half of the achievement distribution. As a result 2006 was set as the baseline scale for the National Assessment Program – Science Literacy and the 2003 results were re-scaled and mapped onto it. The adjusted tables for the 2003 results are shown in the relevant chapters of this Report with the 2006 data.

The following figure shows the mean scores and distributions for 2003 and 2006.

Figure ES.1 Comparison of distributions of Year 6 student performance by State and Territory in 2003 and 2006



Note: 2003 results rescaled to 2006.

In technical terms, the darker coloured bands mark the likely range of the mean of the true population score. This is know as the confidence interval (CI).

The distribution of student performances in the National Assessment Program – Science Literacy shows that the Australian Capital Territory was the only State or Territory which performed significantly above the national mean in 2003 and 2006.

At the national level, the results across the scientific literacy Proficiency Levels showed the following trends:

- for males and females, there were no significant differences in proficiency
- Indigenous students had significantly lower mean achievement than non-Indigenous students (see Table ES.1)
- students in remote and very remote areas had significantly lower mean achievement than students in all other geographic locations (see Table ES.2).

Missing student background data meant that it was not possible to calculate the proficiency of the language background other than English (LBOTE) group.

Table ES.1 2006 mean scores for Indigenous and non-Indigenous students

Student group	Mean score	95% CI
Indigenous	311	±29.4
Non-Indigenous	402	±5.8

Table ES.2 2006 mean scores for students by geographic location

Code	Geographic location	% of students	Mean score	95% CI
1.1	Metropolitan zone capital city	58.8	404	±8.5
1.2	Major urban statistical district	12.4	406	±11.0
2.1	Provincial city statistical district	9.2	395	±12.1
2.2	Inner and outer provincial areas	16.6	396	±8.5
3	Remote and very remote areas	3.0	331	±29.9
	All	100.0	400	±5.4

Note: geographic locations as supplied and defined by MCEETYA.

Standard for Year 6 scientific literacy

A standard for scientific literacy was established after the 2003 assessment to provide parents, educators and the community with a clear picture of the proficiency that students are expected to demonstrate by the end of Year 6.

To identify what students should know and be able to do by the end of Year 6, university science educators, curriculum officers and experienced primary teachers in all States and Territories, from government, Catholic and independent schools, were brought together. The crucial scientific literacy skills and understandings needed by students for the next phase of science learning at school were discussed and debated before consensus was reached on a 'proficient' standard for Year 6. This standard informed the development of the tests for the 2006 assessment.

The proficient standard is a challenging level of performance, with students needing to demonstrate more than minimal or elementary skills to be regarded as reaching it. It is one of several achievement levels that collectively represent a continuum of learning and describe what students know and are able to do. Students who have not achieved the proficient standard have demonstrated only partial mastery of the skills and understandings expected for Year 6; these students are on the way to becoming proficient. There are also students who have shown superior results and exceeded the proficient standard.

Initially, in 2003, three Proficiency Levels, corresponding with Levels 2, 3 and 4 of the Scientific Literacy Progress Map, were identified.

However, as 90 per cent of students' scores were within Level 3 in 2003, three further Proficiency Levels within Level 3 were created, providing five levels for reporting student performance in the assessment. The proficiency standard was deemed to be Level 3.2 on the Proficiency Level continuum.

Minimum standards like the benchmarks in literacy and numeracy have not been set for scientific literacy. These benchmarks are defined as the critical level of skill and understanding without which a student will have difficulty making sufficient progress at school. They are more suited to foundational areas such as reading, writing and numeracy where deficiencies will have significant effects on students' future learning and functioning in society.

Information about students' performances in relation to the Year 6 standard from the second national Year 6 science literacy assessment is summarised below. The results in Table ES.3 show the percentage of students in each of the Proficiency Levels established for scientific literacy, while Figure ES.2 (on page xvii) provides a comparison between the percentages of students who achieved each of the Proficiency Levels in 2003 and 2006.

Table ES.3 Percentages of students in Proficiency Levels by State and Territory 2006

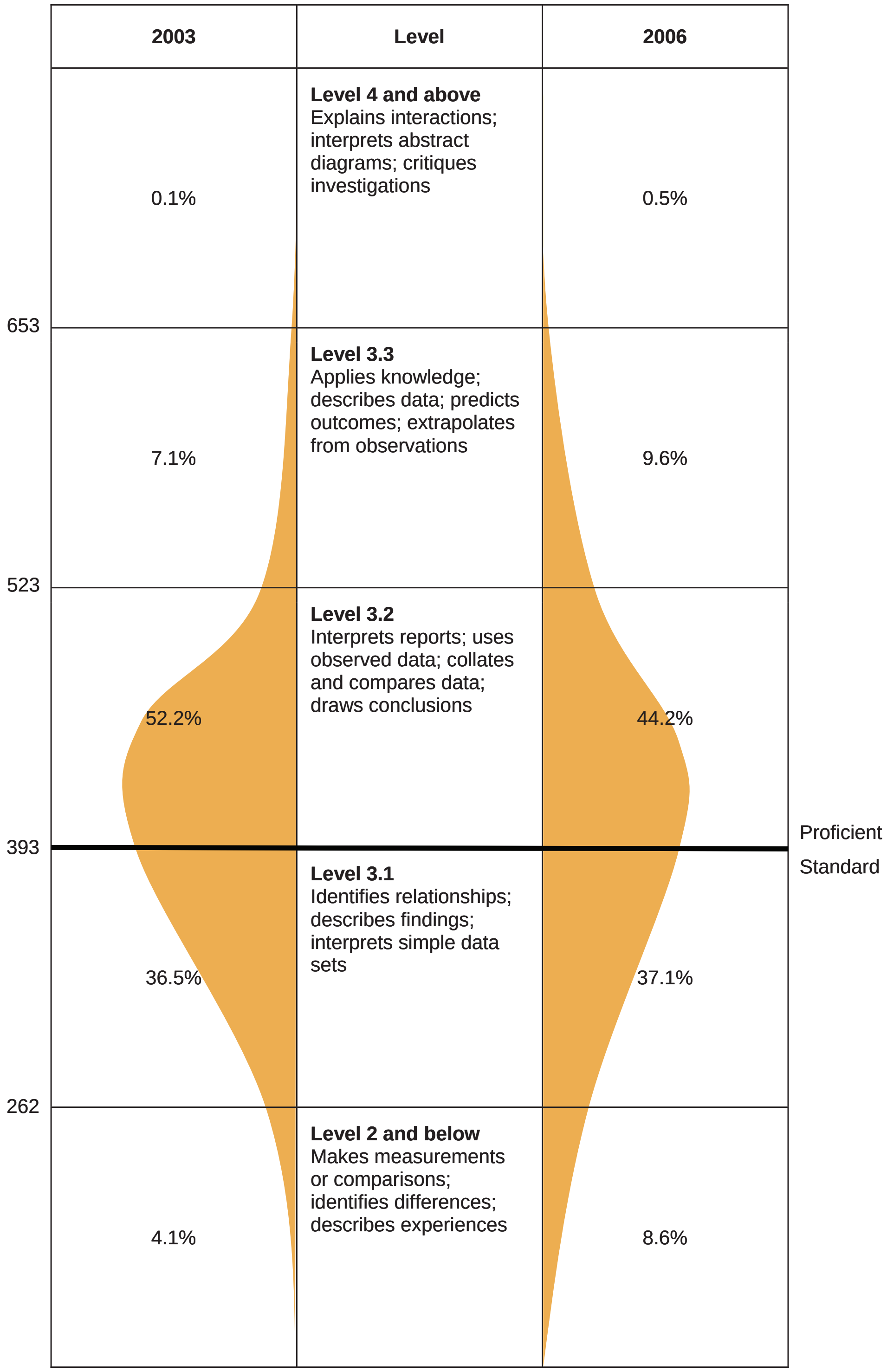
State/ Territory	Proficiency Level					
	2 and below	3.1	3.2	3.3	4 and above	At or above Prof. Std
NSW	7.4 (±2.0)	35.2 (±3.5)	**43.9 (±3.6)**	12.3 (±3.1)	1.2 (±1.2)	57.4 (±4.3)
VIC	6.5 (±2.3)	35.2 (±3.9)	**48.5 (±4.1)**	9.6 (±2.2)	0.2 (±0.4)	58.3 (±5.0)
QLD	10.2 (±2.5)	40.6 (±2.9)	**42.0 (±3.6)**	7.0 (±1.7)	0.2 (±0.2)	49.2 (±3.8)
SA	9.6 (±2.2)	38.7 (±3.7)	**43.6 (±3.9)**	7.9 (±2.3)	0.1 (±0.2)	51.6 (±4.7)
WA	11.5 (±2.6)	42.0 (±3.7)	**39.6 (±4.0)**	6.8 (±2.3)	0.2 (±0.2)	46.6 (±4.7)
TAS	7.6 (±2.3)	34.9 (±4.4)	**46.7 (±4.7)**	10.4 (±3.0)	0.3 (±0.4)	57.4 (±5.5)
NT	28.6 (±7.5)	33.0 (±5.5)	**31.6 (±5.4)**	6.7 (±2.8)	0.2 (±0.4)	38.4 (±6.5)
ACT	7.3 (±2.5)	30.7 (±4.8)	**47.9 (±4.8)**	13.5 (±4.0)	0.6 (±1.1)	62.0 (±5.6)
AUST	8.6 (±1.1)	37.1 (±1.7)	**44.2 (±1.8)**	9.6 (±1.2)	0.5 (±0.4)	54.3 (±2.1)

Note: figures in parentheses refer to 95 per cent confidence intervals.

The results show that the ACT has the highest proportion of students at the proficient standard or above, i.e. Level 3.2 and higher. It should be noted that, as with mean scores, when confidence intervals are taken into account, it is unlikely that there will be a significant difference between the ACT, NSW, Victoria or Tasmania in terms of the proportion of students achieving the proficient standard.

Nationally, 54.3 per cent of students achieved or bettered the proficient standard, compared with 59.4 per cent in 2003. Approximately 91.4 per cent achieved Level 3.1 or above, compared with 95.9 per cent in 2003. The highest Proficiency Levels (Levels 3.3 and 4 and above) were achieved by approximately 10.1 per cent of students, compared with 7.2 per cent in 2003.

Figure ES.2 Distribution of students in Proficiency Levels for 2003 and 2006



Note: 2003 results rescaled to 2006.

The nation-wide percentage of students achieving or bettering the proficient standard (54.3 per cent) was exceeded in the ACT, NSW, Victoria and Tasmania.

Table ES.4 Changes in results achieved from 2003 to 2006

State/ Territory	Level 2	Level 3.1	Level 3.2	Level 3.3	Level 4
NSW	+4.2	+1.7	-10.1	+3.3	+1.0
VIC	+2.6	-1.3	-4.8	+3.3	+0.1
QLD	+5.6	+0.7	-8.1	+1.7	+0.1
SA	+4.2	+0.4	-6.0	+1.4	+0.1
WA	+6.1	+1.6	-9.4	+1.6	+0.1
TAS	+3.5	+1.4	-6.1	+1.1	+0.1
NT	+21.2	-8.8	-11.6	-1.0	+0.2
ACT	+5.4	+4.3	-8.9	-1.1	+0.3
AUST	+4.5	+0.6	-8.0	+2.5	+0.4

Note: 2003 results rescaled to 2006. Sign indicates direction of change from 2003 to 2006.

The percentage of students in the highest Proficiency Level is higher than in 2003. The percentage of students in the lowest Proficiency Level is also higher than in 2003. The percentage of students at or above Level 3.1 is slightly lower than in 2003. None of these differences is statistically significant.

Chapter 1
Overview of the National Assessment

Introduction

In 1999, the State, Territory and Commonwealth Ministers of Education agreed to the new Adelaide Declaration on National Goals for Schooling in the Twenty First Century (Ministerial Council on Education, Employment, Training and Youth Affairs [MCEETYA] 1999) (available online on the MCEETYA website at www.mceetya.edu.au).

The National Goals provide the framework for reporting on student achievement through the annual MCEETYA publication, the *National Report on Schooling in Australia* (ANR).

In July 2001, MCEETYA agreed to the development of assessment instruments and key performance measures for reporting on student skills, knowledge and understandings in primary science. It directed the Performance Measurement and Reporting Taskforce (PMRT) to undertake the national assessment program.

The PMRT set the policy objectives and established a Steering Committee to manage the assessment and a Consultative Committee to facilitate discussion among the jurisdictions and school sectors. The Consultative Committee also provided feedback about the appropriateness of the conceptual framework and reviewed the assessment items to ensure that they were inclusive of all the States and Territories' curricula.

The National Science Assessment was the first assessment program designed specifically to provide information about performance against the National Goals. MCEETYA has also endorsed similar assessment programs to be conducted for Civics and Citizenship, and Information and Communications Technology (ICT). The intention is that each assessment program will be repeated every three years so that performance in these areas of study can be monitored over time.

Apart from being the first subject area, science is the only program that focuses entirely on primary school performance. This is because MCEETYA has agreed to use PISA as the measure of performance for secondary science.

In 2005, PMRT awarded the contract for the second cycle of science testing, due in 2006, to a consortium of Educational Assessment Australia (EAA) and Curriculum Corporation (CC). The Benchmarking and Educational Measurement Unit (BEMU) was nominated by PMRT to liaise between the contractors and PMRT in the delivery of the project.

The National Assessment Program – Science Literacy

Implementation of the National Assessment Program – Science Literacy involved a large number of separate but related steps, including the development of items and instruments to assess the assessment domain; the trialling of those items and assessment instruments; the administration of the assessment to a sample of students; and the marking, analysis and reporting of the results.

This report provides details about the school and student samples used, describes the testing process, presents the results at the national and State and Territory levels and includes comparisons with 2003.

What does the National Assessment Program – Science Literacy measure?

The National Assessment Program – Science Literacy measures scientific literacy.

The OECD-PISA (1999) has defined scientific literacy as:

> *… the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity.*
>
> (OECD 1999, p. 60)

This definition has been adopted for the National Assessment Program – Science Literacy 2006 in accord with the Ball et al. 2000 report recommendation.

The science items and instruments therefore assess outcomes that contribute to scientific literacy, such as conceptual understandings, rather than focusing solely on facts. They also assess student competence in carrying out investigations in realistic situations.

The National Assessment Program – Science Literacy relates to the ability to think scientifically in a world in which science and technology are increasingly shaping children's lives.

A Scientific Literacy Progress Map (see Appendix 1) has been developed based on the construct of scientific literacy and on an analysis of the State and Territory curriculum and assessment frameworks. The Progress Map describes the development of scientific literacy across three strands of knowledge which are inclusive of Ball et al.'s concepts and processes and the elements of the OECD–PISA definition.

What aspects of scientific literacy were assessed?

Three main areas of scientific literacy were assessed:

Strand A: formulating or identifying investigable questions and hypotheses, planning investigations and collecting evidence.

Strand B: interpreting evidence and drawing conclusions from their own or others' data, critiquing the trustworthiness of evidence and claims made by others, and communicating findings.

Strand C: using science understandings for describing and explaining natural phenomena, and for interpreting reports about phenomena.

A conscious effort was made to develop assessment items that related to everyday contexts.

The scientific literacy domain is detailed in Appendix 1. In addition the items drew on four concept areas: Life and Living; Earth and Beyond; Natural and Processed Materials; and Energy and Change. The major scientific concepts found most widely in States and Territories were used by item developers to guide test development. The list of endorsed examples for each of these major concepts is in Table A1.2 of Appendix 1.

The intention was to ensure that all Year 6 students were familiar with the materials and experiences to be used in the National Assessment Program – Science Literacy and so avoid any systematic bias in the instruments being developed.

What is the national scientific literacy standard?

A standard for scientific literacy was established as part of the first cycle of national assessment in 2003 to provide parents, educators and the community with a clear picture of the level of proficiency that students are expected to demonstrate by the end of Year 6.

The standard for scientific literacy used for the second cycle of national assessment in 2006 was consistent with that used in the first cycle in 2003. This made it possible to compare performance in 2003 and 2006 and gauge whether student proficiency had improved.

To identify what students should know and be able to do by the end of Year 6, university science educators, curriculum officers and experienced primary teachers in all States and Territories, from government, Catholic and independent schools, were brought together.

The members of this expert group used their classroom experience and knowledge of the science curricula in the various jurisdictions to examine the test items from the national assessment.

The crucial science-literacy skills and understandings needed by students for the next phase of science learning at school were discussed and debated before consensus was reached on a 'proficient' standard for Year 6.

The proficient standard is a challenging level of performance, with students needing to demonstrate more than minimal or elementary skills to be regarded as reaching it. It is one of several achievement levels that collectively represent a continuum of learning and describe what students know and are able to do.

In terms of the Proficiency Levels described in Chapter 5, the standard was found to be equivalent to Level 3.2: that is, students achieving at Level 3.2 or better are considered to have a sound understanding of Year 6 science. Students at this level demonstrate considerably more skill and understanding than those performing at Levels 3.1 and below.

Year 6 students who exceed the proficient standard (those who perform at Level 3.3 and above) demonstrate exemplary performance.

Students who have not achieved the proficient standard have demonstrated only partial mastery of the skills and understandings expected for Year 6; these students are on the way to becoming proficient.

Minimum standards like the benchmarks in literacy and numeracy have not been set for scientific literacy. These benchmarks are defined as the critical level of skill and understanding without which a student will have difficulty making sufficient progress at school. They are more suited to foundational areas such as reading, writing and numeracy where deficiencies will have significant effects on students' future learning and functioning in society.

The proficient standard (equivalent to Level 3.2) will be the main reference point for monitoring scientific literacy in Australian primary schools over time. Every three years a new national Year 6 science assessment will be conducted to gauge whether student proficiency has improved.

Information about students' performances in relation to the Year 6 standard from the second (2006) National Assessment Program – Science Literacy is reported with comparisons to 2003 data by proficiency levels in Chapter 5.

Who participated in the 2006 National Assessment Program – Science Literacy?

Approximately 5 per cent of the total Australian Year 6 student population was sampled randomly and assessed. The sample was drawn from all States and Territories. Government, Catholic and independent schools participated. Table 1.1 shows the number of schools and students in the final sample for which results were reported.

A grade-based population of students enrolled at schools was chosen. This is consistent with the reporting of literacy and numeracy performance in the Annual National Report (ANR). There are differences between the States and Territories in the structure and organisation of pre-primary education and the age of entry to full-time formal schooling. Information about structural differences that may assist interpretation of the results of the testing is summarised in Table 3.1 (page 18).

