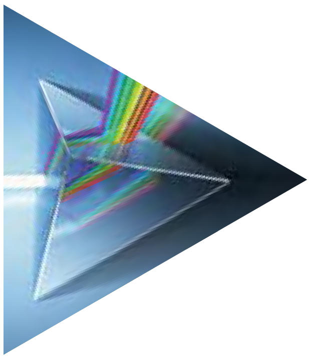
AUSTRALIAN CURRICULUM, ASSESSMENT AND REPORTING AUTHORITY

NAP Sample Assessment Science Literacy 2015



Public report

NAP NATIONAL ASSESSMENT PROGRAM

This public report was written by Nick Connolly with contributions from Mark Dulhunty, John Pedrazzini, Dr Sofia Kesidou, Dr Andrew Stephanou, Frans Boot, Melissa Lennon; some text from the 2012 NAP–SL Public Report was used.

2015 NAP–SL project staff:

Michelle Robins was ACARA's NAP–SL Project Manager and Rose Miller was the Executive Officer. John Pedrazzini was ACARA's Science Curriculum Lead and the author of chapter 9.

Dr Sofia Kesidou and Nick Connolly from Educational Assessment Australia (EAA) were the project directors. Jennifer Cowing was the EAA Project Manager. The EAA test development and marking teams were led by Frans Boot. The EAA NAP–SL helpdesk was run by Sally Matheson and Jessica Fountain. The sampling and data analysis tasks were led by Dr Rassoul Sadeghi, Dr Andrew Stephanou and Dr Kelvin Gregory from EAA; and Mark Dulhunty and Leanne Calvitto from Educational Measurement Solutions (EMS).

The following people all made important contributions at EAA to the 2015 NAP–SL project: Nicholas Marosszeky, Rosita Rastan, Somayeh Mahintorabi, Patrick Bruck, Greg Smith, Jessie Gao, Xiang Liu, Sue Smith, Jessica Conroy, Daisy Wang, Michael Combrink, Nilesh Prasad and Maria Grueva.

Contents

Appendices

NAP–SL TABLES & FIGURES

Acknowledgements

NAP sample assessment – science literacy 2015 working group members

Listed below are the main working group members representing the Australian Government, jurisdictions and school sectors. These members have made a valuable contribution to the project throughout the development and implementation phases.

ACARA
Chair, ACARA
Michelle Robins
Senior Psychometrician, ACARA
Dr Rassoul Sadeghi
Curriculum Specialist, ACARA
John Pedrazzini
Executive Officer, ACARA
Rose Miller

National representatives
Australian Government
Kelly Paxman
Catholic school sector
Michael Doyle
Independent school sector
Glenda Leslie

State and territory representatives
Australian Capital Territory
Niels Hider
New South Wales
Tanya Coli
Northern Territory
Gail Barwick
Queensland
Regan Spence
South Australia
Jan Brooks
Tasmania
Philippa Clymo
Victoria
Maria James
Western Australia
Marilyn McKee

Foreword

The NAP sample assessment–science literacy (NAP–SL) is one of a suite of three national sample assessments, which also include civics and citizenship, and information and communication technology literacy. These assessments are developed and managed by the Australian Curriculum, Assessment and Reporting Authority (ACARA) under the auspices of the Education Council.

Every three years, ACARA assesses a stratified random sample of students. These assessments support measurement and reporting on progress towards the achievement of the objectives outlined in the Melbourne Declaration on Educational Goals for Young Australians (2008). They monitor the extent to which our schooling promotes equity and excellence, and the progress of young Australians towards becoming successful learners, confident and creative individuals, and informed citizens.

The findings from the 2015 NAP sample assessment presented in this report provide valuable information on the science literacy of Year 6 students across Australia. The report also analyses the performance of states and territories and various subgroups, including Indigenous students, those living in remote and very remote areas, and students from language backgrounds other than English. The data provide insight into the level of science knowledge, understandings and skills of our Year 6 students. In addition, the student survey highlights the extent of students' interest in science, their engagement in science-related activities and their understanding of how science is relevant to their lives.

The present report is the fifth in the series of three-yearly reports, which began in 2003. Nationally, the results of the five assessments show little change in average student achievement or the proportion of students performing at or above the defined proficient standard in science literacy. The proficient standard is challenging, and only just over half (55.1 percent) of Year 6 students reach or exceed it.

The results of the student survey show that most students (over 80 per cent) appear to be interested in learning new things in science, learning about science and doing science-based activities. This interest in science is a strong foundation on which to build student awareness of the importance of science in their everyday lives, build confidence, and encourage students to consider rewarding future careers in the field of science.