Appendix 2 provides a comprehensive summary of the sample frame, with exclusions and response rates for participating schools and students by State and Territory for the assessment.

Table 1.1 Number of schools and students by State and Territory in the final sample 2006

State/ Territory	Number of schools in target sample	Number and percentage of schools in final sample	Number of students in target sample	Number and percentage of students in final sample
ACT	57	57 (100%)	1345	1271 (94.4%)
NSW	92	90 (97.8%)	2104	2039 (94.0%)
NT	49	43 (87.8%)	932	740 (88.3%)
QLD	94	94 (100%)	2116	2016 (91.8%)
SA	94	93 (98.9%)	2087	1809 (90.9%)
TAS	64	64 (100%)	1397	1225 (92.1%)
VIC	91	88 (96.7%)	2098	1810 (90.7%)
WA	95	92 (96.8%)	2093	2001 (91.9%)
AUST	636	621 (97.6%)	14172	12911 (92%)

Note: the student participation percentage calculation includes within-school exclusions.

What did the National Assessment Program – Science Literacy participants have to do?

There were seven pen-and-paper (objective) assessments which included multiple-choice and short-answer items. There were also two practical assessment tasks. The assessment papers included common (link) items. The papers were distributed randomly so that each of the students in a class completed one of the seven pen-and-paper assessments.

However, all students in the same class undertook the same practical task. The practical tasks were assigned to classes across Australia in a way that ensured approximately equal numbers of classes attempted each of the two tasks.

The practical tasks required the students to work in groups of three. Teachers allocated students randomly to groups, using a procedure outlined in the Test Administrator's Manual. Students conducted the experiment in these groups and responded as a group to a set of questions designed to stimulate group discussion about the experiment.

The students then answered a further set of items independently. The individual student responses were the only ones used in the analysis and generation of proficiency data.

Equating of the seven objective assessments onto one scale was achieved by the use of common items shared between the assessments. The practical items were then linked onto this scale by results obtained from students doing the same objective assessment and practical task.

Students were allowed 60 minutes to complete the pen-and-paper assessments and 45 minutes for the practical tasks.

The students' regular classroom teachers administered the National Assessment Program – Science Literacy on:

- 18 October 2006 – Northern Territory, Queensland, Tasmania
- 25 October 2006 – Australian Capital Territory, New South Wales, South Australia, Victoria, Western Australia.

How are the National Assessment Program – Science Literacy results reported?

The results of the National Assessment Program – Science Literacy are reported as mean scores and distributions of scores. They are also described in terms of the understandings and skills that students demonstrated in the Assessment: these understandings and skills are mapped against the scientific literacy assessment framework.

Five levels of proficiency are defined and described for scientific literacy. Further details of the proficiency scales, including results in relation to the scales by State and Territory, are contained in Chapter 4, Interpreting the Scientific Literacy Results.

Results for groups such as Indigenous students and students from different geographic locations are also presented in this report.

How is this report organised?

Chapter 2 provides a more detailed discussion of the scientific literacy scale, the assessment domain and the assessment procedures.

Results in terms of means and distributions of student performance are presented in Chapter 3.

Chapter 4 discusses the results in terms of students' proficiency on the scientific literacy scale. The scale links the students' results to descriptions of their understandings and skills in the assessment domain. Further information about the nature and coverage of the assessment tasks accompanies the discussion of students' results.

Chapter 5 examines comparisons in achievement by Proficiency Levels between the tests in 2003 and the 2006 cycle.

Chapter 6 provides an analysis of the results achieved by specific groups of students, including boys and girls, Indigenous and non-Indigenous students and students from diverse geographic locations.

Chapter 2
The Scientific Literacy Scale

Introduction

This chapter provides a brief description of the steps that were used to define and construct the scientific literacy scale.

More detailed information about each of the steps is provided in the various publications that are referred to in this chapter.

Very high standards were set for sampling, constructing assessment materials and undertaking operational procedures in order to ensure the integrity of the data.

Scale construction

In 2003, a measurement and reporting scale for scientific literacy was developed that would enable standards and changes in student proficiency over time to be monitored.

To compare the results of the 2006 and 2003 assessments, the following steps were undertaken, involving a number of inter-related tasks:

1. clarifying the assessment strands for scientific literacy
2. constructing assessments that comprised items and tasks which defined the assessment strands operationally
3. administering the assessments to students
4. using the measurement model and technical standards to analyse the results.

As in 2003, the PMRT established a number of national committees to ensure that the assessments and results were valid across the States and Territories and to advise it on critical aspects of the study.

The main function of these committees and groups was to ensure that the assessment strands of scientific literacy were inclusive of the different State and Territory curricula and that the items comprising the assessments were fair for all students irrespective of where they attended school.

For the 2006 cycle a Science Literacy Review Committee (SLRC) was established, including representatives from all sectors, States and Territories. The SLRC was consulted about item development and review, as well as other issues as they arose.

A brief description of the steps involved in aligning the scientific literacy measurement and reporting scale is provided here.

1. **Clarifying the assessment strands for scientific literacy**

   A common understanding of the Progress Map, the descriptions of each strand, and a hierarchy of students' understandings and skills in the concept areas were developed (see Appendix 1).

2. **Constructing assessments that comprised items and tasks which defined the assessment strands operationally and covered the full range of proficiency expected to be represented in Year 6 classes**

   In consultation with EAA/CC, BEMU (on behalf of PMRT) approved the more technical aspects of the assessment design, including, for example, the number of assessment booklets, the ratio of multiple-choice to open-ended items in the booklets, and the number of items per strand per test booklet.

   Test constructors developed items and tasks that enabled students at different points along the scale to demonstrate what they knew and could do in terms of scientific literacy. The constructors had to ensure that the tasks assessed the outcomes articulated in the assessment strands. They also had to ensure that the tasks intended to assess higher-order understandings and skills at the top of the scale were more difficult than those at the middle and bottom of the scale.

The items were reviewed first by EAA/CC and their internal panels, and then by advisory committees and other key staff in the States and Territories. This was followed by trialling with samples of students in four States and Territories and then a further review by the advisory committees and other key staff in the States and Territories. The emphasis during these reviews was on ensuring that the items and tasks reflected the understandings and skills in the assessment strands and were not biased unduly for or against particular groups of students.

3. **Administering the assessments to students**

   Once the items and tasks had been written, they were trialled with a sample of students in 31 schools selected from the government, Catholic and independent sectors in New South Wales, Victoria, the Australian Capital Territory and South Australia.

   The results were analysed to determine the degree to which the items and tasks measured the scientific literacy domain. The committees then reviewed the data from the trial testing, gauged the validity of the assessments and suggested modifications where necessary. These modifications were included in the revised assessments.

   The final assessments were administered to a stratified random sample of students in October 2006. The total number of students in the final sample was 14 172 at 636 schools. Information about the achieved sample is shown in Table A2.4 and Table A2.5 of Appendix 2.

4. **Using the measurement model and technical standards to analyse the results.**

   Item Response modelling was used to analyse the results from the sample of students who participated in the National Assessment Program – Science Literacy. These statistical models are used in all State and Territory testing programs and in major international testing programs such as PISA and the Trends in International Mathematics and Science Study (TIMSS).

   Details of the application of the Rasch model can be found in the Technical Report for the National Assessment Program – Science Literacy.

In Chapter 4, additional meaning and depth are added to the summary statistics by referencing the data to descriptions of the understandings and skills students were able to demonstrate, using examples of test items.

The assessment booklets

In 2006 the National Assessment Program – Science Literacy involved the use of seven assessment booklets. Only four test booklets had been used in 2003, but for 2006 a balanced incomplete block (BIB) design was used, similar to that used in PISA. BIB is achieved by arranging the items in clusters so that every cluster appears with another cluster once, and every cluster appears in each of the three possible block positions. The effect of the BIB design is to reduce the possibility that an item’s location in a test booklet has an impact on its difficulty.

To achieve the BIB design for the National Assessment Program – Science Literacy, the items were first written in units. Each unit had a context and contained between one and five items. Clusters were then constructed by grouping four to six units together. From there booklets were compiled by arranging three clusters in every booklet. In total there were seven different clusters across the seven booklets. Each booklet had approximately 39 items in the objective section.

The multiple-choice items in the booklets had only single correct answers. The open-ended items required students to construct their own responses. They were categorised into those that required a single word or short sentence response (short-answer items) and those that required more substantive responses (extended-response items).

Each booklet contained an objective (pen-and-paper) test and two practical tasks. Participating students had to complete the objective section of their booklet and one of the two practical tasks. The practical task required students to undertake an activity in small groups and then respond individually to either nine or ten pen-and-paper items related to the activity.

Coverage of scientific literacy

The distribution of items across the assessment domain for scientific literacy (each strand and major conceptual area) is shown in Table 2.1. There were 110 items distributed across the seven pen-and-paper tests and two practical tasks. Each student had to sit for one pen-and-paper test and one practical task.

Table 2.1 Distribution of assessment items across the assessment strands for scientific literacy 2006

Domain	Item type and number of items			
	Multiple-choice	Short-answer	Extended-response	Total
Distribution of items by strand				
Strand A	2	0	6	8
Strand B	31	3	17	51
Strand C	16	10	25	51
Total	49	13	48	110
Distribution of items by major science conceptual area				
Life and Living	12	3	21	36
Earth and Beyond	23	1	13	37
Natural and Processed Materials	11	5	6	22
Energy and Change	3	4	8	15
Total	49	13	48	110

Major science concepts

The scientific literacy strands specify processes and concepts, rather than traditional subject boundaries such as physics, chemistry or biology.

The strands describing these processes and concepts of scientific literacy have been listed in Table 2.1 above. They were considered to be more relevant to students at primary school and, according to PISA, 'to all people in their lives beyond school than the more traditional subject areas ...' (Lokan et al. 2000, p. 97).

It can be seen from Table 2.1 that the items were distributed across four conceptual areas.

The domain

The scientific literacy domain comprises the following three strands:

Strand A involves experimental design and data gathering. More specifically, it involves skills such as formulating or identifying investigable questions and hypotheses, planning investigations and collecting evidence.

Strand B involves interpreting experimental data and requires skills such as interpreting evidence and drawing conclusions from students' own or others' data, critiquing the trustworthiness of evidence and claims made by others, and communicating findings.

Strand C involves using scientific understandings for describing and explaining natural phenomena, and for interpreting reports about phenomena.

Table 2.1 shows that 59 of the items assessed the process strands (8 for Strand A and 51 for Strand B) and 51 assessed the conceptual understanding strand (Strand C).

Types of assessment items

The strands and major concepts of scientific literacy were assessed through a range of item types (Table 2.1). Of the items, 49 were classified as being multiple-choice, 13 as short-answer and 48 as extended-response.

All the items were presented in item sets or units, with between one and five items pertaining to each stimulus text and/or diagram(s).

The sampling procedures

As for 2003, the sample was selected using procedures similar to those followed in PISA and TIMSS. The distribution of schools from the various sectors in each State and Territory was drawn according to each sector's proportion of Year 6 enrolments. The sampling in the National Assessment Program – Science Literacy followed the same rigorous procedures as PISA.

However, there were some important differences in the sampling frame in 2006.

In 2006, as many schools were included in the defined population as possible. Essentially this meant that there were no school-level exclusions from the supplied sampling frame prior to sample selection. If a small school (with fewer than three Year 6 students) was selected, then this school could administer the pen-and-paper tasks only.

The number of students sampled in each jurisdiction was determined with the following considerations in mind:

- results for each jurisdiction should be of similar precision. While this was an ultimate goal, it was recognised that reduced sample sizes would be needed for the smaller jurisdictions (ACT, NT and TAS)
- the nationwide achieved sample was to be approximately equal to 12 000 students who were to be located within approximately 600 schools throughout Australia.

Further information about the characteristics of the sample, including details of students who were granted exemptions or excluded from the sample and the procedures used to determine the standard errors of estimates, is provided in the Technical Report (available online on the MCEETYA website at www.mceetya.edu.au).

Assessment administration procedures

Students' regular class teachers administered the National Assessment Program – Science Literacy, so as to minimise disruption to the normal class environment.

Standardised administration procedures were developed and published in a Test Administrator's Manual. In all schools in which students were to complete the National Assessment Program – Science Literacy, teachers and school administrators were provided with the Manual. Detailed instructions were also given in relation to the participation or exclusion of students with disabilities and students from language background other than English (refer to Table A2.6 in Appendix 2).

Teachers were able to review the Test Administrator's Manual before the assessment date and raise questions with the coordinators of the National Assessment Program – Science Literacy in their jurisdiction. A toll-free telephone number and email address were provided.

A quality-monitoring program was established to gauge the extent to which class teachers followed the specified administration procedures. This involved trained monitors observing the administration of the Assessment in a random sample of classes in 30 of the 630 schools involved. The monitors reported conformity with the administration procedures.

Marking of responses to open-ended items

Over half of the items were open-ended and required marking by trained markers. Some items involved single answers or phrases that could be marked objectively.

Marking Guides were prepared by EAA and CC, and refined during the trialling process. The marking team included experienced teacher-markers employed by EAA.

The markers participated in a five-hour training session conducted by a member of the test construction team. The session involved formal presentations by the trainers, followed by hands-on practice with sample student answer books. In addition, the markers undertook a further two hours of marking in which a pair of markers marked the same student answer books and moderators reconciled differences in discussion with the markers.

Markers were monitored constantly for reliability by having samples of their student answer books check-marked by group leaders. In cases where there were differences between markers and group leaders, the scoring was reconciled jointly in consultation with the professional leader. This procedure, coupled with the intensive training at the beginning of the marking exercise, ensured that markers applied the scoring criteria consistently.

Data entry procedures

The multiple-choice responses and teacher-marked scores were data processed. A validation of the data processing was performed that ensured accuracy in data capture.

Scanning software was used to capture images of all the student responses. These have been indexed and provided to BEMU for future reference.

School reports

Schools that participated in the National Assessment Program – Science Literacy were provided with feedback about the performance of their students on the Assessment prior to the close of the 2006 school year. The reports showed the results for each student on an item-by-item basis and comparative data showing the percentage of the school and the national sample of students responding correctly to the item or, in the case of items that had more than one mark available for the response to the item, the percentage of students achieving the maximum score on the item.

Scientific Literacy School Release Materials

Some assessment items have been released from the 2006 National Assessment Program – Science Literacy to enable teachers to administer the tasks under similar conditions and gauge their own students' proficiency in relation to the national standards. The Scientific Literacy School Release Materials comprise an objective test containing 37 multiple-choice and short-answer questions, and a practical task. The Scientific Literacy School Release Materials will be made available on the MCEETYA website at www.mceetya.edu.au

The remaining 2006 assessment items have been secured for the purpose of equating the next National Assessment Program – Science Literacy (which is to be undertaken in 2009) and, together with the 2003 assessment, will allow longitudinal data on student performance to be obtained.

Chapter 3
Profile of Student Performance in Scientific Literacy for 2006 with comparisons to 2003

Introduction

In this chapter, summary statistics for the National Assessment Program – Science Literacy are shown in terms of students' mean scores and distributions of scores by State and Territory.

In Chapter 4, additional meaning and depth are added to the summary statistics by referencing the data to descriptions of the understandings and skills students were able to demonstrate, using sample assessment items.

Interpreting the results

The sample

As described in Chapter 2, in 2006 the sampling frame for the Assessment was slightly different from that used in 2003. The population definition for the National Assessment Program – Science Literacy 2006 was more inclusive, and this change did have an impact on the composition of the Northern Territory cohort in particular.

Remote schools are often educationally disadvantaged and the Northern Territory has many schools of this type. It is likely that the inclusion of these schools in the sample in 2006 had some effect on driving down the mean scores for Northern Territory student groups, in comparison both to other States and Territories and to the Northern Territory means reported in 2003.

Structural differences

Structural differences include the length of time that students have spent in formal schooling by the time they are in Year 6 and their age at the time of the Assessment.

Table 3.1 Distributions of ages of students in the sample by State and Territory

State/Territory	Average age at time of testing	Average time at school
ACT	12 yrs 4 m	7 years 1 m
NSW	12 yrs 4 m	7 years 1 m
NT	12 yrs 4 m	7 years 1 m
QLD	11 yrs 4 m	6 years 1 m
SA	12 yrs 4 m	7 years 1 m
TAS	12 yrs 4 m	7 years 2 m
VIC	12 yrs 4 m	7 years 1 m
WA	11 yrs 4 m	6 years 1 m

It can be seen that students in Western Australia and Queensland are on average 12 months younger than students in the other States and Territories and have had 12 months' less formal schooling.

Metric for reporting summary performance

As for 2003, the PISA definition of scientific literacy formed the basis for the assessment domain.

A scientific literacy scale was constructed in 2003, using the Rasch model (see Technical Report for more information). The Rasch analysis produced information about the relative difficulty of items, as well as information about students' abilities. This data was located on a continuum to form the scientific literacy scale, and a national mean was set at 400 with a standard deviation of 100.

In 2006, the test instrument more accurately measured students at the higher end of achievement than did the 2003 test. Results from 2003 were able to be mapped to the 2006 results, enabling interpretation of comparative achievement.

More information about the construction of the 2006 scale and comparisons to 2003 can be found in the Technical Report.

Reading the bar charts

Figure 3.1 is an example of the bar chart used to display the scaled mean scores and distributions for States and Territories.

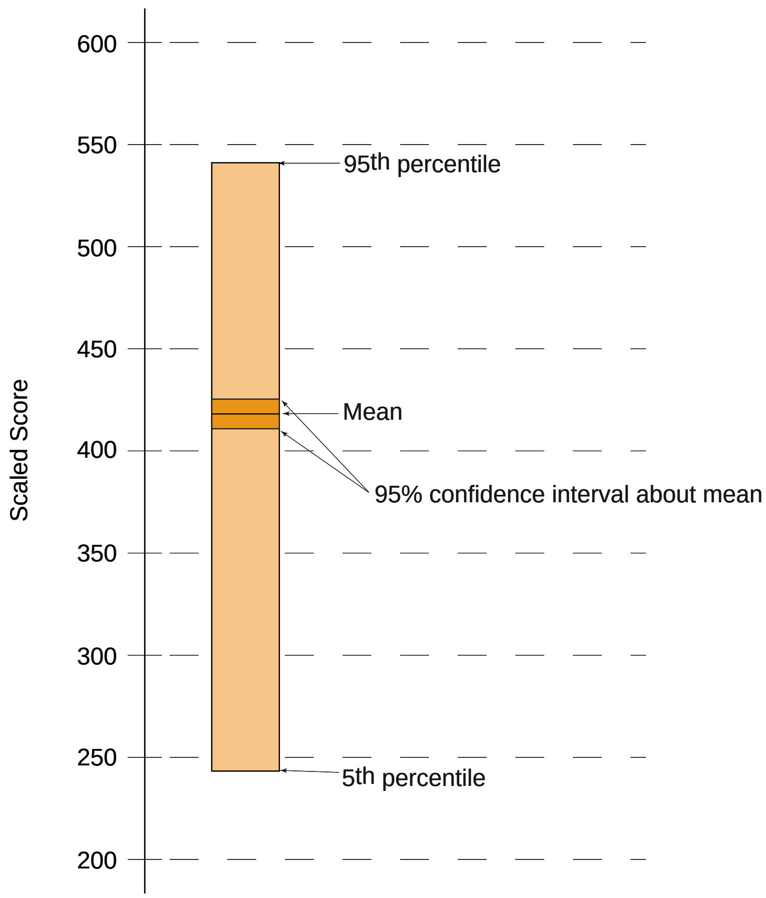
A vertical bar shows the range of student performance.

The highest point on the bar is the 95th percentile, which is the point above which the highest-scoring 5 per cent of the students are located.

The lowest point on the vertical bar is the 5th percentile – the point below which the lowest-scoring 5 per cent of students are located.

Located in the middle region of each bar is a darker gold band that contains a thin horizontal black line. This black line denotes the mean score, while the darker region on either side gives an indication, through the height of the band, of the level of accuracy with which the mean was measured (the smaller the band, the more accurate the measurement).

Figure 3.1 Sample bar chart

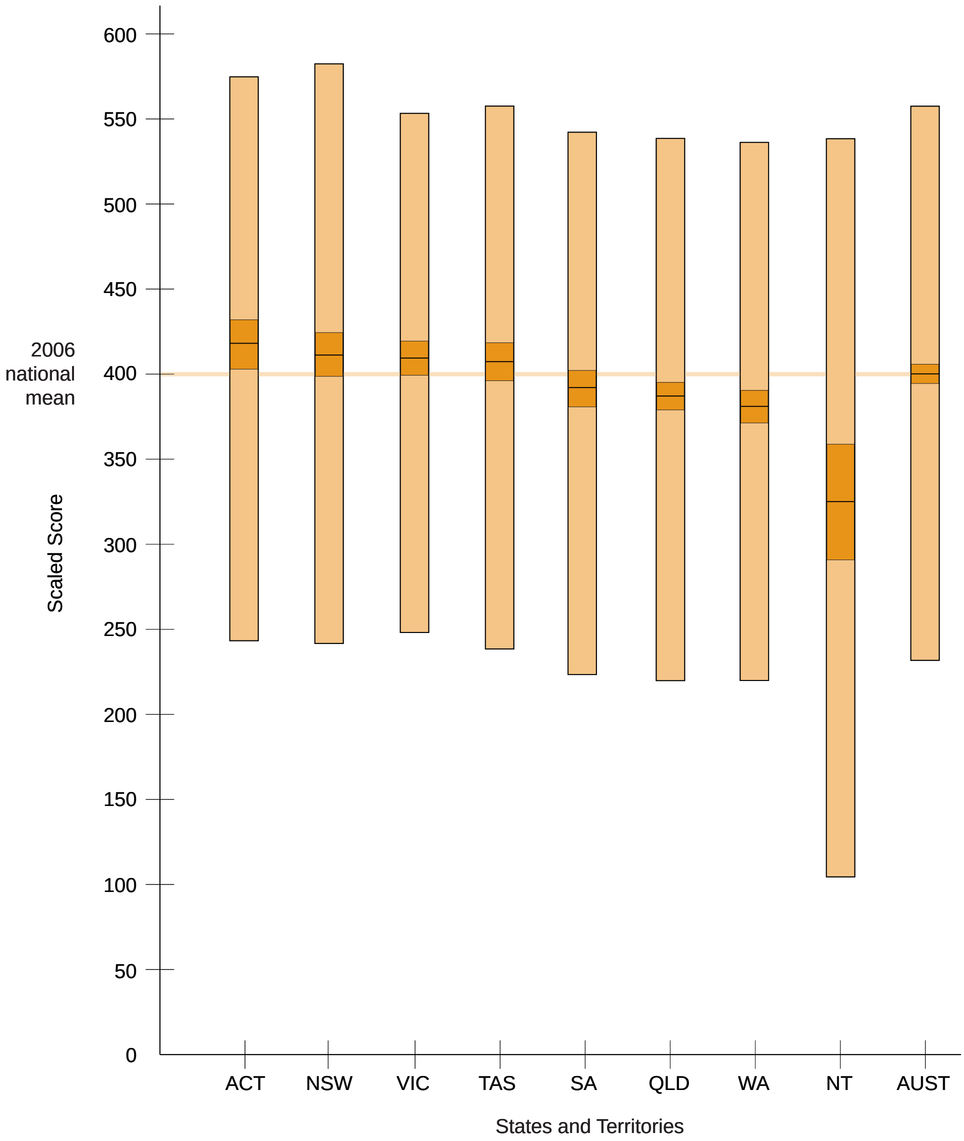


In technical terms, the darker band marks the likely range of the true population score. This is know as the confidence interval (CI).

A comparison of scientific literacy between States and Territories: 2006

Figure 3.2 shows student performance in scientific literacy for each State and Territory.

Figure 3.2 2006 distribution of student performance by State and Territory



	ACT	NSW	VIC	TAS	SA	QLD	WA	NT	AUST
Mean	418	411	408	406	392	387	381	325	400
95% CI	±14.3	±12.5	±10.2	±12.1	±10.0	±8.6	±10.0	±33.7	±5.4

Each State and Territory's result is an estimate of the total population value, inferred from the result obtained from the sample of students tested. Because it is an estimate, it is subject to uncertainty.

If the mean scores were estimated from different samples drawn from the same population of students, the actual results for the mean would vary a little. However, the reader may be confident that the population mean lies between the value obtained and about two standard errors on either side of it.

The estimate of the mean from repeated sampling would be expected to fall within that range for 95 of each 100 samples that were drawn.

The darker bands (confidence intervals) vary in size from one State or Territory to the next. The bars show the spread of scores for each State/Territory, that is, the range of scores achieved by the middle 90 per cent of the population. A large spread of scores suggests that the students' skills and knowledge vary widely; a narrower range indicates more consistency in the students' achievement. Their width is a function of the State or Territory's sample size and the spread of achievement scores on the test. The sample sizes vary in proportion to population, so the States and Territories with the smallest populations have the smallest samples and therefore the widest confidence intervals.

It can be seen that the Northern Territory had the widest spread of scores achieved by the middle 85 per cent of students (those between the 10th and 95th percentiles). Note that the 5th percentile point for NT cannot be calculated as a positive scaled score, so the 10th percentile (for a scaled score of 104) is used as the lowest data point for NT. Table 3.2 shows that the actual difference was 435 points.

New South Wales had the next widest spread of scores between the 5th and 95th percentiles, with a difference of 340 points.

The States and Territories with the least spread of achievement scores for the middle 90 per cent of students were Victoria (305 points), Western Australia (317 points), Queensland and South Australia (both 319 points). The average spread across Australia was approximately 336 points.

It can also be seen from Table 3.2 that the Australian Capital Territory had the highest mean score (418), followed by New South Wales (411) and Victoria (408).

Table 3.2 Percentile scores by State and Territory 2006

State/ Territory	Mean Score	95% confidence interval	Percentile						
			5th	10th	25th	50th	75th	90th	95th
ACT	418	±14.3	243	280	353	423	485	541	575
NSW	411	±12.5	242	279	343	411	479	542	582
VIC	408	±10.2	247	284	347	413	473	523	552
TAS	406	±12.1	237	278	341	410	473	527	557
SA	392	±10.0	223	265	329	396	461	514	542
QLD	387	±8.6	220	261	325	390	454	507	539
WA	381	±10.0	220	254	315	384	448	505	537
NT	325	±33.7	NA	104	242	355	434	499	539
AUST	400	±5.4	232	271	335	403	468	524	558

The Northern Territory has a mean that is statistically significantly different from the mean for the other Territory and States and for Australia as a whole. That is, taking into account the measurement and sampling effect, the mean for students in the Northern Territory is significantly lower than for other States.

The highest-achieving students (those at the 95th percentile for their States and Territories) were from the Australian Capital Territory, New South Wales and Tasmania. Conversely, the lowest-performing students were students from the Northern Territory, Queensland and Western Australia.

Comparisons of achievement

Table 3.3 enables a comparison of States and Territories to be made. The statistical technique compares the results of the States and Territories on a pairwise–comparison basis. Refer to the Technical Report for more information.

Table 3.3 Multiple comparisons of scientific literacy results by State and Territory for 2006 *without* Bonferroni adjustment

			ACT	NSW	VIC	TAS	SA	QLD	WA	NT	AUST
		Mean score	418	411	408	406	392	387	381	325	400
	Mean score	95% CI	±14.3	±12.5	±10.2	±12.1	±10.0	±8.6	±10.0	±33.7	±5.4
ACT	418	±14.3		•	•	•	▲	▲	▲	▲	▲
NSW	411	±12.5	•		•	•	▲	▲	▲	▲	•
VIC	408	±10.2	•	•		•	▲	▲	▲	▲	•
TAS	406	±12.1	•	•	•		•	▲	▲	▲	•
SA	392	±10.0	▼	▼	▼	•		•	•	▲	•
QLD	387	±8.6	▼	▼	▼	▼	•		•	▲	▼
WA	381	±10.0	▼	▼	▼	▼	•	•		▲	▼
NT	325	±33.7	▼	▼	▼	▼	▼	▼	▼		▼
AUST	400	±5.4	▼	•	•	•	•	▲	▲	▲	

▲	Mean performance that is statistically significantly higher than in comparison State/Territory
•	No statistically significant difference from comparison State/Territory
▼	Mean performance that is statistically significantly lower than in comparison State/Territory

By reading across the lines it is possible to draw a comparison between a specific State or Territory and any other State or Territory.

It can be seen in Figure 3.2 and Table 3.2 that the students from the Australian Capital Territory achieved a significantly higher mean score than those from all the other States and Territories except New South Wales, Victoria and Tasmania.

The students from New South Wales and Victoria achieved a significantly higher mean score than those from all the other States and Territories except the Australian Capital Territory and Tasmania.

Students from Tasmania achieved a significantly higher mean score than those from Queensland, Western Australia and the Northern Territory.

The statistics indicate that students from the Northern Territory achieved a significantly lower mean score than students in all other States and Territories.

Table 3.4 Multiple comparisons of scientific literacy results by State and Territory for 2006 *with* Bonferroni adjustment

			ACT	NSW	VIC	TAS	SA	QLD	WA	NT	AUST
		Mean score	418	411	408	406	392	387	381	325	400
	Mean score	95% CI	±14.3	±12.5	±10.2	±12.1	±10.0	±8.6	±10.0	±33.7	±5.4
ACT	418	±14.3		•	•	•	▲	▲	▲	▲	•
NSW	411	±12.5	•		•	•	•	▲	▲	▲	•
VIC	408	±10.2	•	•		•	•	▲	▲	▲	•
TAS	406	±12.1	•	•	•		•	•	▲	▲	•
SA	392	±10.0	▼	•	•	•		•	•	▲	•
QLD	387	±8.6	▼	▼	▼	•	•		•	▲	•
WA	381	±10.0	▼	▼	▼	▼	•	•		▲	▼
NT	325	±33.7	▼	▼	▼	▼	▼	▼	▼		▼
AUST	400	±5.4	•	•	•	•	•	•	▲	▲	

When making multiple comparisons (that is, comparing the performance of one jurisdiction with those of all the others), a more cautious approach is required. Multiple comparison significance tests that limit the probability of mistakenly finding a difference in performance to 5 per cent were applied (Bonferroni adjustment). For further details on Bonferroni adjustment refer to the Technical Report.

In five instances, differences which were statistically significant when analysed without Bonferroni adjustment were no longer significant when analysis was conducted using Bonferroni correction.

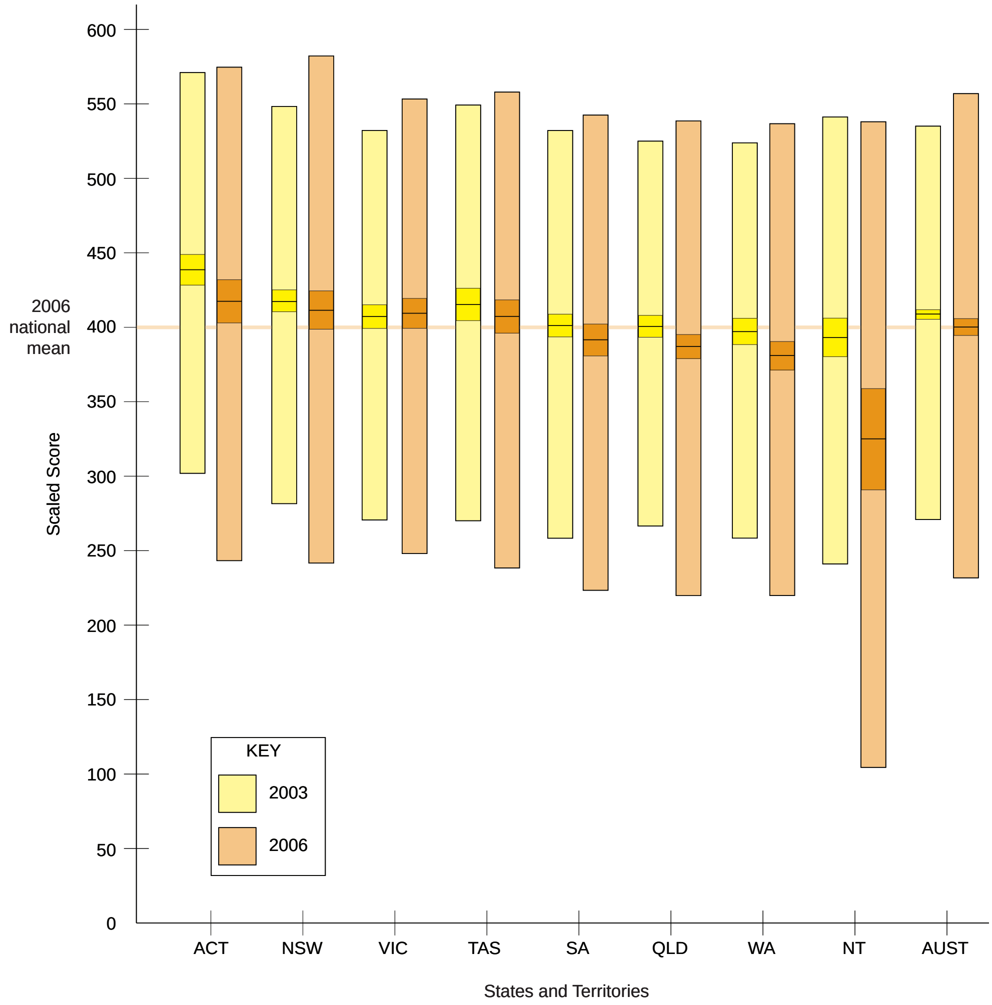
Comparisons of means to 2003

The 2006 National Assessment Program – Science Literacy was the second time the science domain has been assessed in the national assessment programs. The first assessment was carried out in 2003 (Primary Science Assessment Program). Consequently, some comparisons can be made between the 2003 and 2006 findings. Over time, the program will be monitored and trends reported. With only two years' results to date, it is premature to describe trends but legitimate to look at comparisons.

To enable comparisons between 2003 and 2006, the 2006 test instruments contained link items from the 2003 test instruments. These link items formed the basis for equating the 2003 and 2006 tests.

The 2006 results were used to set a scale with a mean score of 400 and a standard deviation of 100 (described in more detail in Chapter 2). The 2003 results were then equated onto the 2006 scale. For details of the equating process, refer to the Technical Report.

Figure 3.3 Comparison of distributions of Year 6 student performance by State and Territory in 2003 and 2006



Note: 2003 results rescaled to 2006.

In technical terms, the darker-coloured bands mark the likely range of the true population score. This is known as the confidence interval (CI).

As previously discussed and illustrated in Figure 3.3, the distribution of student performance between 2003 and 2006 is not statistically significant with the exception of the Northern Territory.

Jurisdiction mean scores

Table 3.5 shows 2003 and 2006 mean scores for each State and Territory, after 2003 results have been equated onto the 2006 scale.

Table 3.5 Comparison of 2003 and 2006 jurisdiction mean scores

State/Territory	Mean score		Change from 2003 to 2006	Statistically significant
	2003 (on 2006 scale)	2006		
ACT	439 (±10.9)	418 (±14.3)	-21	NO
NSW	417 (±7.6)	411 (±12.5)	-6	NO
VIC	407 (±7.9)	408 (±10.2)	1	NO
TAS	415 (±11.3)	406 (±12.1)	-9	NO
SA	401 (±8.0)	392 (±10.0)	-10	NO
QLD	400 (±7.1)	387 (±8.6)	-14	NO
WA	397 (±8.8)	381 (±10.0)	-16	NO
NT	393 (±13.1)	325 (±33.7)	-68	YES
AUST	409 (±3.7)	400 (±5.4)	-9	NO

Note: figures in parentheses refer to 95 per cent confidence intervals. Mean scores have been rounded. 2003 results rescaled to 2006.

While the 2003 mean scores differ from the 2006 mean scores by varying amounts across the jurisdictions, the only jurisdiction where the difference is statistically significant is the Northern Territory. It should be noted that the sampling frame for the Northern Territory in 2006 differs from the sampling frame in 2003, as very remote schools were excluded in 2003 but not in 2006 (see Chapter 2 and Appendix 2 for more detail concerning the changed sampling frame). To further examine trends in the Northern Territory, students' performance was examined separately for remote and provincial districts. Table 3.6 shows Northern Territory mean scores for the two groups separately, in 2003 and 2006.

Table 3.6 Northern Territory student performance by geographic location

Northern Territory	Mean score		Change from 2003 to 2006	Statistically significant
	2003 (on 2006 scale)	2006		
Provincial	395 (±17.6)	383 (±18.9)	-12	NO
Remote	389 (±25.2)	238 (±65.0)	-151	YES
Total	393 (±13.1)	325 (±33.7)	-68	YES

Note: figures in parentheses refer to 95 per cent confidence intervals. Mean scores have been rounded. 2003 results rescaled to 2006.