The results from this assessment, along with those from international assessments in which our students participate (including the Programme for International Student

Assessment and Trends in International Mathematics and Science Study), highlight the need for improvements in primary school science teaching. Australia needs more than 55.1 per cent of Year 6 students at or above the proficient standard if we are to remain creative and competitive economically and socially. That is why, for the first time, this report includes a chapter prepared specifically for teachers and curriculum specialists. It contains suggestions about how to improve science learning in the classroom, using the Australian Curriculum.

I wish to acknowledge the collaboration and dedication of educators across all states and territories that have contributed to the development of the 2015 NAP–SL sample assessment.

Thanks as well to the principals, teachers, and students at government, Catholic and independent schools around Australia whose participation provided valuable information about science literacy in schools.

I commend this report to teachers, educators and community members engaged in achieving improved educational outcomes for all young Australians, and to those with a specific interest in helping young Australians to participate in a society that is increasingly dependent on science.

Emeritus Professor Steven Schwartz AM
Chair, Australian Curriculum, Assessment and Reporting Authority Board

1

EXECUTIVE SUMMARY

Executive Summary

The NAP sample assessment – science literacy (NAP–SL) defines science literacy as a student's ability to apply broad conceptual understandings of science in order to make sense of the world; to understand natural phenomena; and to interpret media reports about scientific issues. It also measures the ability to ask investigable questions, conduct investigations, collect and interpret data and make informed decisions. This 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, page 60)*

NAP–SL is one of a suite of three national sample assessments with civics and citizenship, and information and communication technology (ICT) literacy, which are conducted with stratified random samples of students in three-year cycles. The results contribute to an understanding of student progress towards the achievement of the Educational Goals for Young Australians specified in the Melbourne Declaration.

In July 2001, the Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA, now superseded by the Education Council) agreed to the development of assessment instruments and key performance measures for reporting on student skills, knowledge and understandings in primary science. The development and implementation of this national assessment in science literacy has been undertaken by the Australian Curriculum, Assessment and Reporting Authority (ACARA).

The first science literacy assessment was conducted in 2003. The assessment is repeated with a new sample of Year 6 students every three years in order to identify trends over time. The findings in this report describe the science literacy of Year 6 students from the 2015 assessment, with comparisons made to the 2003, 2006, 2009 and 2012 cohorts.

ACARA established the Science Literacy Working Group (SLWG) in 2014. Its role was to give advice throughout the course of the project on critical aspects of the assessment program and ensure that the assessments and results were valid across the states and territories. The main function of the SLWG was to ensure that the science literacy assessment domain was inclusive of the different state and territory curricula and that the items comprising the assessments were fair for all students.

Assessment domain

The science literacy assessment domain was developed in the first assessment cycle in consultation with curriculum experts from each state and territory, and representatives from the Catholic and independent school sectors. During each cycle, it has been reviewed and amended progressively in consultation with the NAP–SL Working Group.

For the 2015 cycle, a mapping exercise was conducted to determine the level of alignment between the Australian Curriculum: Science and the NAP–SL assessment domain. A high degree of alignment was evident. The science as a human endeavour strand of the curriculum (nature and development of science; use and influence of science) also served to inform stimulus contexts and aspects of the student survey.

Historical description

For past and present NAP–SL cycles, three main areas of scientific literacy have informed the item development process:

- Strand A: formulating or identifying investigable questions and hypotheses; planning investigations; and collecting evidence
- Strand B: 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: using science understandings for describing and explaining natural phenomena, and for interpreting reports about phenomena.

The science literacy assessment domain is detailed in appendix 1.

In addition, the item development process drew on four major scientific concept areas: earth and space; energy and force; living things; and matter. These concept areas, found most widely in state and territory curriculum documents and mapped to the Australian Curriculum, were used by item developers to guide item and test development.

A conscious effort was made to develop assessment items that would relate to everyday contexts. The intention was to ensure that all Year 6 students were familiar with the materials and experiences to be used in NAP–SL and so avoid any systematic bias in the instruments developed.

Australian Curriculum description

The Australian Curriculum: Science has three interrelated strands – science understanding, science as a human endeavour, and science inquiry skills – that are designed to be taught in an integrated way. Together, these three strands provide students with understanding, knowledge and skills through which they can develop a scientific view of the world.

These three strands are then further divided into sub-strands, year levels and specific content descriptions.

During the initial test development phase, ACARA's NAP–SL Working Group and Educational Assessment Australia worked collaboratively to develop a document that would provide a bridge between the existing NAP–SL assessment framework and that would be required for NAP–SL 2018. The results of the mapping initiative appear in the 2015 NAP–SL Assessment Framework bridging document, now available on ACARA's website. Appendix 1 of this report includes an extract from the bridging document.

As an additional step, every item used in the 2015 cycle was mapped against the curriculum with an emphasis on Year 5 and Year 6 content descriptions. Items were also classified against the general capabilities in the Australian Curriculum: Science, including the Critical and Creative Thinking capability.

Assessment instrument

The assessment instrument was administered to a stratified random sample consisting of approximately five per cent of the total Australian Year 6 student population.