It can be seen from Table 3.6 that, for provincial districts in the Northern Territory, the mean score in 2006 is not statistically significantly different from the 2003 mean score. However, there is a statistically significant difference in the mean scores for remote areas between 2003 and 2006. This adds further evidence that the more inclusive sampling frame, to include very remote areas in 2006, resulted in a difference in mean performance for remote areas between 2003 and 2006. As a consequence, this difference in performance in remote areas led to a significant difference between 2003 and 2006 in the overall mean scores for the Northern Territory as a jurisdiction.

Further, the differential performance between provincial districts and remote areas in the Northern Territory as shown in the 2006 National Assessment Program – Science Literacy is consistent with the results from other studies. For example, the published ANR benchmark reports for numeracy and literacy also showed marked differences in performance between provincial districts and remote areas in the Northern Territory (http://www.mceetya.edu.au/verve/_resources/2005_Benchmarks.pdf). By contrast, the 2003 Primary Science Assessment Program results showed that the difference between provincial districts and remote areas in the Northern Territory was not statistically significant. This was probably due to the exclusion of very remote schools in 2003.

These results show that overall mean scores must be interpreted with caution. The composition of the samples has an impact on the overall mean achievement levels. A closer examination of the achievement distributions by sub-groups may provide more informative pictures of the performance of students.

Ranking of jurisdictions by mean scores

Table 3.7 shows a jurisdiction-by-jurisdiction comparison of the mean scores in rank order for 2003 and 2006.

Table 3.7 Comparison of 2003 and 2006 jurisdiction rankings by mean scores

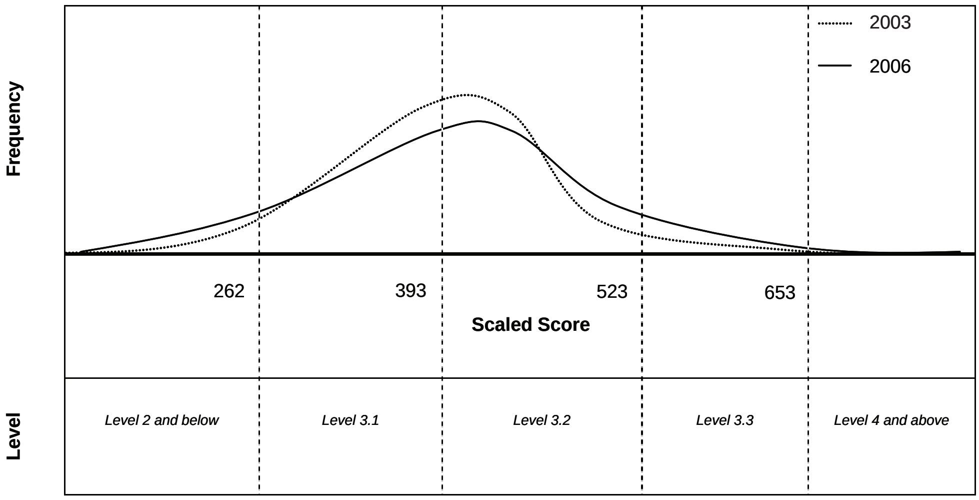
Rank by jurisdiction mean score	2003		2006	
	State/Territory	Mean	State/Territory	Mean
1	ACT	439 (±10.9)	ACT	418 (±14.3)
2	NSW	417 (±7.6)	NSW	411 (±12.5)
3	TAS	415 (±11.3)	VIC	408 (±10.2)
4	VIC	407 (±7.9)	TAS	406 (± 12.1)
5	SA	401 (±8.0)	SA	392 (±10.0)
6	QLD	400 (±7.1)	QLD	387 (±8.6)
7	WA	397 (±8.8)	WA	381 (±10.0)
8	NT	393 (±13.1)	NT	325 (±33.7)

Note: figures in parentheses refer to 95 per cent confidence intervals. Mean scores have been rounded. 2003 results rescaled to 2006.

It can be seen from Table 3.7 that the rank order of jurisdictions by mean scores is largely the same between 2003 and 2006, with the ACT leading the nation, followed by NSW. In 2006, Victoria and Tasmania have changed places when compared with their 2003 ranks, but the other jurisdictions have maintained their rank order. This consistency in rank ordering between 2003 and 2006 provides us with some confidence in interpreting the comparative performance of jurisdictions. The ranking order of jurisdictions in the National Assessment Program – Science Literacy is also consistent with the results from TIMSS 2003 Science Grade 4 cohort for Australia (http://timss.bc.edu/timss2003i/intl_reports.html).

A comparison of the distributions of the 2003 and 2006 student achievement levels is shown in Figure 3.4. Note that 2003 results have been equated onto the 2006 scale. It can be seen that the proficiencies of students are more spread out in 2006 than in 2003. This is probably due to a more inclusive sampling frame in 2006, and also more discriminating test items at the higher-ability end of the scale. See Chapter 4 and the Technical Report for more detail.

Figure 3.4 Comparison of student performance in scientific literacy distributions in 2003 and 2006



Note: 2003 results rescaled to 2006.

Summary

In summary, the trends analysis shows that, at the jurisdiction and national levels, there are no significant differences between 2003 and 2006 in terms of mean scores, except for the Northern Territory where the sampling frame included more small remote schools in 2006 than in 2003. However, there are some differences in the shape of the ability distributions in 2003 and 2006, where the 2006 distribution is more spread out and the 2003 distribution is more centralised, resulting in more students in 2006 at Level 2 and below and at Level 3.3 and above.

Chapter 4
Interpreting the Scientific Literacy Results

Introduction

Chapter 3 showed students' score distributions on the scientific literacy scale. The results can also be referenced directly to the assessment domain, by the items comprising the tests, to reveal the understandings and skills demonstrated by students.

For the purposes of this report the scientific literacy scale has been partitioned into levels called 'Proficiency Levels'.

The next section discusses the establishment of the Proficiency Levels and the cut-off scores for each of the levels.

Establishing Proficiency Levels

One of the main objectives of the National Assessment Program – Science Literacy is to monitor trends in scientific literacy performance over time. One convenient and informative way of doing so is to reference the results to the Proficiency Levels.

Typically, students whose results are located within a particular Proficiency Level are able to demonstrate the understandings and skills associated with that level and possess the understandings and skills of lower Proficiency Levels.

The cut-off points established in the 2003 scale, located onto the 2006 scale, were used to determine the Proficiency Levels for the 2006 test. The cut-off points, which denote the boundaries between Proficiency Levels, were established in 2003 using a combination of experts' knowledge of the skills required to answer each scientific literacy item and information from the analysis of students' responses. Initially, in 2003, three Proficiency Levels, corresponding with Levels 2, 3 and 4 of the assessment domain, were identified.

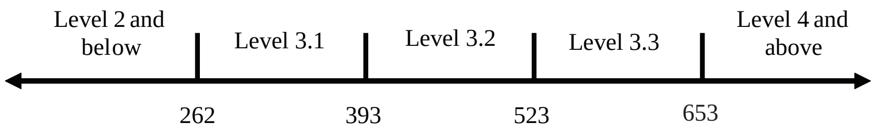
However, as 90 per cent of students' scores fell within Level 3 in 2003, three Proficiency Levels within Level 3 were created, providing five levels for reporting student performance in the Assessment.

The difficulty range spanned by each level was such that students whose scores were at the top of a level had a 65 per cent chance of answering the hardest items in that level correctly, and an 87 per cent chance of answering the easiest items correctly. On average these students would be expected to answer about 76 per cent of the items in that level correctly.

Students who were at the bottom of a level had a 65 per cent chance of answering the easiest items in the level correctly, and a 35 per cent chance of success on the hardest items. On average these students would be expected to answer about 50 per cent of the items in that level correctly.

The cut-off scores for each level are shown in Figure 4.1.

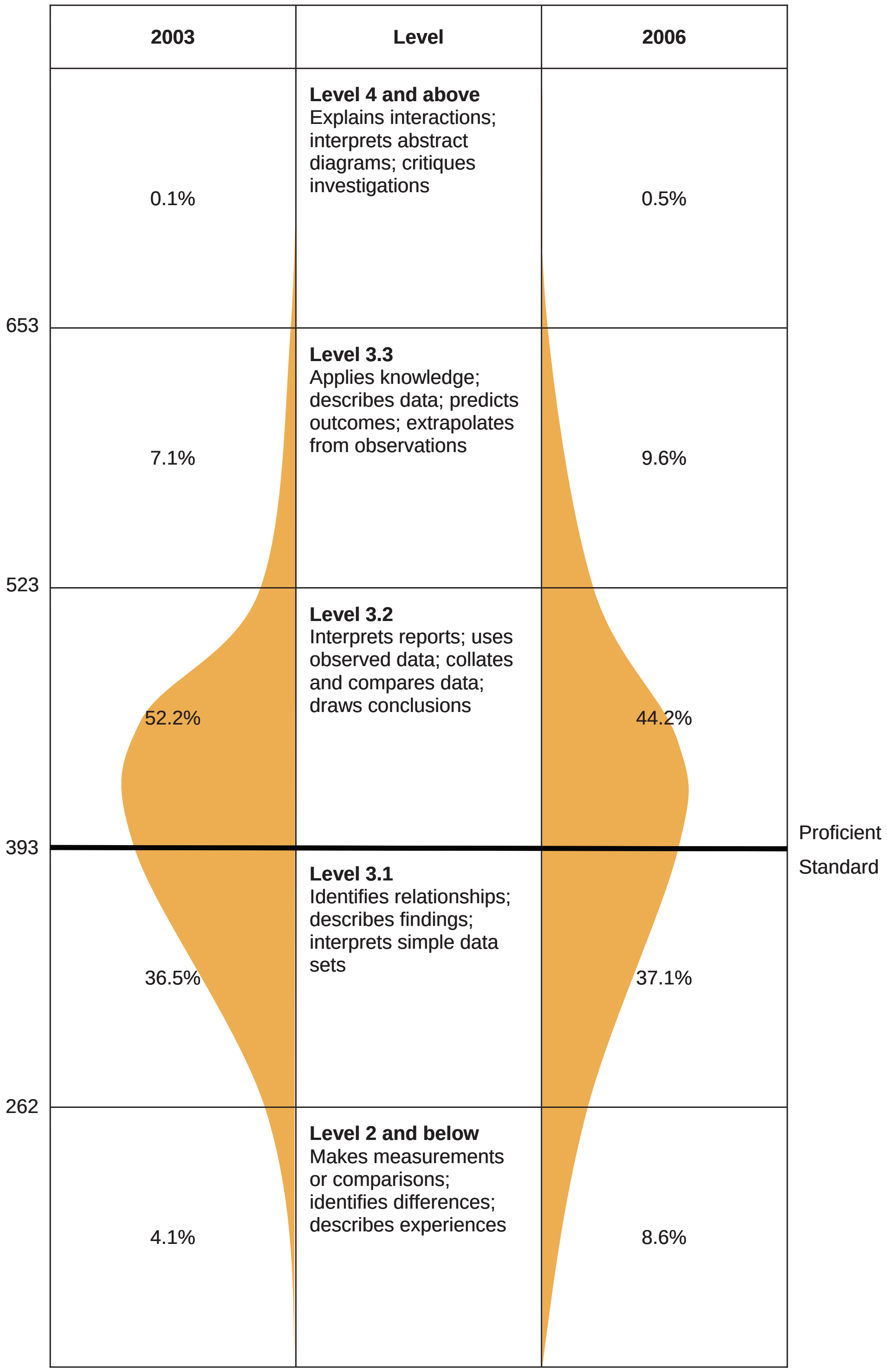
Figure 4.1 Cut-off scores



A score of 653 or more locates students in Proficiency Level 4 and above.

Similarly, scores in the range of 262 to 653 relate to Proficiency Level 3 on the Assessment framework.

Figure 4.2 Distribution of students in Proficiency Levels for 2003 and 2006



Note: 2003 results rescaled to 2006.

Describing Proficiency Levels

Appendix 3 provides the descriptions of knowledge and skills required of students at each Proficiency Level. The descriptions come from the scientific literacy assessment domain presented in Appendix 1, but Level 3 has been divided into sub-levels 3.1, 3.2 and 3.3. Table A3.1 in Appendix 3 also includes descriptors for example items from the 2006 testing at each Proficiency Level.

Below, sample items from the 2006 test are used to illustrate the skill expectations of each Proficiency Level.

Sample items illustrating Proficiency Levels

The following sections provide sample items that illustrate the types of understandings and skills that students at a particular Proficiency Level are likely to display successfully.

At each Proficiency Level, a wide range of items that varied in context, format and difficulty was used to give students the best opportunity to provide evidence of what they knew and could do in relation to scientific literacy.

Only a small number of items have been released in this report – others have been retained for future national science assessments (i.e. held secure).

Those items chosen for presentation here are recorded in Table A3.1 on page 82 as 'illustrative items'. From the descriptors included with the items, it is possible to see that items at the higher Proficiency Levels require more demanding skills and understandings to answer correctly than do those at the lower Levels.

The results for the sample items can be found in Table 4.1 at the end of this chapter. A table of results by State and Territory on the illustrative items is provided in Appendix 4.

Sample items illustrating performance at Proficiency Level 4 and above

Question 3 (see page 35) in the item set 'Native grasslands and the striped legless lizard' illustrates performance at Level 4 and above. This extended response item assesses Strand A, and the concepts are from Life and Living. It assesses students' ability to identify the purpose of each of two interrelated aspects of an experimental design in the context of grasslands as the habitat of the striped legless lizard, a threatened species. This item is located at 701 on the scientific literacy scale.

Figure 4.3 Items illustrating performance at Proficiency Level 4 and above and also Level 3.2

Native grasslands and the striped legless lizard

Native grasslands

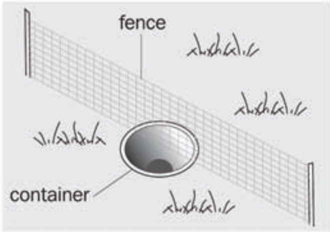
Native grasslands are one of the most threatened natural ecosystems in Australia. Some native grasslands are located on the edges of growing cities.

The striped legless lizard is a rare and threatened species that inhabits native grasslands. It spends most of its time out of sight in grass clumps, in cracks in the ground, or under rocks.

Q1 Explain what humans have done that has destroyed native grasslands.

Studying the striped legless lizard

A scientist studied an area of native grassland to investigate threatened striped legless lizards.



Pitfall trap

The scientist used a pitfall trap which is a smooth-sided container. It is buried so that its top is level with the ground. A small mesh fence directs animals into the trap. Invertebrates such as spiders, crickets, slaters, moth larvae and centipedes are trapped when they fall into the pit.

The scientist also collected samples of the droppings of two legless lizards.

The droppings showed that the striped legless lizard ate kangaroo grass, black crickets, spiders and moth larvae.

Q2 Scientists use diagrams called food chains to show what food is eaten by animals.

A food chain can be shown like this:

> spider ➡ striped legless lizard
>
> The arrow means 'is eaten by'.
>
> This means spiders are eaten by striped legless lizards.

Which one of the following shows another food chain for the striped legless lizard?

- ◯ kangaroo grass ➡ black-winged hawk ➡ striped legless lizard ➡ black cricket
- ◯ kangaroo grass ➡ black cricket ➡ striped legless lizard ➡ black-winged hawk
- ◯ kangaroo grass ➡ black cricket ➡ black-winged hawk ➡ striped legless lizard
- ◯ kangaroo grass ➡ striped legless lizard ➡ black cricket ➡ black-winged hawk

Q3 What was the purpose of collecting information from both the pitfall traps and the lizard's droppings?

a Pitfall traps

____

b Lizard's droppings

____

Question 3 required students to demonstrate both an understanding that the information collected from the pitfall trap indicates the types of invertebrates which inhabit the area in general, and that the collection of lizard droppings provides information about the food sources specific to the striped legless lizard. At a more general level, students who can complete items requiring the same level of scientific literacy as this item would be able to identify the questions being investigated when provided with an experimental design involving multiple variables.

Questions 1 and 2 of the item set 'Native grasslands and the striped legless lizard' are illustrative of Proficiency Level 3.2 (see further below).

A second example illustrative of Level 4 and above assesses Strand B. This item is from a practical activity undertaken by groups of three students, with each student answering the associated questions in the booklet individually. The practical activity was situated within the context of Earth and Beyond and required the students to investigate gravity effects. Students constructed a simple pendulum and then determined the effects of both changing the mass of the pendulum bob and changing the length of the pendulum string on the number of swings of the pendulum in a given time.

Figure 4.4 Item illustrating performance at Proficiency Level 4 and above

Georgia's experiment

Georgia conducted an experiment using 5 different lengths of string. The mass of the bob and the release position stayed the same.

Here are her results.

Length of string (cm)	Average time for 10 swings (seconds)
20	9
40	13
60	16
80	18
100	20

Q8 What do Georgia's results show?

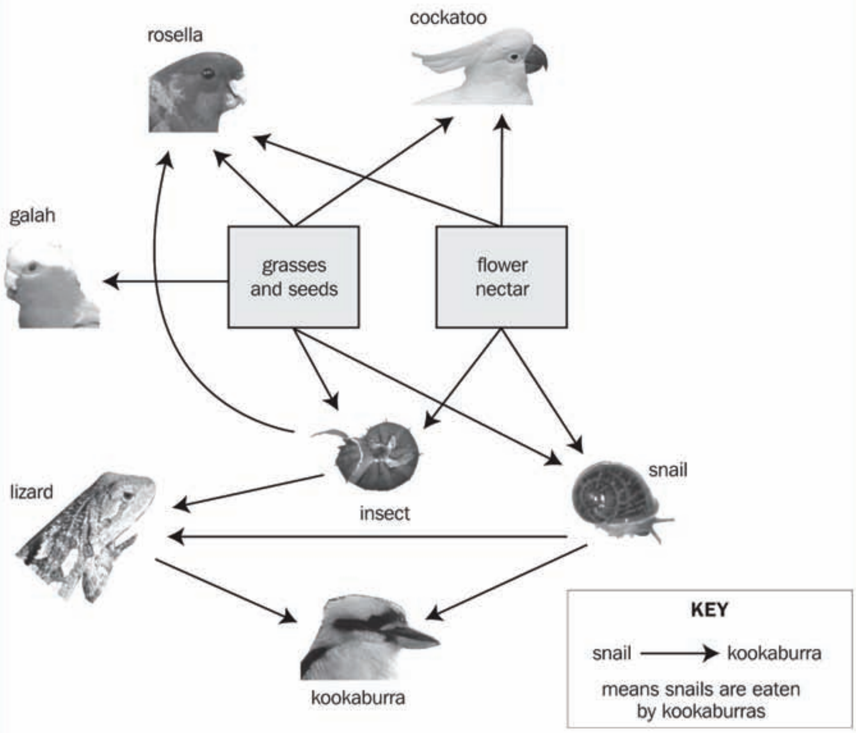
This item is located at 695 on the scientific literacy scale. For this item, students were presented with second-hand data and were required to draw a conclusion that summarised the patterns in the data. This extended-response item required a response which indicated that the trend is for the swing time to increase as the string is lengthened. At a more general level, students who can complete items requiring the same level of scientific literacy as this item would be able to draw a conclusion that is consistent with data provided (in tabular or graphical form) and that summarises the patterns in the data in the form of a rule.

A third example of Level 4 and above assesses Strand C. This item is located at 687 on the scientific literacy scale. This question addresses concepts from Life and Living, specifically, competition for food sources. Students must interpret an abstract representation (a food web) illustrating the interactions between organisms in an Australian bush habitat.

Figure 4.5 Item illustrating performance at Proficiency Level 4 and above

Food web of native animals

The food web shows the flow of energy as one living thing is eaten by another living thing.



Q2 Competition with other animals for food can threaten the survival of an animal.

Cross the boxes for all the animals in the food web that might threaten the survival of the rosella.

☐ cockatoo ☐ galah ☐ kookaburra ☐ lizard

Students needed to consider the food sources of the rosella (grasses and seeds, flower nectar and insects) and then identify any animals that compete directly with the rosella for these same food sources. Students should have selected cockatoo, galah and lizard. The cockatoo feeds on flower nectar and grasses and seeds, which are food sources for the rosella. The galah's only food source is grasses and seeds, again a food source for the rosella. Lizards feed on snails and insects. Insects are another food source for the rosella. Therefore cockatoos, galahs and lizards threaten the survival of the rosella. At a more general level, students who can complete items requiring the same level of scientific literacy as this item would be able to explain interactions in terms of a non-observable or abstract concept.

Sample items illustrating performance at Proficiency Level 3.3

Question 5 of the 'Gravity effects' practical task assesses Domain A and is illustrative of Level 3.3. This item is located at 624 on the scientific literacy scale.

The students completed 'Experiment 2: Checking the effect of changing the length of the string' (reproduced in Figure 4.6), collecting and recording their data in Table 2 of Figure 4.6.

Figure 4.6 Stimulus for an item illustrating performance at Proficiency Level 3.3

Experiment 2:
Checking the effect of changing the length of the string

You will collect data to test this idea (hypothesis):

'The length of a string does not change the number of swings in a given time.'

You will release the two-clip bob using two different string lengths.

1 **Using the longer (40 cm) length of string**, hold the pendulum up, with the string stretched out, so it is level with the pencil.

2 Release the bob and record how many times the pendulum swings in 10 seconds. One student says 'go' at the start of 10 seconds and 'stop' at the end.

3 Repeat this two more times, so that you have three results for the long piece of string.

4 Record all of these results in Table 2.

5 **Now change to the shorter (20 cm) length of string** and repeat the experiment releasing the two-clip bob.

6 Record your results in Table 2.

Table 2: Changing the **length** of the string

Number of swings in 10 seconds (two-clip bob)		
Trial	**40 cm piece of string**	**20 cm piece of string**
1		
2		
3		

Students were then asked to respond to two extended-response items, Questions 4 and 5. Question 4 (shown on page 43) is included as an illustrative item of Proficiency Level 3.1 and is discussed further below. In response to Question 5, students were required to indicate an awareness of the need for fair testing where, specifically, one variable (the mass of the bob) must remain unchanged while the second variable (the length of the string) is changed.

Figure 4.7 Item illustrating performance at Proficiency Level 3.3

> Experiment 2:
> **Checking the effect of changing the length of the string**
>
> **Q5** In this experiment you used two different lengths of string.
> Why were you instructed to use the two-clip bob each time?
>
> ____
>
> ____

For example, students responded that, because otherwise you 'wouldn't know if it was the length of string or the mass', or that the same clip was used so that 'the weight of the two strings [pendulums] would be the same and the weight would therefore not matter'. At a more general level, students who can complete items requiring the same level of scientific literacy as this item would be able to demonstrate an awareness of the principles of conducting an experiment and controlling variables.

An item from the item set 'The effect of temperature on animal survival' is also illustrative of Level 3.3; this is shown in Figure 4.8 on page 40. Question 2 is a multiple-choice item assessing Strand B. This item is located at 556 on the scientific literacy scale and addresses concepts from Life and Living.

Figure 4.8 Item illustrating performance at Proficiency Level 3.3

The effect of temperature on animal survival

Some animals keep their body temperature constant. The body temperature of other animals changes as the temperature in their environment changes.

A scientist measured the body temperature of some animals in a hot and a cold environment. The results are summarised in this table.

Table 1

Name of animal	Type of animal	Animal body temperature (°C)	
		When placed in a **hot** environment (temperature = 35 °C)	When placed in a **cold** environment (temperature = 8 °C)
platypus	mammal	32	32
shark	fish	27	12
bat	mammal	33	33
snake	reptile	30	9
pigeon	bird	40	40
lizard	reptile	32	12

Q2 The scientist then placed the lizard and the pigeon in an environment at 20 °C.

Which body temperatures are they most likely to have after a few days?

- ◯ lizard 12 °C and pigeon 22 °C
- ◯ lizard 22 °C and pigeon 22 °C
- ◯ lizard 32 °C and pigeon 40 °C
- ◯ lizard 22 °C and pigeon 40 °C

Students were required to identify the relationship (rule) between internal body temperature and external environmental temperature for two specific animals: the lizard and the pigeon. Students had to identify the lizard as having a changing body temperature and the pigeon as having a constant body temperature, and then predict the body temperatures of each given an external temperature of 20 °C. At a more general level, students who can complete items requiring the same level of scientific literacy as this item would be able to interpolate or extrapolate from an observed pattern to predict or describe an expected outcome or event.

A third example of Level 3.3 assesses Strand C, and is an item from the item set ‘Musical instruments’. This item is located at 565 on the scientific literacy scale and addresses concepts from Energy and Change.

Figure 4.9 Item illustrating performance at Proficiency Level 3.3

Musical instruments

Objects produce sound when they vibrate, for example, when a guitar string is plucked, it makes a sound.

The frequency of the sound produced is the number of times something vibrates to and fro in one second.

Pitch is the highness or lowness of a note. The pitch of a sound rises as its frequency increases.

A high pitched sound is caused by something vibrating at a high frequency. A low pitched sound is caused by something vibrating at a low frequency.

Q3 The photograph shows a small drum.

It is possible to change the pitch of the sound from the drum by tightening or loosening the skin.

Explain how this would change the pitch.

In response to this extended-response item, students were required to explain how to change the pitch of a small drum (an example of a cause-and-effect relationship). Appropriate student responses required application of the science knowledge about vibrations provided in the stimulus. Statements such as: 'As the drum skin is loosened, it vibrates more slowly producing a lower pitch' and 'As the drum skin tightened, it vibrates quicker producing a higher pitch' were acceptable responses. At a more general level, students who can complete items requiring the same level of scientific literacy as this item would be able to apply knowledge of a relationship to explain a reported phenomenon.

Sample items illustrating performance at Proficiency Level 3.2

Questions 1 and 2 of the item set 'Native grasslands and the striped legless lizard' are illustrative of Proficiency Level 3.2 and both address concepts from Life and Living. Question 1 (see Figure 4.3 on page 33) asked students to explain how humans have impacted on the native grasslands. This extended-response item assesses Strand C and is located at 416 on the scientific literacy scale. Students were required to provide a plausible reason which reflects an understanding that human activities have negatively impacted on the habitat of the striped legless lizard (a cause-and-effect relationship). Appropriate responses included those that indicated that native grasslands were being used to grow crops or used as sites for property development (thus destroying the native habitat). Other acceptable responses referred to the fact that some activities may have resulted in pollution that affects the soil and kills the native grasses or that fires may have been intentionally lit to clear the land. At a more general level, students who can complete items requiring the same level of scientific literacy as this item would be able to interpret contextualised information and provide a reason in terms of cause and effect, drawing on science knowledge.

Question 2 (see Figure 4.3 continued on page 34) is a multiple choice item assessing Strand B, and is located at 458 on the scientific literacy scale. Students were required to interpret the results from a pitfall trap survey to identify a number of possible feeding relationships, and then select a food chain for the striped legless lizard. At a more general level, students who can complete items requiring the same level of scientific literacy as this item would be able to interpret data and identify patterns in – and/or relationships between – elements of the data.

Sample items illustrating performance at Proficiency Level 3.1

Question 4 of the 'Gravity effects' practical task assesses Strand A and is illustrative of Level 3.1. This item is located at 373 on the scientific literacy scale.

The students completed 'Experiment 2: Checking the effect of changing the length of the string' (reproduced in Figure 4.6 on page 38), collecting and recording their data in Table 2 of Figure 4.6.

Figure 4.10 Item illustrating performance at Proficiency Level 3.1

Experiment 2:
Checking the effect of changing the length of the string

Q4 Look at your results for the long and the short lengths of string in Table 2: *Changing the* ***length*** *of the string.*

Write a sentence to describe what you observed.

____

____

Students were required to respond to this extended-response item by describing their observations, based on the data collected previously. They were required to focus on one aspect of the data: the number of swings of the pendulum for two string lengths (with the mass of the bob unchanged). Student responses were also required to be consistent with the data they had collected. At a more general level, students who can complete items requiring the same level of scientific literacy as this item would be able to make simple standard measurements and records of data as descriptions.

Questions 2 and 4 of the item set 'Properties of plastics' are illustrative of Proficiency Level 3.1 and address concepts from Natural and Processed Materials. This item set also included an illustrative item at Level 2 and below (Question 1) and is discussed further on page 45.

Figure 4.11 Items illustrating performance at Proficiency Level 3.1 and Level 2

Properties of plastics

Some plastics can be recycled. They can be formed into many different products.

The table lists:

- properties of some plastics
- products that are made from these plastics when they are first formed
- products that are made when these plastics are recycled.

Type of plastic	Properties			Products made from these plastics	
		Resistance to		when first formed	when recycled
	Clarity	solvents	heat		
PET	clear	good	poor	soft drink bottles, food containers	soft drink bottles, detergent bottles
HDPE	not clear	poor	good	milk containers, shopping bags	flower pots, recycle bins
PVC	clear	good	fair	clear food wrappings	floor mats, drain pipes
PP	not clear	good	good	sauce bottles, yogurt tubs	DVD cases, paint buckets

Note: A solvent is a substance, usually a liquid, capable of dissolving another substance.

Q1 According to the table, DVD cases are made from recycled

- ○ PET.
- ○ HDPE.
- ○ PVC.
- ○ PP.

Q2 Each type of plastic listed can be

- ○ recycled to make containers which clearly show their contents.
- ○ used to safely store solvents.
- ○ used to contain food when heating it in a microwave oven.
- ○ recycled to make other useful products.

Q4 Give one reason why plastic shopping bags cause problems for marine life.

____

Question 2 of the ‘Properties of plastics’ item set assesses Strand B and is illustrative of Level 3.1. This multiple-choice item is located at 283 on the scientific literacy scale. Students were required to interpret and summarise patterns in tabular data with respect to a number of simple criteria specifically related to common properties of plastics. At a more general level, students who can complete items requiring the same level of scientific literacy as this item would be able to interpret a simple dataset requiring an element of comparison.

Question 4 of the ‘Properties of plastics’ item set assesses Strand C and is also illustrative of Level 3.1. This extended-response item is located at 363 on the scientific literacy scale. Students were required to provide a reason for a cause-and-effect relationship given a familiar context (why shopping bags cause problems for marine life). Acceptable responses included those that indicated that shopping bags choke or suffocate marine life, or that bags trap and entangle marine life. Students were required to indicate how bags caused the effect, so that a response such as ‘It kills them’ was inadequate. At a more general level, students who can complete items requiring the same level of scientific literacy as this item would be able to select an appropriate reason to explain a reported observation related to personal experience.

Sample items illustrating performance at Proficiency Level 2 and below

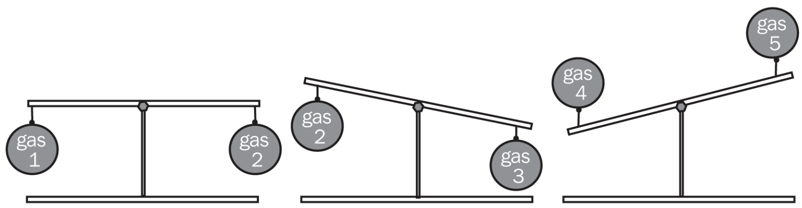
Question 1 of the ‘Properties of plastics’ item set (see Figure 4.11) assesses Strand B and is illustrative of Level 2 and below. This multiple-choice item is located at 203 on the scientific literacy scale. Students were required to compare aspects of a data table providing information on the properties of four types of plastics and the products made from these plastics. The students had to identify the type of plastic that, when recycled, is used to make DVD cases. At a more general level, students who can complete items requiring the same level of scientific literacy as this item would be able to identify simple patterns in data and/or interpret a data set containing some interrelated elements.

Question 1 of the item set ‘States of matter’ (see Figure 4.12 on page 46) assesses Strand C and is also illustrative of Level 2 and below. This multiple-choice item is located at 204 on the scientific literacy scale. This item is also in the context of Natural and Processed Materials.

Figure 4.12 Item illustrating performance at Proficiency Level 2 and below

States of matter

Gas has volume and mass. Gas that is heavier than air sinks. Gas that is lighter than air floats.
The mass of two objects can be compared using balance scales.
Ingrid filled six identical balloons with gases.
She tied the balloons to the ends of balance scales with light string.
Gas 1 was oxygen.



Q1 Ingrid compared the mass of equal volumes of gas 1 and gas 2.
She found that

- ○ gas 1 had more mass than gas 2.
- ○ gas 1 had less mass than gas 2.
- ○ gas 1 had equal mass to gas 2.

Students were required to interpret a diagram representing gas-filled balloons on balance scales to identify the relationship between two gases. The students had to choose whether one gas had more, less or equal mass compared with a second gas. The context was familiar both in terms of the balloons and in the use of balance scales. At a more general level, students who can complete items requiring the same level of scientific literacy as this item would be able to describe or make a choice for a situation based on first-hand concrete experience, requiring the application of limited knowledge.

Table 4.1 summarises the results for the sample items. A breakdown by State and Territory for these items can be found in Appendix 4. This table has been ordered by scaled score.

Table 4.1 Summary of results for sample items 2006

Page	Figure	Unit	Question	% correct	Level	Strand	Scaled score
32-35	Figure 4.3	Native grasslands and the striped legless lizard	3	12.6	≥ 4	A	701
36	Figure 4.4	Gravity effects	8	12.9	≥ 4	B	695
36-37	Figure 4.5	Food web of native animals	2	13.7	≥ 4	C	687
38-39	Figure 4.7	Gravity effects	5	21.2	3.3	A	624
40-41	Figure 4.9	Musical instruments	3	31.3	3.3	C	565
39-40	Figure 4.8	The effect of temperature on animal survival	2	33.1	3.3	B	556
33-35 + 42	Figure 4.3	Native grasslands and the striped legless lizard	2	52.1	3.2	B	458
33-35 + 42	Figure 4.3	Native grasslands and the striped legless lizard	1	60.4	3.2	C	416
42-43	Figure 4.10	Gravity effects	4	67.6	3.1	A	373
43-45	Figure 4.11	Properties of plastics	4	69.5	3.1	C	363
43-45	Figure 4.11	Properties of plastics	2	81.5	3.1	B	283
45-46	Figure 4.12	States of matter	1	90.3	≤ 2	C	204
44-45	Figure 4.11	Properties of plastics	1	89.7	≤ 2	B	203

Chapter 5
Distribution of students within Proficiency Levels for 2006 with comparisons to 2003

Introduction

In 2003 the National Assessment Program – Science Literacy determined that student achievement would be reported against three broad levels of achievement, with Level 3 being further segmented into three sub-levels represented by 3.1, 3.2 and 3.3. The Proficiency Level was situated at the boundary between Level 3.1 and 3.2.

The 2006 distributions of students within Proficiency Levels are shown in Table 5.1.

Student performance by Proficiency Level

At the national level, approximately 8.6 per cent of students performed at Proficiency Level 2 and below. In 2003 the figure was 4.1 per cent. The assessment was constructed with the expectation that most Year 6 students would demonstrate the understandings and skills of Proficiency Level 3.

Approximately 54 per cent of students were proficient at Level 3.2 and above. In 2003, 59 per cent of students were proficient at Level 3.2 and above but this difference is not statistically significant.

Table 5.1 2006 percentage of students in Proficiency Levels by State and Territory

State/ Territory	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above
NSW	7.4 (±2.0)	35.2 (±3.5)	43.9 (±3.6)	12.3 (±3.1)	1.2 (±1.2)
VIC	6.5 (±2.3)	35.2 (±3.9)	48.5 (±4.1)	9.6 (±2.2)	0.2 (±0.4)
QLD	10.2 (±2.5)	40.6 (±2.9)	42.0 (±3.6)	7.0 (±1.7)	0.2 (±0.2)
SA	9.6 (±2.2)	38.7 (±3.7)	43.6 (±3.9)	7.9 (±2.3)	0.1 (±0.2)
WA	11.5 (±2.6)	42.0 (±3.7)	39.6 (±4.0)	6.8 (±2.3)	0.2 (±0.2)
TAS	7.6 (±2.3)	34.9 (±4.4)	46.7 (±4.7)	10.4 (±3.0)	0.3 (±0.4)
NT	28.6 (±7.5)	33.0 (±5.5)	31.6 (±5.4)	6.7 (±2.8)	0.2 (±0.4)
ACT	7.3 (±2.5)	30.7 (±4.8)	47.9 (±4.8)	13.5 (±4.0)	0.6 (±1.1)
AUST	8.6 (±1.1)	37.1 (±1.7)	44.2 (±1.8)	9.6 (±1.2)	0.5 (±0.4)

Note: figures in parentheses refer to 95 per cent confidence intervals.