Test dates

The students' regular classroom teachers administered NAP–SL during the following time periods:

- Monday 12 October – Friday 30 October 2015 in NSW, NT, Qld and Vic.
- Monday 19 October – Friday 6 November 2015 in the ACT, SA, Tas. and WA.

Online delivery

The 2015 cycle marked the first online delivery of NAP–SL. Over the course of many months, a bespoke platform was developed to accommodate the various requirements of the assessment. The student test player was configured to present the following components to students:

- one set of practice items designed to help students use the online system
- one tests consisting of a variety of item types that tested students' science literacy skills in a broad range of contexts
- one inquiry task consisting of a variety of item types that tested students' science literacy skills within the context of a single inquiry activity
- one student survey of students' attitudes towards science.

Students were allowed 60 minutes for the test and 35 minutes for the inquiry task. While the survey was an untimed component of the assessment, approximately 20 minutes was recommended.

Adapting the assessment for online and the impact of the change in mode

One purpose of NAP–SL is to monitor trends over time. To do this, a selection of secure items used in previous cycles is used in later assessments. In 2015 these 'historical' items needed to be adapted from paper versions to online versions.

Making any change to an item can result in changes to its difficulty or other properties. To understand the effect of these changes better, a special study called 'a mode-effect study' was conducted. In this study, a sample of Year 6 students was selected and then randomly divided into two groups. One group completed a paper version of the historical items and the second group completed an online version. The performance of both groups was then compared and any differences were noted.

This study found that the online, adapted, versions of the historical NAP–SL link items used in the study were more difficult for the students than the original paper versions. Given that information about performance of these link items is used in the longitudinal equating of NAP–SL tests, the observed differences were taken into account during the analysis and reporting of the 2015 NAP–SL outcomes.

It is important to note that these differences apply only to the set of adapted online test items and are therefore only applicable to the 2015 NAP–SL cycle. In addition, and equally important, the observed difference in mode of testing are not expected to occur in other NAP assessments including NAPLAN tests as they have different items and test design compared with NAP–SL tests. Further, NAPLAN online tests have been developed to maximise the comparability of test items across the online and paper tests.

Student performance in science literacy

Student performance is reported against a scale that was developed in 2006 (rather than the original 2003 scale). The 2006 sample was more inclusive of remote schools and contained items that catered for a greater range of student ability. Consequently, the results of the 2006 assessment were used to establish a new baseline science literacy scale and has been used since then.

The 2015 student performance has been scaled to the 2006 baseline. As a result, discussion of trends in this report is primarily based on comparisons between the 2006, 2009, 2012 and 2015 assessments.

Table ES.1 shows the national, state and territory mean scores and indicates whether the differences in means between 2012 and 2015, between 2009 and 2015, and between 2006 and 2015 are statistically significant.

Table ES.1

Comparison of 2006, 2009, 2012 and 2015 jurisdiction mean scores

State/ territory	Mean score				Change and statistical significance					
	2006	2009	2012	2015	2012–2015		2009–2015		2006– 2015	
NSW	411 (±12.5)	396 (±12.1)	395 (±9.9)	411 (±8.6)	16	NO	15	NO	0	NO
Vic.	408 (±10.2)	398 (±9.2)	393 (±9.7)	399 (±8.9)	6	NO	1	NO	–9	NO
Qld	387 (±8.6)	385 (±8.9)	392 (±6.4)	398 (±10.6)	6	NO	13	NO	11	NO
WA	381 (±10.0)	393 (±9.6)	406 (±9.5)	408 (±7.5)	2	NO	15	NO	27	YES
SA	392 (±10.0)	380 (±10.4)	392 (±7.9)	392 (±8.8)	0	NO	12	NO	0	NO
Tas.	406 (±12.1)	386 (±13.5)	395 (±12.3)	414 (±11.7)	19	NO	28	YES	8	NO
ACT	418 (±14.3)	415 (±10.6)	429 (±13.2)	414 (±12.1)	–15	NO	–1	NO	–4	NO
NT	325 (±33.7)	326 (±28.6)	319 (±31.1)	320 (±25.6)	1	NO	–6	NO	–5	NO
Aust.	**400 (±5.4)**	**392 (±5.1)**	**394 (±4.4)**	**403 (±4.3)**	**9**	**NO**	**11**	**NO**	**3**	**NO**

Note: Figures in parentheses refer to 95 per cent confidence intervals. Mean scores have been rounded.

As can be seen in table ES.1, there is no statistically significant difference between 2012 and 2015 results at the national level or across the states or territories.