Table 5.1 shows that NSW had the largest proportion of students working at Proficiency Level 4 or above (1.2%), followed by the ACT (0.6%). But, taking the confidence intervals into account, these differences are not significant. Table 5.2 shows that in 2003 the ACT had the highest proportion at this level (0.3%), with NSW and TAS following (0.2%) but taking the confidence intervals into account, these differences are also not significant.

The Northern Territory had approximately 28.6 per cent of students working at Proficiency Level 2 or below (7.3% in 2003), and approximately 38.5 per cent working at Level 3.2 and above (50.8% in 2003). The shifts in results for the Northern Territory from 2003 to 2006 may be due to the more inclusive sampling design in 2006 (see Chapter 2).

Table 5.2 2003 percentage of students in Proficiency Levels by State and Territory

State/ Territory	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above
NSW	3.2 (±1.3)	33.5 (±3.7)	54.0 (±3.4)	9.0 (±2.2)	0.2 (±0.3)
VIC	3.9 (±1.5)	36.5 (±4.3)	53.3 (±3.9)	6.3 (±1.6)	0.1 (±0.2)
QLD	4.6 (±1.4)	39.9 (±3.4)	50.1 (±3.5)	5.3 (±1.4)	0.0 (±0.1)
SA	5.5 (±2.2)	38.4 (±3.8)	49.7 (±4.0)	6.4 (±1.7)	0.1 (±0.2)
WA	5.4 (±2.2)	40.3 (±3.9)	49.0 (±4.1)	5.2 (±1.5)	0.0 (±0.1)
TAS	4.2 (±2.0)	33.5 (±5.4)	52.9 (±6.0)	9.3 (±4.2)	0.2 (±0.4)
NT	7.3 (±3.2)	41.8 (±6.3)	43.1 (±6.5)	7.7 (±4.4)	0.0 (±0.0)
ACT	1.9 (±1.5)	26.4 (±5.0)	56.8 (±5.9)	14.6 (±4.9)	0.3 (±0.7)
AUST	4.1 (±0.7)	36.5 (±1.7)	52.2 (±1.7)	7.1 (±0.9)	0.1 (±0.1)

Note: figures in parentheses refer to 95 per cent confidence intervals. 2003 results rescaled to 2006.

Table 5.3 Comparison of 2003 and 2006 jurisdiction rankings by percentages of students achieving the proficient standard

Rank by jurisdiction	2003		2006	
	State/Territory	Proficient standard or above	State/Territory	Proficient standard or above
1	ACT	71.6 (±2.6)	ACT	62.0 (±5.6)
2	NSW	63.0 (±2.0)	VIC	58.3 (±5.0)
3	TAS	62.2 (±2.9)	NSW	57.4 (±4.3)
4	VIC	59.9 (±2.2)	TAS	57.4 (±5.5)
5	SA	56.3 (±2.0)	SA	51.6 (±4.7)
6	QLD	55.5 (±2.0)	QLD	49.2 (±3.8)
7	WA	54.3 (±2.4)	WA	46.6 (±4.7)
8	NT	50.7 (±3.4)	NT	38.4 (±6.5)
	AUST	59.4 (±1.0)	AUST	54.3 (±2.1)

Note: figures in parentheses refer to 95 per cent confidence intervals. 2003 results rescaled to 2006.

Table 5.3 shows a comparison of the percentages of students in each State and Territory who achieved the proficiency standard or above. The only jurisdiction that showed a significant difference was the Northern Territory where the more inclusive sampling design resulted in more students from remote locations being included in the program.

Chapter 6
Sub-group results and comparisons by mean and Proficiency Levels

Introduction

In this chapter, the differences in achievement in terms of mean scores and the distribution of results for gender, Indigenous students and students from diverse geographic locations are considered across the States and Territories.

Gender results by mean

Table 6.1 shows that on the scientific literacy scale the mean score for males was 4 points above the mean score for females. This difference is 3 points lower than in 2003 and is not statistically significant.

Table 6.1 also indicates that the confidence intervals are slightly wider for males than for females, indicating a greater spread in scores for males than for females.

Table 6.1 2006 scientific literacy of males and females by State and Territory

State/Territory	% of males in sample	Mean scores	
		Males	Females
ACT	53.0	413 (±16.6)	423 (±14.4)
NSW	51.1	415 (±15.3)	408 (±11.4)
NT	46.7	339 (±35.2)	313 (±37.4)
QLD	49.8	387 (±11.2)	387 (±9.1)
SA	50.7	388 (±12.3)	395 (±11.7)
TAS	52.0	408 (±15.4)	403 (±10.8)
VIC	51.8	411 (±11.3)	405 (±11.1)
WA	49.6	381 (±12.1)	382 (±11.1)
AUST	50.8	402 (±6.4)	398 (±5.1)

Note: figures in parentheses refer to 95 per cent confidence intervals.

It can be seen from Table 6.1 above that females in the Australian Capital Territory were the highest-performing group (mean 423), followed by males from New South Wales (mean 415). However, only the Northern Territory showed a difference in mean scores between males and females which is statistically significant.

The tendency for males to perform better than females was less consistent in 2006 than in 2003.

Table 6.2 2003 scientific literacy of males and females by State and Territory

State/Territory	% of males in sample	Mean scores	
		Males	Females
ACT	51.2	438 (±12.8)	439 (±18.3)
NSW	49.8	420 (±9.9)	414 (±7.8)
NT	48.6	392 (±18.0)	394 (±13.7)
QLD	51.5	404 (±8.6)	396 (±7.8)
SA	53.0	402 (±9.7)	400 (±9.6)
TAS	49.6	419 (±14.8)	411 (±11.4)
VIC	51.9	410 (±10.3)	404 (±9.5)
WA	51.1	401 (±9.8)	394 (±9.9)
AUST	51.1	412 (±4.7)	405 (±4.0)

Note: figures in parentheses refer to 95 per cent confidence intervals.
2003 results rescaled to 2006.

In 2003 the highest performing group was females from the ACT.

In 2006, the mean score for males was higher than that for females in New South Wales, the Northern Territory, Tasmania and Victoria. The mean score for females was higher than that for males in the Australian Capital Territory, South Australia and Western Australia.

The highest differences were observed in the Northern Territory (mean difference 26), the ACT (10), NSW (7) and Tasmania (5). In Queensland the mean scores were identical for both males and females (387).

Table 6.1 also provides information about the proportion of the sample composed of male students – 50.8 per cent. In 2006 the proportion of the sample composed of male students was highest in the ACT (53%) and lowest in the NT (46.7%). However, the sample design was such that this was counterbalanced by information from other jurisdictions.

Gender results by Proficiency Levels

Table 6.3 shows the distributions of results across the Proficiency Levels for males and females and confirms that there were no significant differences in performance.

Table 6.3 2006 distribution of results by gender

State/ Territory	Gender	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above	At or above prof. std
NSW	Male	7.8 (±2.4)	33.6 (±4.6)	43.2 (±4.8)	13.9 (±4.2)	1.6 (±3.1)	58.6 (±5.3)
	Female	7.1 (±2.5)	36.7 (±4.0)	44.8 (±4.0)	10.6 (±3.2)	0.8 (±1.6)	56.2 (±4.6)
VIC	Male	6.3 (±2.6)	34.9 (±5.3)	48.0 (±4.9)	10.5 (±2.7)	0.4 (±0.8)	58.9 (±5.8)
	Female	6.7 (±3.0)	35.4 (±4.8)	49.1 (±5.1)	8.7 (±2.7)	0.1 (±0.1)	57.8 (±5.6)
QLD	Male	10.9 (±3.2)	39.8 (±4.6)	41.7 (±5.0)	7.5 (±2.7)	0.2 (±0.4)	49.3 (±5.0)
	Female	9.5 (±2.9)	41.3 (±3.7)	42.5 (±4.1)	6.6 (±1.9)	0.1 (±0.2)	49.2 (±4.4)
SA	Male	11.6 (±3.1)	37.0 (±4.5)	43.2 (±5.4)	8.0 (±3.1)	0.2 (±0.3)	51.4 (±5.6)
	Female	7.7 (±2.5)	40.5 (±5.4)	44.0 (±4.7)	7.7 (±3.3)	0.1 (±0.1)	51.8 (±5.8)
WA	Male	12.6 (±3.5)	40.9 (±4.5)	38.7 (±5.0)	7.7 (±2.7)	0.2 (±0.4)	46.5 (±5.8)
	Female	10.4 (±3.1)	43.0 (±5.2)	40.5 (±4.7)	5.9 (±2.7)	0.2 (±0.3)	46.6 (±5.4)
TAS	Male	8.4 (±3.2)	33.0 (±5.5)	45.8 (±5.7)	12.5 (±4.2)	0.4 (±0.7)	58.6 (±6.6)
	Female	6.8 (±2.8)	37.1 (±5.5)	47.7 (±5.8)	8.2 (±3.1)	0.2 (±0.3)	56.1 (±5.9)
NT	Male	26.3 (±9.3)	33.2 (±7.4)	32.4 (±8.3)	7.7 (±4.0)	0.4 (±0.7)	40.5 (±9.1)
	Female	30.7 (±7.6)	32.7 (±6.2)	30.9 (±5.5)	5.8 (±3.0)	0.1 (±0.1)	36.7 (±6.3)
ACT	Male	9.2 (±3.8)	30.6 (±5.6)	46.4 (±5.7)	13.4 (±4.4)	0.5 (±0.9)	60.3 (±6.4)
	Female	5.3 (±2.2)	30.8 (±5.9)	49.5 (±5.1)	13.7 (±4.6)	0.8 (±1.5)	63.9 (±6.3)
AUST	Male	9.0 (±1.4)	36.1 (±2.2)	43.6 (±2.2)	10.6 (±1.7)	0.7 (±0.6)	54.9 (±2.5)
	Female	8.2 (±1.3)	38.2 (±2.1)	44.8 (±2.1)	8.5 (±1.3)	0.3 (±0.3)	53.7 (±2.3)

Note: figures in parentheses refer to 95 per cent confidence intervals.

Table 6.4 2003 distribution of results by gender

State/ Territory	Gender	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above
NSW	Male	3.2 (±1.5)	32.8 (±4.5)	53.6 (±4.2)	10.1 (±3.2)	0.3 (±0.4)
	Female	3.2 (±1.7)	34.8 (±4.3)	54.1 (±4.2)	7.7 (±2.2)	0.1 (±0.3)
VIC	Male	4.3 (±2.1)	34.2 (±4.8)	54.2 (±4.4)	7.3 (±2.8)	0.1 (±0.2)
	Female	4.1 (±2.0)	38.2 (±5.6)	52.7 (±5.3)	5.0 (±2.1)	0.1 (±0.3)
QLD	Male	4.5 (±1.9)	38.5 (±4.4)	50.6 (±4.7)	6.3 (±2.2)	0.0 (±0.1)
	Female	4.7 (±1.7)	41.3 (±4.1)	49.7 (±4.1)	4.2 (±1.5)	0.0 (±0.1)
SA	Male	5.1 (±2.8)	38.0 (±4.3)	50.5 (±4.5)	6.4 (±2.0)	0.0 (±0.1)
	Female	5.6 (±2.5)	38.8 (±5.0)	49.5 (±5.1)	6.1 (±2.4)	0.1 (±0.3)
WA	Male	4.9 (±2.2)	39.6 (±5.0)	49.6 (±5.4)	5.9 (±2.0)	0.0 (±0.2)
	Female	5.9 (±2.7)	41.1 (±4.8)	48.8 (±5.2)	4.1 (±1.6)	0.0 (±0.1)
TAS	Male	4.3 (±2.5)	31.8 (±6.0)	53.8 (±9.0)	9.9 (±7.0)	0.2 (±0.7)
	Female	4.2 (±2.4)	35.2 (±5.7)	52.4 (±5.7)	8.1 (±3.6)	0.1 (±0.3)
NT	Male	7.5 (±4.7)	42.6 (±8.8)	42.8 (±9.8)	7.1 (±5.4)	0.0 (±0.0)
	Female	7.1 (±4.2)	41.5 (±9.1)	43.9 (±9.5)	7.4 (±4.8)	0.0 (±0.0)
ACT	Male	1.5 (±1.5)	26.8 (±7.5)	56.5 (±6.7)	15.0 (±5.6)	0.2 (±0.7)
	Female	2.5 (±2.4)	25.6 (±5.8)	56.0 (±8.5)	15.5 (±8.0)	0.4 (±1.2)
AUST	Male	4.1 (±0.8)	35.3 (±2.2)	52.5 (±2.0)	8.1 (±1.4)	0.1 (±0.2)
	Female	4.2 (±0.9)	37.6 (±2.1)	52.0 (±2.1)	6.1 (±1.0)	0.1 (±0.1)

Note: figures in parentheses refer to 95 per cent confidence intervals. 2003 results rescaled to 2006.

Gender comparison to 2003

For all jurisdictions except the Northern Territory, comparisons between 2003 and 2006 mean scores for male students showed that there was no statistically significant difference. Similarly, at the jurisdiction level, no statistically significant difference was found in the performance of female students between 2003 and 2006. However, at the national level, while the difference between males and females was not statistically significant in either 2003 or 2006, male students' mean score was slightly higher than the mean score of female students in both 2003 and 2006.

Indigenous students

Sampling of schools was undertaken to enable reliable estimates of achievement by Indigenous students to be made at the national level. Indigenous students' results relative to non-Indigenous results are shown in Table 6.5.

Table 6.5 2006 mean scores for Indigenous and non-Indigenous students

Student group	Mean score	95% CI
Indigenous	311	±29.4
Non-Indigenous	402	±5.8

Indigenous students, with a mean score of 311, did not perform as well as non-Indigenous students, with a mean score of 402. The statistics are significant, as they were in 2003 when the Indigenous mean score was 350 and the non-Indigenous mean score was 412.

Table 6.6 2003 mean scores for Indigenous and non-Indigenous students

Student group	Mean score	95% CI
Indigenous	350	±11.3
Non-Indigenous	412	±3.7

Note: 2003 results rescaled to 2006.

Nevertheless, as Table 6.7 shows, 3.2 per cent of Indigenous students achieved Proficiency Level 3.3 or above, which is an increase over the 1.7 per cent achieved at this level by Indigenous students in 2003. Table 6.7 also shows that 25.5 per cent of Indigenous students achieved at or above the proficient standard compared to 54.7 per cent of non-Indigenous students.

Table 6.7 also shows that 31.4 per cent of Indigenous students were working at Level 2 or below, whereas only 8.1 per cent of non-Indigenous students were working at the same level. This percentage is greater than the figure of 15.9 per cent for Indigenous students at this level in 2003.

Table 6.7 2006 percentages of Indigenous and non-Indigenous students at Proficiency Levels

Student group	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above	At or above prof. std
Indigenous	31.4 (±8.1)	43.1 (±7.5)	22.3 (±7.4)	3.1 (±3.9)	0.1 (±0.4)	25.5 (±10.0)
Non-Indigenous	8.1 (±1.1)	37.3 (±1.8)	44.3 (±1.9)	9.8 (±1.4)	0.6 (±0.5)	54.7 (±2.2)
AUST	8.6 (±1.1)	37.1 (±1.7)	44.2 (±1.8)	9.6 (±1.2)	0.5 (±0.4)	54.3 (±2.1)

Note: figures in parentheses refer to 95 per cent confidence intervals.

Percentages of students performing at the proficiency levels in 2003 are shown in Table 6.8.

Table 6.8 2003 percentages of Indigenous and non-Indigenous students at Proficiency Levels

Student group	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above
Indigenous	15.9 (±1.3)	51.6 (±6.3)	30.9 (±6.7)	1.7 (±2.0)	0.0 (±0.0)
Non-Indigenous	3.6 (±0.6)	35.7 (±1.7)	53.3 (±1.7)	7.4 (±0.9)	0.1 (±0.1)
AUST	4.1 (±0.7)	36.5 (±1.7)	52.2 (±1.7)	7.1 (±0.9)	0.1 (±10)

Note: figures in parentheses refer to 95 per cent confidence intervals. 2003 results rescaled to 2006.

School location

The five geographic locations used to derive and report the sampled students were defined and provided by MCEETYA. The classifications used to gather data based on geographic location changed between 2003 and 2006, making any comparison invalid.

Table 6.9 shows the distributions of mean scaled scores with 95 per cent confidence intervals. It shows that differences between four of the locations were not statistically significant. However, the 3 per cent of students in the 'Remote and very remote areas' had significantly lower performances in scientific literacy than students from any other location. Students attending schools in 'Major urban statistical district' achieved the highest mean scaled scores. Similar results were found in 2003.

Table 6.9 2006 mean scores of students by geographic location

MCEETYA geographic location category	% of students	Mean score
Metropolitan zone capital city	58.8	404 (±8.5)
Major urban statistical district	12.4	406 (±11.0)
Provincial city statistical district	9.2	395 (±12.1)
Inner and outer provincial areas	16.6	396 (±8.5)
Remote and very remote areas	3.0	331 (±29.9)
ALL	100.0	400 (±5.4)

Note: figures in parentheses refer to 95 per cent confidence intervals and the percentages of students in geographic location regions are weighted to reflect the population percentages. They are not the percentages of students in the sample.

Table 6.10 2006 distribution of results across Proficiency Levels for the five geographic locations

MCEETYA geographic location category	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above	At or above prof. std
Metropolitan zone capital city	8.3 (±1.5)	36.5 (±2.6)	43.9 (±2.5)	10.6 (±2.0)	0.7 (±0.7)	55.3 (±3.2)
Major urban statistical district	6.5 (±2.0)	37.5 (±4.2)	46.0 (±3.9)	9.5 (±2.7)	0.5 (±0.7)	56.0 (±4.9)
Provincial city statistical district	7.9 (±2.4)	40.5 (±4.9)	43.4 (±5.1)	8.0 (±2.6)	0.2 (±0.6)	51.6 (±6.0)
Inner and outer provincial areas	8.7 (±2.2)	37.0 (±3.5)	46.4 (±3.9)	7.8 (±2.0)	0.1 (±0.2)	54.3 (±4.3)
Remote and very remote areas	26.5 (±9.0)	38.0 (±7.0)	31.1 (±8.3)	4.3 (±3.7)	0.1 (±0.3)	35.5 (±9.2)
ALL	8.6 (±1.1)	37.1 (±1.7)	44.2 (±1.8)	9.6 (±1.2)	0.5 (±0.4)	54.3 (±2.1)

Note: figures in parentheses refer to 95 per cent confidence intervals.

A comparison of mean scores indicates that schools in major urban areas had the highest percentages of students in the top levels.

Tables 6.11 and 6.12 show 2003 percentages of students by geographic location and their distribution across the five Proficiency Levels.

Table 6.11 2003 mean scores of students by geographic location

MCEETYA geographic location category	% of students	Mean score
Metropolitan zone, mainland state capital city	57.4	409 (±5.2)
Metropolitan zone, major urban district	12.5	415 (±10.4)
Provincial zone, inner provincial	11.0	405 (±11.0)
Provincial zone, outer provincial	8.4	409 (±9.8)
Provincial zone, city 25K–50K	3.0	413 (±13.9)
Provincial zone, city 50K–100K	5.9	405 (±13.2)
Remote zone, remote areas	1.8	375 (±20.0)
ALL	100.0	409 (±3.7)

Note: figures in parentheses refer to 95 per cent confidence intervals and the percentages of students in geographic location regions are weighted to reflect the population percentages. They are not the percentages of students in the sample. 2003 results rescaled to 2006.

Table 6.12 2003 distribution of results across Proficiency Levels for the seven geographic locations

MCEETYA geographic location category	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above
Metropolitan zone, mainland state capital city	4.3 (±0.9)	36.4 (±2.3)	52.0 (±2.3)	7.2 (±1.4)	0.1 (±0.2)
Metropolitan zone, major urban district	3.3 (±1.7)	33.9 (±5.6)	54.5 (±4.8)	8.3 (±2.5)	0.1 (±0.2)
Provincial zone, inner provincial	4.4 (±2.4)	37.7 (±4.2)	51.7 (±4.5)	6.0 (±2.3)	0.2 (±0.4)
Provincial zone, outer provincial	3.4 (±2.0)	36.7 (±5.4)	53.4 (±5.5)	6.5 (±2.8)	0.0 (±0.2)
Provincial zone, city 25K–50K	3.7 (±2.7)	34.0 (±8.6)	54.5 (±9.8)	7.6 (±4.5)	0.1 (±0.6)
Provincial zone, city 50K–100K	3.8 (±2.1)	38.6 (±8.0)	50.9 (±6.8)	6.5 (±3.3)	0.1 (±0.3)
Remote zone, remote areas	10.5 (±5.6)	46.5 (±8.3)	38.3 (±9.3)	4.6 (±3.4)	0.0 (±0.1)
ALL	4.1 (±0.7)	36.5 (±1.7)	52.2 (±1.7)	7.1 (±0.9)	0.1 (±0.1)

Note: figures in parentheses refer to 95 per cent confidence intervals. 2003 results rescaled to 2006.

Other student background information

In 2003 the only other student background information collected related to language background. This was collected from information supplied by students on the front covers of the test booklets.

In 2006 BEMU commissioned an online software application from Curriculum Corporation, called the Online Student Registration System (OSRS). School Contact Officers of schools selected for the sample were informed that they were to register their students online or, for a few jurisdictions, that this task had been done centrally. State and Territory Liaison Officers were briefed in providing support to principals to use the site. OSRS was designed to capture information that had previously been provided by students on the test book covers in 2003. Pre-registration meant that test books could be overprinted with individual student details, ensuring that every student received the correct test form and that student details were correct. It should be noted, however, that much data that schools were requested to provide on OSRS proved to be missing. Thus the data was incomplete when supplied for analysis, preventing the inclusion of some demographic variables in the item response model e.g. Language Background other than English (LBOTE).

It was therefore not possible to calculate the proficiency of the LBOTE sub-group nationally because of the omissions and inconsistencies in the data provided, or to make comparisons with English Speaking Background students.

Tables 6.13 and 6.14 show data for 2003 only.

Table 6.13 2003 mean scores of scientific literacy scores by main language spoken at home

Student group	Mean score	95% CI
Home language not English	387	±8.5
Home language English	413	±3.6

Note: 2003 results rescaled to 2006.

Table 6.14 2003 percentage of students at each Proficiency Level on the scientific literacy scale by main language spoken at home

Student group	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above
Home language not English	7.4 (±2.4)	43.1 (±4.0)	44.7 (±4.1)	4.9 (±1.9)	0.0 (±0.0)
Home language English	3.5 (±0.6)	35.3 (±1.8)	53.5 (±1.7)	7.5 (±0.9)	0.1 (±0.1)

Note: 2003 results rescaled to 2006.

Similarly information related to parent occupation or other socio-economic factors could not be reported because data was missing.

References

Australian Government Department of Education, Science and Training (2004) Science Education Assessment Resource http://cms.curriculum.edu.au/sear/

Ball, S. et al. (2000) 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: National Education Performance Monitoring Taskforce.

Bigg, J.B. & Collis, K.F. (1982) Evaluating the quality of learning: The SOLO taxonomy. New York: Academic Press.

Goodrum, D., Hackling, M.W. & Rennie, L. (2001) The status and quality of teaching and learning of science in Australian schools. Canberra: Department of Education, Training and Youth Affairs.

Hackling, M.W., Goodrum, D. & Rennie, L. (2001) The state of science in Australian secondary schools. *Australian Science Teachers Journal*, 47(4), 6–17.

Laugksch, R.C. (2000) Scientific literacy: A conceptual overview. *Science Education*, 84(1), 71–94.

OECD Programme for International Student Assessment (1999) Measuring student knowledge and skills: A new framework for assessment. Paris: OECD.

Copyright Acknowledgements

We would like to thank the following copyright holders who have granted permission to use the texts and graphics cited below.

Item	Page	Acknowledgement
Striped legless lizard	33	© Nick Clemann/Viridans Images
Body temperature table	40	Adapted table from Heinemann Interactive Science 1 by C. (Christine) Watson © Reed International Books, Australia. First published 1998. Permission granted by Pearson Education Australia
Drum	41	© Corbis

Appendix 1
National Year 6 Primary Science Assessment Domain

Assessment Strands: Scientific literacy

The national review of the status and quality of teaching and learning of science in Australian schools (Goodrum, Hackling & Rennie 2001) argued that the broad purpose of science in the compulsory years of schooling is to develop scientific literacy for all students.

Scientific literacy is a high priority for all citizens, helping them to:

- be interested in and understand the world around them
- engage in the discourses of and about science
- be sceptical and questioning of claims made by others about scientific matters
- be able to identify questions, investigate and draw evidence-based conclusions
- make informed decisions about the environment and their own health and wellbeing.

Scientific literacy is important because it contributes to the economic and social wellbeing of the nation and improved decision making at public and personal levels (Laugksch 2000).

PISA focuses on aspects of preparedness for adult life in terms of functional knowledge and skills that allow citizens to participate actively in society. It is argued that scientifically-literate people are 'able to use scientific knowledge and processes not just to understand the natural world but also to participate in decisions that affect it' (OECD 1999, p. 13).

The OECD-PISA defined scientific literacy as:

> *... the capacity to use scientific knowledge, to identify questions (investigate)*[1] *and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity.*
>
> (OECD 1999, p. 60)

This definition has been adopted for the National Assessment Program – Science Literacy in accord with the Ball et al. 2000 report recommendation.

1 Because of the constraints of large-scale testing, PISA was not able to include performance tasks such as conducting investigations. Consequently, its definition of scientific literacy omitted reference to investigating. The word 'investigate' was inserted into the definition for the purposes of the National Science Assessment, as the sample testing methodology to be used allowed for assessments of students' ability to conduct investigations.

Scientific literacy: Progress Map

A scientific literacy progress map was developed based on the construct of scientific literacy and an analysis of State and Territory curriculum and assessment frameworks. The progress map describes the development of scientific literacy across three strands of knowledge which are inclusive of Ball et al.'s concepts and processes and the elements of the OECD–PISA definition.

The five elements of scientific literacy, including concepts and processes used in PISA 2000 (OECD–PISA 1999), include:

- demonstrating understanding of scientific concepts
- recognising scientifically investigable questions
- identifying evidence needed in a scientific investigation
- drawing or evaluating conclusions
- communicating valid conclusions.

These elements have been clustered into three more holistic strands which have been described below. The second and third elements and conducting investigations to collect data are encompassed in Strand A; the fourth and fifth elements and conducting investigations to collect data are included in Strand B; and the first element is included in Strand C.

Strand A: Formulating or identifying investigable questions and hypotheses, planning investigations and collecting evidence.

This process strand includes posing questions or hypotheses for investigation or recognising scientifically investigable questions; planning investigations by identifying variables and devising procedures where variables are controlled; gathering evidence through measurement and observation; and making records of data in the form of descriptions, drawings, tables and graphs using a range of information and communications technologies.

Strand B: Interpreting evidence and drawing conclusions from their own or others' data, critiquing the trustworthiness of evidence and claims made by others, and communicating findings.

This process strand includes identifying, describing and explaining the patterns and relationships between variables in scientific data; drawing conclusions that are evidence-based and related to the questions or hypotheses posed; critiquing the trustworthiness of evidence and claims made by others; and communicating findings using a range of scientific genres and information and communications technologies.

Strand C: Using science understandings for describing and explaining natural phenomena, and for interpreting reports about phenomena.

This conceptual strand includes demonstrating conceptual understandings by being able to describe, explain and make sense of natural phenomena; understand and interpret reports (e.g. TV documentaries, newspaper or magazine articles or conversations) related to scientific matters; and make decisions about scientific matters in students' own lives which may involve some consideration of social, environmental and economic costs and benefits.

Scientific literacy has been described here in three strands to facilitate the interpretation of student responses to assessment tasks. However, authentic tasks should require students to apply concepts and processes together to address problems set in real-world contexts. These tasks may involve ethical decision making about scientific matters in students' own lives and some consideration of social, environmental and economic costs and benefits.

The scientific literacy progress map describes progression in six levels from 1 to 6 in terms of three aspects:

- increasing complexity, from explanations that involve one aspect to several aspects, through to relationships between aspects of a phenomenon
- progression from explanations that refer to and are limited to directly experienced phenomena (concrete) to explanations that go beyond what can be observed directly and involve abstract scientific concepts (abstract)
- progression from descriptions of 'what' happened in terms of objects and events, to explanations of 'how' it happened in terms of processes, to explanations of 'why' it happened in terms of science concepts.

The process strands (Strands A and B) are based on the Western Australian and Victorian assessment profiles, as these most clearly describe these learning outcomes.

The conceptual strand (Strand C) has been abstracted across conceptual strands and makes no reference to particular science concepts or contexts. As the progression in the conceptual strand is based on increasing complexity and abstraction, links have been made to the Structure of Observed Learning Outcomes (SOLO) taxonomy (Biggs & Collins 1982).

The taxonomy was written to describe levels of student responses to assessment tasks. The basic SOLO categories include:

prestructural	no logical response
unistructural	refers to only one aspect
multistructural	refers to several independent aspects
relational	can generalise (describe relationships between aspects) within the given or experienced context
extended abstract	can generalise to situations not experienced.

The three main categories of unistructural, multistructural and relational can also be applied, as cycles of learning, to the four modes of representation:

sensorimotor	the world is understood and represented through motor activity
iconic	the world is represented as internal images
concrete	writing and other symbols are used to represent and describe the experienced world
formal	the world is represented and explained using abstract conceptual systems.

The conceptual strand, Strand C, of the progress map therefore makes links to the SOLO categories of concrete unistructural (level 1), concrete multistructural (level 2), concrete relational (level 3), abstract unistructural (level 4), abstract multistructural (level 5) and abstract relational (level 6).

The SOLO levels of performance should not be confused with Piagetian stages of cognitive development. Biggs and Collins (1982, p. 22) explain that the relationship between Piagetian stages and SOLO levels 'is exactly analogous to that between ability and attainment' and that level of performance depends on quality of instruction, motivation to perform, prior knowledge and familiarity with the context. Consequently, performance for a given individual is highly variable and often sub-optimal.

The agreed proficiency standards serve to further elaborate the progress map. Level 3 is now described as 3.1, 3.2, and 3.3. A 'proficient' standard is a challenging level of performance, with students needing to demonstrate more than minimal or elementary skills.

Table A1.1 Scientific Literacy Progress Map – July 2004 version from DEST Science Education Assessment Resource (SEAR) project

Level	**SOLO taxonomy**	**Strands of scientific literacy**		
		Strand A Formulating or identifying investigable questions and hypotheses, planning investigations and collecting evidence. Process strand: experimental design and data gathering.	**Strand B** Interpreting evidence and drawing conclusions from their own or others' data, critiquing the trustworthiness of evidence and claims made by others, and communicating findings. Process strand: interpreting experimental data.	**Strand C** Using understandings for describing and explaining natural phenomena, and for interpreting reports about phenomena. Conceptual strand: applies conceptual understanding.
6	Abstract relational	Uses scientific knowledge to formulate questions, hypotheses and predictions and to identify the variables to be changed, measured and controlled. Trials and modifies techniques to enhance reliability of data collection.	Selects graph type and scales that display the data effectively. Conclusions are consistent with the data, explain the patterns and relationships in terms of scientific concepts and principles, and relate to the question, hypothesis or prediction. Critiques the trustworthiness of reported data (e.g. adequate control of variables, sample or consistency of measurements, assumptions made in formulating the methodology), and consistency between data and claims.	Explains complex interactions, systems or relationships using several abstract scientific concepts or principles and the relationships between them. SOLO: Abstract relational
5	Abstract multi-structural	Formulates scientific questions or hypotheses for testing and plans experiments in which most variables are controlled. Selects equipment that is appropriate and trials measurement procedure to improve techniques and ensure safety. When provided with an experimental design involving multiple independent variables, can identify the questions being investigated.	Conclusions explain the patterns in the data using science concepts, and are consistent with the data. Makes specific suggestions for improving/extending the existing methodology (e.g. controlling an additional variable, changing an aspect of measurement technique). Interprets/compares data from two or more sources. Critiques reports of investigations noting any major flaw in design or inconsistencies in data.	Explains phenomena, or interprets reports about phenomena, using several abstract scientific concepts. SOLO: Abstract multistructural
4	Abstract unistructural	Formulates scientific questions, identifies the variable to be changed, the variable to be measured and in addition identifies at least one variable to be controlled. Uses repeated trials or replicates. Collects and records data involving two or more variables.	Calculates averages from repeat trials or replicates, plots line graphs where appropriate. Interprets data from line graph or bar graph. Conclusions summarise and explain the patterns in the science data. Able to make general suggestions for improving an investigation (e.g. make more measurements).	Explains interactions, processes or effects that have been experienced or reported, in terms of a non-observable property or abstract science concept. SOLO: Abstract unistructural

3	Concrete relational	Formulates simple scientific questions for testing and makes predictions. Demonstrates awareness of the need for fair testing and appreciates scientific meaning of 'fair testing'. Identifies variable to be changed and/or measured but does not indicate variables to be controlled. Makes simple standard measurements. Records data as tables, diagrams or descriptions.	Displays data as tables or constructs bar graphs when given the variables for each axis. Identifies and summarises patterns in science data in the form of a rule. Recognises the need for improvement to the method. Applies the rule by extrapolating and predicting.	Describes the relationships between individual events (including cause and effect relationships) that have been experienced or reported. Can generalise and apply the rule by predicting future events. SOLO: Concrete relational
2	Concrete multi-structural	Given a question in a familiar context, identifies that one variable/factor is to be changed (but does not necessarily use the term 'variable' to describe the changed variable). Demonstrates intuitive level of awareness of fair testing. Observes and describes or makes non-standard measurements and limited records of data.	Makes comparisons between objects or events observed. Compares aspects of data in a simple supplied table of results. Can complete simple tables and bar graphs given table column headings or prepared graph axes.	Describes changes to, differences between or properties of objects or events that have been experienced or reported. SOLO: Concrete multistructural
1	Concrete unistructural	Responds to the teacher's questions and suggestions, manipulates materials and observes what happens.	Shares observations; tells, acts out or draws what happened. Focuses on one aspect of the data.	Describes (or recognises) one aspect or property of an individual object or event that has been experienced or reported. SOLO: Concrete unistructural

A comparison of the 2003 and 2004 conceptual frameworks shows that the changes are elaborations that serve to clarify the content of the cells of the Map. In particular, the elaborations assist in further describing the progression from student descriptions of 'what' happened, to 'how' it happened (concrete), to explanations of 'why' it happened (abstract).

Major scientific concepts in the National Assessment Program – Science Literacy

A table of the major scientific concepts found most widely in the various State and Territory curriculum documents has been developed to accompany the scientific literacy progress map (see Table A1.2).

These major concepts are broad statements of scientific understandings that Year 6 students would be expected to demonstrate. They provided item writers with a specific context in which to assess scientific literacy. An illustrative list of examples for each of the major concepts provides elaboration of these broad conceptual statements and, in conjunction with the scientific literacy progress map, which describes the typical developmental stages for scientific literacy, was used as a guide for the development of assessment items.

It should be noted that, because the National Assessment Program – Science Literacy test instruments are constructed within the constraints of test length, it will not be feasible to include all the listed concepts in instruments constructed for a specific testing cycle.

Table A1.2 Major scientific concepts in the National Assessment Program – Science Literacy 2006

Major scientific concepts	Examples
Earth and Beyond Earth, sky and people: Our lives depend on air, water and materials from the ground; the ways we live depend on landscape, weather and climate. The changing Earth: The Earth is composed of materials that are altered by forces within and upon its surface. Our place in space: The Earth and life on Earth are part of an immense system called the universe.	Features of weather, soil and sky and effects on me. Changes in weather, weather data, seasons, soil landscape and sky (e.g. moon phases), weathering and erosion, movement of the Sun and shadows, bush fires, land clearing. People use resources from the Earth; need to use them wisely. Rotation of the Earth and night/day, spatial relationships between Sun, Earth and Moon. Planets of our solar system and their characteristics.
Energy and Change Energy and us: Energy is vital to our existence and our quality of life as individuals and as a society. Transferring energy: Interaction and change involve energy transfers; control of energy transfer enables particular changes to be achieved. Energy sources and receivers: Observed change in an object or system is indicated by the form and amount of energy transferred to or from it.	Uses of energy, patterns of energy use and variations with time of day and season. Sources, transfers, carriers and receivers of energy, energy and change. Types of energy, energy of motion – toys and other simple machines – light, sound. Forces as pushes and pulls, magnetic attraction and repulsion.
Life and Living Living together: Organisms in a particular environment are interdependent. Structure and function: Living things can be understood in terms of functional units and systems. Biodiversity, change and continuity: Life on Earth has a history of change and disruption, yet continues generation to generation.	Living vs non-living. Plant vs animal and major groups. Major structures and systems and their functions. Dependence on the environment: Survival needs – food, space and shelter. Change over lifetime, reproductions and lifecycles. Interactions between organisms and interdependence, e.g. simple food chains. Adaptation to physical environment.
Natural and Processed Materials Materials and their uses: The properties of materials determine their uses; properties can be modified. Structure and properties: The substructure of materials determines their behaviour and properties. Reactions and change: Patterns of interaction of materials enable us to understand and control those interactions.	Materials have different properties and uses. The properties of materials can be explained in terms of their visible substructure, such as fibres. Materials can change their state and properties. Solids, liquids and gases.