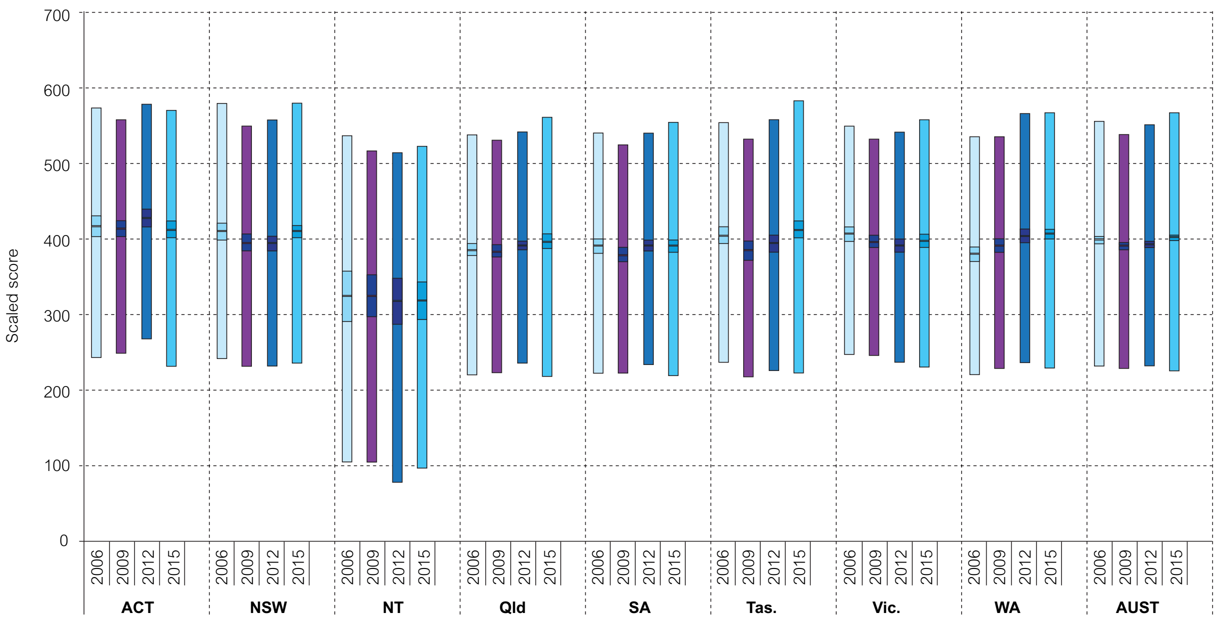
Similarly, the comparison between 2006 and 2015 results shows that there is no statistically significant difference at the national level. The exception was Western Australia, which maintained the cumulative gains achieved between 2006 and 2012, showing a significant increase of 27 points in 2015 compared with the mean student achievement in 2006.

When comparing 2015 results with 2009, there is no significant difference at the national level. However, at the jurisdiction level, Tasmania was the only jurisdiction to achieve significantly higher results accumulated over time since the 2009 cycle.

Figure ES.1 also shows a comparison of the student results. Shaded bands around the mean within each bar mark the 95 per cent confidence interval (CI) around the mean (the upper and lower five per cent of the distribution are excluded due to the large variability associated with the extreme scores at the two ends of the scale).

Figure ES.1

Comparison of distributions of student scores by state and territory in 2006, 2009, 2012 and 2015



Note: Owing to the fact that the NAP-SL scale is designed to have mean of 400 (and standard deviation of 100) and that NT have students who perform at more than five units of standard deviation below such a mean, the fifth percentile point for NT 2006 is a negative score. For the graphical presentation purposes the tenth percentile (a scaled score of 104) is used as the lowest data point for NT.

In terms of state and territory performance in 2015, when the test of statistical significance including the Bonferroni adjustment[1] was applied, students from the Northern Territory achieved significantly lower mean score than students in all other states and territories. All other differences between jurisdictions were not statistically significant.

1 The Bonferroni adjustment to statistical significance testing is conducted in order to account for the possibility that a difference can be deemed to be statistically significant by chance when multiple comparisons are conducted using the same data. The Bonferroni adjustment increases the strictness of the criterion for establishing statistical significance relative to a pair-wise comparison, hence making it harder to claim that a difference is statistically significant.

Group performance

At the national level, a comparison of mean achievement between student groups showed the following results:

- For male and female students, there was a significant difference in mean achievement. Female students performed better than male students nationally. This is a change from 2012, when the mean for females was not significantly higher than males.
- In every jurisdiction except Queensland, the mean score for female students was higher than male students (see table ES.2). Western Australia had the largest difference between gender groups, with the mean score for female students 28 points higher than that for male students.
- Indigenous students had a statistically significant lower mean achievement than non-Indigenous students (see table ES.3). These results are consistent with the 2006, 2009 and 2012 results.
- Students from metropolitan areas had significantly higher mean scores than students in provincial areas, who in turn had higher mean scores than students in remote and very remote areas (see table ES.4). This is consistent with the 2012 results.