Appendix 2
Sampling

Sampling results

The target population for National Assessment Program – Science Literacy consisted of all students enrolled in Year 6 in Australian schools in 2006.

The nationwide achieved sample was to be approximately 12 000 students located within approximately 600 schools throughout Australia. The 2006 sample differed from that drawn in 2003 in the following ways:

- the sample frame, by definition, was more closely aligned to the national desired population than the sample frame in 2003, since the 2006 sample frame contained very small and very remote schools that were excluded in 2003
- target sample sizes across the jurisdictions were determined so that the precisions of estimates were as similar across jurisdictions as possible
- ACT, TAS and NT all had smaller sample sizes than other States, but their sample sizes were comparable with or larger than their corresponding sample sizes in 2003
- the target sample sizes for the larger jurisdictions (NSW, VIC, QLD, SA and WA) were reduced in 2006 compared to those of 2003.

The sample design for National Assessment Program – Science Literacy was a two-stage stratified cluster sample. (Stratification involves ordering and grouping schools according to different school characteristics [e.g. state, sector, urban/rural] which helps ensure adequate coverage of all desired school types in the sample.) Stage 1 consisted of selecting schools that had Year 6 students. In this stage, schools were selected with probabilities proportional to their estimated Year 6 enrolment. Stage 2 involved the random selection of an intact Year 6 class from the sampled schools selected in Stage 1.

No school-level exclusions from the supplied sampling frame were made prior to sample selection. However, the inclusion of schools that would previously (in the 2003 method) have been excluded was expected to result in an increased non-response rate for 2006 compared to 2003. Consequently, a slightly inflated sample size was required to deal with this expected increase in non-response rate at the school level, so that the actual achieved number of schools and students in the sample would be adequate.

Table A2.1 shows the number of educational institutions and students in the sampling frame for each jurisdiction.

Table A2.1 Estimated 2006 Year 6 enrolment figures as provided by BEMU

State	Institutions	Students	Student %
ACT	108	4364	2%
NSW	2345	86 961	33%
NT	148	3002	1%
QLD	1378	55 712	21%
SA	618	18 837	7%
TAS	223	6462	2%
VIC	1805	64 405	24%
WA	872	27 673	10%
AUST	7497	267 416	100%

Table A2.2 shows the proportions of large, moderately small and very small schools within each jurisdiction. It can be seen that there are many small schools in each jurisdiction. It was important that an appropriate strategy was utilised to prevent an over-selection of small schools, resulting in a sample size lower than the desired target sample size.

Table A2.2 Proportions of schools by school size and jurisdiction

State/ Territory	School size	No. Schools	% Schools	No. Students	% Students
ACT	Large	69	64	3766	86%
	Moderately small	26	24	515	12%
	Very small	13	12	83	2%
	Total	108	100	4364	100%
NSW	Large	1394	59	76 913	88%
	Moderately small	360	15	6712	8%
	Very small	591	25	3336	4%
	Total	2345	100	86 961	100%
NT	Large	53	36	2256	75%
	Moderately small	21	14	363	12%
	Very small	74	50	382	13%
	Total	148	100	3001	100%
QLD	Large	747	54	49 652	89%
	Moderately small	204	15	3662	7%
	Very small	427	31	2397	4%
	Total	1378	100	55 712	100%
SA	Large	322	52	15 259	81%
	Moderately small	140	23	2580	14%
	Very small	156	25	999	5%
	Total	618	100	18 837	100%
TAS	Large	117	52	5145	80%
	Moderately small	54	24	977	15%
	Very small	52	23	340	5%
	Total	223	100	6462	100%
VIC	Large	1072	59	55 520	86%
	Moderately small	342	19	6464	10%
	Very small	391	22	2421	4%
	Total	1805	100	64 405	100%
WA	Large	470	54	23 523	85%
	Moderately small	144	17	2656	10%
	Very small	258	30	1494	5%
	Total	872	100	27 673	100%

Class selection

One classroom containing Year 6 students was sampled per school. Classrooms generally had equal probabilities of selection. The overall procedure for class selection was as follows:

1. each class in a school was assigned a random number
2. the classes in a school were ordered by the assigned random numbers
3. the first class on each school's ordered list was chosen for the sample.

More detail about the sampling process may be found in the Technical Report.

Sample achieved

The National Assessment Program – Science Literacy specifications set a total achieved sample size at 12 000 students (down from 14 000 in 2003). The total achieved sample size for 2006 was 12,911.

Table A2.3 School participation rates by jurisdiction

State/ Territory	School population	Number of schools sampled	Number of excluded schools	Number of schools that participated	School participation (percentage)
ACT	108	57	0	57	100.0
NSW	2345	92	0	90	97.8
NT	148	49	3	43	87.8
QLD	1378	94	0	94	100.0
SA	618	94	0	93	98.9
TAS	223	64	0	64	100.0
VIC	1805	91	2	88	96.7
WA	872	95	2	92	96.8
AUST	7497	636	7	621	97.6

Seven schools were excluded. Various reasons were given for non-participation by other sampled schools, including several small schools with no Year 6 students in attendance on test day. More detail may be found in the Technical Report.

Table A2.4 provides a breakdown of the sample according to jurisdiction. The target sample is the number of Year 6 students enrolled at the time of testing. The achieved sample is the number of Year 6 students who participated.

Table A2.4 National Assessment Program – Science Literacy target and achieved sample sizes by jurisdiction

State/Territory	Target sample		Achieved sample	
	Students	Per cent	Students	Per cent
ACT	1346	9.5	1271	9.8
NSW	2212	15.6	2039	15.8
NT	867	6.1	740	5.7
QLD	2195	15.5	2016	15.6
SA	2002	14.1	1809	14.0
TAS	1330	9.4	1225	9.5
VIC	2020	14.3	1810	14.0
WA	2184	15.4	2001	15.5
AUST	14 156	100.0	12 911	100.0

Table A2.5 provides a breakdown of the achieved sample in comparison with the number of Year 6 students in each jurisdiction.

Table A2.5 Achieved sample by student participation

State/ Territory	Student population	Number of students in participating schools	Number of students that participated	Within-school exclusions	Within-school exclusions (percentage)	Within-school student participation (percentage)
ACT	4364	1345	1271	4	0.3	94.4
NSW	86 961	2104	2039	10	0.5	94.0
NT	3001	932	740	7	0.8	88.3
QLD	55 712	2116	2016	19	0.9	91.8
SA	18 837	2087	1809	12	0.6	90.9
TAS	6462	1397	1225	10	0.8	92.1
VIC	64 405	2098	1810	13	0.7	90.7
WA	27 673	2093	2001	19	0.9	91.9
AUST	267 414	14 172	12 911	94	0.7	92.0

School-level student exclusions

Within-school exclusions may have occurred for the following reasons:

Table A2.6 Within-school exclusion categories

Functional disability	Student has a moderate to severe permanent physical disability such that he/she cannot perform in the National Assessment Program – Science Literacy testing situation. Functionally-disabled students who could respond to the Assessment were included.
Intellectual disability	Student has a mental or emotional disability and is cognitively delayed such that he/she cannot perform in the National Assessment Program – Science Literacy testing situation. This includes students who are emotionally or mentally unable to follow even the general instructions of the Assessment.
Limited language proficiency	The student is unable to read or speak any of the languages of the Assessment in the country and would be unable to overcome the language barrier in the testing situation. Typically a student who has received less than one year of instruction in the languages of the Assessment may be excluded.
Refusal	Parent requested that student not participate OR student refused.

The numbers of non-participating students are provided in Table A2.7, broken down by jurisdiction and reason for non-participation.

Table A2.7 Student non-participation by jurisdiction

State/ Territory	Non-inclusion code					Total
	Absent	Functional disability	Intellectual disability	Limited language proficiency	Student or parent refusal	
ACT	70	1	1	2	1	75
NSW	162	1	7	2	1	173
NT	112	0	3	10	2	127
QLD	155	1	10	8	5	179
SA	169	1	9	2	12	193
TAS	95	2	6	2	0	105
VIC	191	0	8	5	6	210
WA	164	1	8	10	0	183
AUST	1118	7	52	41	27	1245

Appendix 3
Proficiency Levels, Assessment Strand Descriptors, Illustrative Items and Item Descriptors

Table A3.1 Proficiency Levels, assessment strand descriptors, illustrative items and item descriptors

Proficiency Level (scaled location)	Assessment strand descriptors	Descriptor: a student at this level may display skills like	Illustrative items and item descriptors
Level 4 and above (scale score > 653)	Strand A: Formulates scientific questions, identifies the variable to be changed, the variable to be measured and in addition identifies at least one variable to be controlled. Uses repeated trials or replicates. Collects and records data involving two or more variables.	When provided with an experimental design involving multiple variables can identify the questions being investigated.	Identifies the purpose of each of two interrelated aspects of an experimental design [in the context of grasslands as habitat of threatened species]. Q3 Native grasslands and the striped legless lizard
	Strand B: Calculates averages from repeat trials or replicates, plots line graphs where appropriate. Interprets data from line graph or bar graph. Conclusions summarise and explain the patterns in the science data. Able to make general suggestions for improving an investigation (e.g. make more measurements)	Conclusions summarise and explain the patterns in the data in the form of a rule and are consistent with the data.	Summarises the pattern in second-hand data in the form of a rule: trend is for swing time to increase as string is lengthened [in the context of an experiment using a pendulum to investigate the force of gravity acting on objects]. Q8 Gravity effects practical task
	Strand C: Explains interactions, processes or effects that have been experienced or reported, in terms of a non-observable property or abstract science concept.	Explains interactions that have been observed in terms of an abstract science concept.	Explains interaction between organisms in a food web in terms of an abstract science concept: competition for resources [in context of food web for a bush habitat]. Q2 Food web of native animals

Proficiency Level (scaled location)	Assessment strand descriptors	Descriptor: a student at this level may display skills like	Illustrative items and item descriptors
Level 3.3 (scaled score 523–653)	Strand A: Formulates simple scientific questions for testing and makes predictions. Demonstrates awareness of the need for fair testing and appreciates scientific meaning of 'fair testing'. Identifies variable to be changed and/or measured but does not indicate variables to be controlled. Makes simple standard measurements. Records data as tables, diagrams or descriptions.	Demonstrates an awareness of the principles of conducting an experiment and controlling variables.	Demonstrates awareness of the need for fair testing by keeping a variable controlled (mass of bob) when changing a second variable: length of string [in the context of an experiment using a pendulum to investigate the force of gravity acting on objects]. Q5 Gravity effects practical task
	Demonstrates awareness of the need for fair testing by keeping a variable controlled (mass of bob) when changing a second variable: length of string [in the context of an experiment using a pendulum to investigate the force of gravity acting on objects]. Q5 Gravity effects practical task	Extrapolates from an observed pattern to describe an expected outcome or event.	Applies the rule (relationship between body temperature and external environment) to predict body temperatures [in context of animal body function]. Q2 Effects of temperature on animal survival
	Applies the rule (relationship between body temperature and external environment) to predict body temperatures [in context of animal body function]. Q2 Effects of temperature on animal survival	Applies knowledge of relationship to explain reported phenomenon.	Explains cause and effect relationship: how to change pitch of a drum [in context of a small drum as a type of musical instrument]. Q3 Musical instruments

Proficiency Level (scaled location)	Assessment strand descriptors	Descriptor: a student at this level may display skills like	Illustrative items and item descriptors
Level 3.2 (scaled score 523–393)	Strand A: Formulates simple scientific questions for testing and makes predictions. Demonstrates awareness of the need for fair testing and appreciates scientific meaning of 'fair testing'. Identifies variable to be changed and/or measured but does not indicate variables to be controlled. Makes simple standard measurements. Records data as tables, diagrams or descriptions.	Collates and compares data set of collected information, Gives reason for controlling a single variable.	All items addressing this strand at this level have been held secure.
	Strand B: Displays data as tables or constructs bar graphs when given the variables for each axis. Identifies and summarises patterns in science data in the form of a rule. Recognises the need for improvement to the method. Applies the rule by extrapolating and predicting.	Interprets data and identifies patterns in – and/or the relationships between – elements of the data.	Interprets information provided (sequence of feeding relationships) to identify pattern in science data: selects specific food chain [in the context of grasslands as habitat of threatened species]. Q2 Native grasslands and the striped legless lizard
	Strand C: Describes the relationships between individual events (including cause and effect relationships) that have been experienced or reported. Can generalise and apply the rule by predicting future events.	Interprets information in a contextualised report by application of relevant science knowledge.	Describes cause and effect relationship (human impact on grasslands) [in the context of grasslands as habitat of threatened species]. Q1 Native grasslands and the striped legless lizard

Proficiency Level (scaled location)	Assessment strand descriptors	Descriptor: a student at this level may display skills like	Illustrative items and item descriptors
Level 3.1 (scaled score 393–262)	Strand A: Formulates simple scientific questions for testing and makes predictions. Demonstrates awareness of the need for fair testing and appreciates scientific meaning of 'fair testing'. Identifies variable to be changed and/or measured but does not indicate variables to be controlled. Makes simple standard measurements. Records data as tables, diagrams or descriptions.	Makes simple standard measurements and records data as descriptions.	Focuses on one aspect of the experimental data to indicate a specific result (related to length of pendulum string) that is consistent with the data collected [in the context of an experiment using a pendulum to investigate the force of gravity acting on objects]. Q4 Gravity effects practical task
	Strand B: Displays data as tables or constructs bar graphs when given the variables for each axis. Identifies and summarises patterns in science data in the form of a rule. Recognises the need for improvement to the method. Applies the rule by extrapolating and predicting.	Interprets simple data set requiring an element of comparison.	Summarises patterns in the data to find common properties of plastics [in context of properties of plastics and their use]. Q2 Properties of plastics
	Strand C: Describes the relationships between individual events (including cause and effect relationships) that have been experienced or reported. Can generalise and apply the rule by predicting future events.	Selects appropriate reason to explain reported observation related to personal experience.	Explains cause and effect of presence of waste plastics and their impact on marine life [in context of human impact on marine environment]. Q4 Properties of plastics

Proficiency Level (scaled location)	Assessment strand descriptors	Descriptor: a student at this level may display skills like	Illustrative items and item descriptors
Level 2 and below (scaled score ≤ 262)	Strand A: Given a question in a familiar context, identifies that one variable/factor is to be changed (but does not necessarily use the term 'variable' to describe the changed variable). Demonstrates intuitive level of awareness of fair testing. Observes and describes or makes non-standard measurements and limited records of data.	Makes measurements or comparisons involving information or stimulus in a familiar context.	All items addressing this strand at this level have been held secure.
	Strand B: Makes comparisons between objects or events observed. Compares aspects of data in a simple supplied table of results. Can complete simple tables and bar graphs given table column headings or prepared graph axes.	Identifies simple patterns in the data and/or interprets a data set containing some interrelated elements.	Compares aspects of data in simple supplied table of results containing information about four types of plastic [in context of recycling of plastics]. Q1 Properties of plastics
	Strand C: Describes changes to, differences between or properties of objects or events that have been experienced or reported.	Describes a choice for a situation based on first-hand concrete experience, requiring the application of limited knowledge.	Recognises a common property of two objects: gases of equal mass represented in a diagram [in the context of the properties of gases]. Q1 States of matter

Appendix 4
Sample Item Statistics

Table A4.1 Performance of students from each State/Territory on sample items

Page	Figure	Unit	Question	Level	Strand	Scaled score	Percentage correct							
							NSW	VIC	QLD	SA	WA	TAS	NT	ACT
35	Figure 4.3	Native grasslands and the striped legless lizard	3	≥4	A	701	14.9	12.1	12.3	14.4	11.2	11.6	11.8	13.8
36	Figure 4.4	Gravity effects	8	≥4	B	695	14.8	11.0	11.9	11.6	11.2	14.3	8.4	16.9
37	Figure 4.5	Food web of native animals	2	≥4	C	687	15.6	12.9	9.7	10.7	13.2	16.4	13.6	16.1
39	Figure 4.7	Gravity effects	5	3.3	A	624	23.2	19.3	20.2	18.3	23.1	24.1	17.3	23.6
41	Figure 4.9	Musical instruments	3	3.3	C	565	33.9	28.4	35.8	29.7	30.2	30.0	33.5	33.5
40	Figure 4.8	The effect of temperature on animal survival	2	3.3	B	556	32.5	36.5	31.4	32.3	30.8	32.7	31.5	33.7
35	Figure 4.3	Native grasslands and the striped legless lizard	2	3.2	B	458	48.1	52.5	53.0	49.0	50.8	53.7	50.9	56.8
33	Figure 4.3	Native grasslands and the striped legless lizard	1	3.2	C	416	57.2	63.1	59.1	58.4	60.4	60.7	60.0	67.4
43	Figure 4.10	Gravity effects	4	3.1	A	373	70.8	71.8	63.4	66.5	62.3	72.8	57.2	70.9
44	Figure 4.11	Properties of plastics	4	3.1	C	363	69.2	70.3	72.3	67.6	65.7	66.7	64.4	73.7
44	Figure 4.11	Properties of plastics	2	3.1	B	283	81.5	86.4	79.3	79.5	79.2	84.6	76.7	82.8
46	Figure 4.12	States of matter	1	≤2	C	204	91.9	90.4	90.6	90.3	89.3	90.3	84.5	90.2
44	Figure 4.11	Properties of plastics	1	≤2	B	203	91.3	92.8	89.7	88.3	87.3	91.1	85.3	90.4

Appendix 5
Performance by Geographic Location

The following table summarises data from Table 6.10 (2006 distribution of results across Proficiency Levels for the five geographic locations) into three broad geographic location categories.

Table A5.1 2006 distribution of results across Proficiency Levels for three geographic location categories

Geographic location category	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above	At or above prof. std
Metropolitan and major urban districts	7.9	36.7	44.3	10.4	0.7	55.4
Provincial areas	8.4	38.2	45.3	7.9	0.2	53.4
Remote and very remote areas	26.5	38.0	31.1	4.3	0.1	35.5
ALL	8.6	37.0	44.2	9.6	0.5	54.3