Table ES.2

Mean scores for male and female students by state and territory in 2015

State/territory	Percentage of male students in the sample	Male students: mean score (± 95 per cent CI)	Female students: mean score (± 95 per cent CI)	Mean
NSW	52.9	405 (±10.4)	418 (±9.8)	411 (±8.6)
Vic.	52.4	395 (±10.4)	403 (±11.1)	399 (±8.9)
Qld	49.3	399 (±11.9)	398 (±12.5)	398 (±10.6)
WA	47.8	393 (±9.9)	421 (±8.5)	408 (±7.5)
SA	52.3	388 (±9.9)	397 (±12.2)	392 (±8.8)
Tas.	49.5	403 (±14.0)	424 (±13.7)	414 (±11.7)
ACT	51.6	410 (±11.8)	419 (±14.8)	414 (±12.1)
NT	49.3	313 (±29.6)	327 (±30.9)	320 (±25.6)
Aust.	**50.8**	**398 (±5.1)**	**408 (±5.1)**	**403 (±4.3)**

Table ES.3

Mean scores for Indigenous and non-Indigenous students in 2015

Indigenous status	Mean score (± 95 per cent CI)
Indigenous	315 (±13.7)
Non-Indigenous	408 (±4.2)

Table ES.4

Mean scores for students by school geographic location in 2015

Geographic location	Percentage of students	Mean score
Metropolitan areas	72.1	410 (±5.1)
Provincial areas	25.3	389 (±8.3)
Remote and very remote areas	2.6	348 (±35.8)
Aust.	**100.0**	**403 (±4.3)**

Notes: Figures in parentheses refer to 95 per cent confidence intervals. The percentages of students in geographic locations are weighted to reflect the population of Year 6 students in Australia. They are not the percentage of students in the sample.

Distribution of students across proficiency levels

One of the key objectives of the National Assessment Program is to monitor trends in science literacy performance over time. One convenient and informative way of describing student performance is to reference the results to proficiency levels. A similar process is used in other NAP Sample programs and several international large-scale assessment programs, including PISA. Students whose results are located within a particular level of proficiency are typically able to demonstrate the understandings and skills associated with that level, and also typically have the understandings and skills defined as applying at lower levels.

The development of proficiency levels

To establish the proficiency levels in the 2003 assessment cycle, a combination of experts' knowledge of the understandings and skills required to answer each science literacy item, and information from the analysis of students' responses was used. This defined five proficiency levels for reporting student performance from the assessment.

The 'proficient standard' and what it represents

The proficient standard for Year 6 science literacy was also established in the 2003 assessment cycle to provide parents, educators and the community with a clear picture of what students should know and can do by the end of Year 6.

To set the standard, an expert group, comprising university science educators, curriculum officers and experienced primary teachers from all states and territories, from government, Catholic and independent schools, was brought together. The crucial science literacy skills and understandings needed by students for the next phase of science learning at school were discussed and debated before a consensus was reached on a proficient standard for Year 6.

The proficient standard is defined as a challenging level of performance, with students required to demonstrate more than minimal or elementary skills to be regarded as having reached it. The standard is situated at the boundary between 3.1 and 3.2.

Students who do not achieve the proficient standard demonstrate only partial mastery of the skills and understandings expected for Year 6. There are also students who show superior results and exceed the proficient standard.

The proficient standard is one of several levels of student achievement in science that collectively represent a continuum of learning and describe what students know and can do. Initially, in 2003, three levels of achievement corresponding with Levels 2, 3 and 4 of the progress map were identified for reporting on science literacy performance. However, as 90 per cent of students' scores were within Level 3, three further levels within Level 3 were created in 2006. This provided five levels for reporting student performance in the assessment. The proficient standard was deemed to be Level 3.2 on the continuum. This standard informed the development of the tests for the 2006, 2009, 2012 and 2015 assessments.

Proficiency levels by state and territory in 2015

Table ES.5 shows a comparison at the jurisdiction level of the proportion of students in each of the proficiency levels and the proportion of students performing at or above the proficient standard in 2015. The table also shows the mean score on the NAP–SL scale for each state and territory and for Australia as a whole, and the standard deviation. The five levels of proficiency are defined and described in chapter 5.

Table ES.5

Percentages of students at proficiency levels by state and territory in 2015

State/ territory	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above	At or above the proficient standard	Mean	Standard deviation
NSW	8.2 (±1.7)	34.5 (±3.3)	43.0 (±3.4)	13.3 (±2.9)	1.0 (±0.7)	57.2 (±3.6)	411 (±8.6)	105.1 (±5.0)
Vic.	9.1 (±1.7)	37.3 (±3.3)	43.1 (±3.0)	10.1 (±2.4)	0.4 (±0.5)	53.6 (±3.8)	399 (±8.9)	100.7 (±4.1)
Qld	11.1 (±3.2)	34.6 (±3.6)	42.8 (±4.1)	11.3 (±2.5)	0.3 (±0.3)	54.3 (±4.6)	398 (±10.6)	105.1 (±6.3)
WA	9.3 (±2.1)	33.0 (±3.1)	44.7 (±3.0)	12.5 (±2.4)	0.4 (±0.4)	57.7 (±3.3)	408 (±7.5)	103.1 (±5.0)
SA	11.0 (±2.4)	38.3 (±3.5)	40.8 (±3.3)	9.6 (±2.2)	0.3 (±0.3)	50.7 (±3.9)	392 (±8.8)	102.1 (±5.0)
Tas.	9.6 (±2.6)	31.4 (±4.1)	42.5 (±4.1)	15.7 (±3.7)	0.8 (±1.0)	59.1 (±4.7)	414 (±11.7)	110.5 (±6.8)
ACT	8.0 (±3.5)	31.5 (±3.9)	46.3 (±4.8)	13.8 (±2.5)	0.4 (±0.5)	60.5 (±5.1)	414 (±12.1)	102.2 (±9.7)
NT	32.8 (±8.3)	35.4 (±6.5)	26.7 (±5.3)	5.0 (±2.2)	0.1 (±0.2)	31.8 (±5.6)	320 (±25.6)	133.6 (±18.7)
Aust.	**9.7 (±1.0)**	**35.2 (±1.6)**	**42.9 (±1.5)**	**11.7 (±1.2)**	**0.6 (±0.3)**	**55.1 (±1.8)**	**403 (±4.3)**	**104.7 (±2.5)**

Note: percentages may not add up to 100 per cent due to rounding.

Table ES.5 shows that the Australian Capital Territory and Tasmania recorded similar percentage distributions with around 60 per cent of students attaining the proficient standard (that is, at or above Level 3.2). The smallest proportion of such students achieving the standard was observed in the Northern Territory.

Historical comparison of proficiency levels in 2006, 2009, 2012 and 2015

In 2015, 55.1 per cent of students at the national level attained the proficient standard or higher in science literacy. The data in table ES.6 show the proportion of students in each level and the proportion of students performing at or above the proficient standard in 2006, 2009, 2012 and 2015 at the national level.

Table ES.6

Historical comparison of percentages of students at proficiency levels in 2006, 2009, 2012 and 2015

Australia	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above	At or above the proficient standard
2006	8.6 (±1.1)	37.1 (±1.7)	44.2 (±1.8)	9.6 (±1.2)	0.5 (±0.4)	54.3 (±2.1)
2009	9.1 (±1.2)	39.0 (±1.7)	44.5 (±1.8)	7.2 (±1.1)	0.1 (±0.1)	51.9 (±2.2)
2012	9.0 (±1.0)	39.6 (±1.6)	42.1 (±1.7)	9.0 (±1.1)	0.3 (±0.2)	51.4 (±2.0)
2015	9.7 (±1.0)	35.2 (±1.6)	42.9 (±1.5)	11.7 (±1.2)	0.6 (±0.3)	55.1 (±1.8)

Note: Due to a change in the structure of the assessment and the sampling in 2006, results from 2003 are not shown. Percentages may not add up to 100 per cent due to rounding.

Table ES.6 shows that at the national level, the difference between 2015 and 2006 in the proportion of students performing at or above the proficient standard is 0.8 percentage points. The corresponding difference between 2015 and 2009 is 3.2 percentage points and between 2015 and 2012 is 3.7 percentage points. These differences were not statistically significant. Similarly, there was no significant difference in the distribution of students across proficiency levels at the national level for the assessments in 2006, 2009 and 2012.

Analysis conducted for states and territories in terms of the difference in the percentages of students achieving at or above the proficient standard shows that the percentage of Western Australian students achieving this standard has increased significantly from 2006. Other differences between cycles for all other states and territories are not statistically significant.

Survey results

Since 2009, students have also been required to complete a survey. The survey canvassed students' perceptions of, and attitudes to, science. It also asked students about their science-learning experiences at school.

Survey items were also scaled using similar techniques as were used to develop the science literacy scale.

Survey items related to students' self-concept in science (that is, the level of belief that students have in their own science competencies) had significant correlations with test performance in all states and territories. That is, a higher science self-concept was associated with a higher mark in science literacy. Interestingly, over 85 per cent of students responded that they would like to learn more science at school and 69 per cent of students indicated that they believe it would be interesting to be a scientist, indicating that a positive attitude towards this subject area exists. Disappointingly, only 31 per cent of students indicated that guest speakers are invited to their school to talk about science topics. There is a growing body of research that clearly points to the benefits of contextualisation of science, which can be achieved quite simply by the inclusion of outside specialists. The survey also showed that more than half of students disagreed with the idea that science is easy for most people to understand.

New items were added to the survey for 2015 that related to the science as a human endeavour strand of the Australian Curriculum: Science. These items also showed significant correlations with test performance in all states and territories. Over 95 per cent of students agreed that both women and men are involved in science and 94 per cent of students agreed that people from many different countries have made important contributions to science.

NAP–SL for teachers

Mapping the NAP–SL construct to the Australian Curriculum: Science in 2015 represented a significant milestone. As a result, there are now new opportunities to provide information that is more directly relevant to teachers to help support student learning. The suggestions shown in chapter 9, which provide linkages to the curriculum should help to identify specific learning opportunities.

Conclusion

The results of the 2015 NAP–SL assessment at the national level remained similar to those observed in the previous assessment cycles, both in terms of mean student achievement and the proportion of students performing at or above the proficient standard in science literacy. Similarly, the comparison between the results from previous cycles shows that there are few statistically significant differences in mean student achievement and in the proportion of students performing at or above the proficient standard in most jurisdictions.

Results of the student survey showed that Australian students continue to have positive views of science and about learning science.

How is this report organised?

This report provides educators and policy-makers with the main findings of the 2015 NAP–SL assessment. The 2015 NAP–SL Technical Report provides more detailed information about the development of the assessment instruments, data collection and the analyses that underpin the findings presented in this report.

Chapter 1 – provides an overview of the national assessment program and includes important historical and contextual information.

Chapter 2 – describes the development of the assessment framework, assessment instruments and survey, including the piloting and trialling processes.

Chapter 3 – describes the sampling processes, test administration and how results were reported to schools.

Chapter 4 – provides a description of the science literacy scale. It includes results in terms of means and distributions of student performance for the Australian population as well as for each state and territory. The chapter also contains comparisons of the performance of Year 6 students over the five cycles of NAP–SL assessments.

Chapter 5 – discusses the results in terms of the students' proficiency as levels on the science 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 items accompanies the discussion of students' results.

Chapter 6 – examines comparisons in achievement by proficiency levels over the five assessment cycles.

Chapter 7 – provides an analysis of the results achieved by specific groups of students, including male and female students, Indigenous and non-Indigenous students, as well as students from various geographic locations or language backgrounds.

Chapter 8 – presents the survey results about the students' opinions and ideas about science and the role of science in their lives and society. The chapter also reports on the relationship between students' responses to the survey and their achievement in the assessment.

Chapter 9 – provides a historical perspective to the Australian Curriculum: Science, the links between the Australian Curriculum and NAP–SL and suggestions to improve student science-learning in the classroom, using the Australian Curriculum.

Chapter 10 – provides a brief summary of the main findings of 2015 NAP–SL and the implications of those findings.

The report is further enhanced with additional information in the form of appendices. Each appendix provides more specific details about an aspect of the project.

Note on reading tables and figures in this report

This report includes multiple tables and figures providing insights into the results of the NAP sample assessment – science literacy for 2015. When reading these tables and figures, the following issues should be noted:

- Percentages may not always add up to 100 per cent due to rounding.
- Due to a change in the structure of the assessment and the sampling in 2006, results from 2003 are indicative only. In some cases, results from 2003 have not been included.
- Some statistics have been weighted to reflect proportions in the population. For more information about the applied weights and the sampling design, refer to the 2015 NAP–SL Technical Report.
- In several tables, numbers in parentheses refer to 95 per cent confidence intervals; for example, (±2.5).
- ‘Geographic location’ refers to whether a student attended school in a metropolitan, provincial or remote / very remote area (Jones, 2004).
    - Metropolitan areas include state capital cities and major urban areas with populations above 100,000, such as Geelong, Wollongong and the Gold Coast.
    - Provincial areas include provincial cities (including Darwin) and provincial areas.
    - Remote and very remote areas include areas of low accessibility, such as Katherine and Coober Pedy.

Chapter 1

OVERVIEW OF THE NATIONAL ASSESSMENT

Chapter 1. Overview of the National Assessment

Introduction

The first national science assessment, designed, developed and carried out under the auspices of the Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA), was conducted in 2003. This assessment represented a new direction in monitoring, and reporting on, progress towards achievement of goals for schooling on a nationally comparable basis.

The development of nationally comparable assessments arose from an agreement by MCEETYA in July 2001 to develop assessment instruments and key performance measures for reporting on student skills, knowledge and understandings in primary school science. The NAP sample assessment – science literacy (NAP–SL) was the first assessment program designed specifically to provide information about student performance against nationally agreed goals; now called the Educational Goals.

In 2004 and 2005, similar national assessments were introduced for students in Years 6 and 10 in civics and citizenship, and information and communications technology (ICT) literacy. Each of these programs assesses a representative sample of Australian students and is repeated every three years.

In December 2008, the state, territory and Commonwealth ministers of education released the Melbourne Declaration on Educational Goals for Young Australians. The new Educational Goals for Young Australians set the direction for Australian schooling for the next 10 years (MCEETYA, 1999 and 2008).

The role of MCEETYA is now carried out by the Education Council, which provides a forum through which strategic policy on school education is coordinated at a national level.

Under the auspices of the Education Council, the development and implementation of the national assessment in science literacy is now undertaken by the Australian Curriculum, Assessment and Reporting Authority (ACARA), the independent statutory authority responsible for the overall management and development of a national curriculum, the National Assessment Program and a national data collection and reporting program that supports 21st century learning for all Australian students.

The previous four NAP–SL assessments were conducted in 2003, 2006, 2009 and 2012. In 2014, ACARA awarded the contract for the fifth cycle of NAP–SL to Educational Assessment Australia (EAA). This report provides the findings of the fifth cycle of NAP–SL conducted in 2015.

NAP sample assessment – science literacy

Implementation of NAP–SL involves many steps that include the development of items and tasks to assess the science literacy domain; the trialling of those items and tasks; the administration of the final 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 in the 2015 NAP–SL cycle; describes the testing process; presents the results at national, state and territory levels; and includes comparisons with previous testing cycles.

What does NAP–SL measure?

NAP–SL measures the science literacy of primary school students in Australian schools. Unlike other assessments that are part of the National Assessment Program, NAP–SL only assesses Year 6 students. Science literacy in secondary education in Australia is assessed as part of the Organisation for Economic Co-operation and Development (OECD) – Programme for International Student Assessment (PISA).

In the first cycle of PISA, 'scientific literacy' was defined 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 was adopted for the purpose of monitoring primary school science in the first cycle of NAP–SL. NAP–SL assesses the ability to think scientifically in a world in which science and technology are increasingly shaping children's lives. Specifically, it assesses students' ability to apply broad conceptual understandings of science in order to make sense of the world, to understand natural phenomena, and to interpret media reports about scientific issues. It also assesses the ability to ask investigable questions, conduct investigations, collect and interpret data, and make informed decisions.

A science literacy progress map (see appendix 1) developed during the first NAP–SL cycle was based on this construct of science literacy and on an analysis of the state and territory curriculum and assessment frameworks. The progress map describes the development of science literacy across three strands: experimental design and data gathering, interpreting experimental data, and applying conceptual understanding.

For 2015 an assessment framework bridging document was developed to guide the transition of the assessment program toward the Australian Curriculum: Science in preparation for NAP–SL in 2018.

What aspects of science literacy are assessed?

The 2015 cycle marks a major transition point for NAP–SL. The adoption of the Australian Curriculum: Science by all states and territories allows for a more consistent description of the science skills and content taught in Australian schools.

To enable the results of 2015 NAP–SL to be compared with those of previous cycles, two complementary descriptions of science literacy have been used:

- the historical description used in previous cycles
- a new description based on the Australian Curriculum: Science.

Historical description

In the previous and current NAP–SL cycles, 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 students' 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.

In addition, the item development process was informed by four major scientific concept areas: earth and space; energy and force; living things; and matter. These concept areas, found most widely in state and territory curriculum documents, were used by item developers to guide item and test development.

A conscious effort was made to develop assessment items that related to everyday contexts. The intention was to ensure that all Year 6 students were familiar with the materials and experiences to be used in NAP–SL and so avoid any systematic bias in the instruments developed.

The science literacy domain is detailed in appendix 1.

Australian Curriculum: Science description

The Australian Curriculum: Science has three interrelated strands – science understanding; science as a human endeavour; and science inquiry skills – designed to be taught in an integrated way. Together, these three strands provide students with understanding, knowledge and skills through which they can develop a scientific view of the world.

These three strands are then further divided into sub-strands, year levels and specific content descriptors.

Every item used in the 2015 cycle was mapped against the curriculum from Year 2 to Year 6, with an emphasis on Year 5 and Year 6 content descriptors. After classifying every item, the spread of items against year level, sub-strand and strand could be monitored.

In addition to this, items were also classified against the general capabilities of the Australian Curriculum: Science, including the critical and creative thinking capability.

What is the national science literacy standard?

A standard for science literacy was established as a part of the first assessment cycle 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.

To identify what students should know and be able to do by the end of Year 6, an expert group, comprising university science educators, curriculum officers and experienced primary school teachers in all states and territories, from government, Catholic and independent schools, was brought together. 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 students were required to have for the next phase of science-learning at school were discussed and debated before a consensus was reached on a proficient standard for Year 6.

The proficient standard

The proficient standard is defined as a challenging level of performance, with students required to demonstrate more than minimal or elementary skills. It is one of several achievement levels that collectively represent a continuum of learning and describe what students know and can 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 Level 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 do not achieve the proficient standard demonstrate only partial mastery of the skills and understandings expected for Year 6.

Differences from the National Minimum Standards in NAPLAN

Minimum standards like the National Minimum Standards in literacy and numeracy have not been set for science literacy. These minimum standards are defined as the critical level of skill and understanding without which a student will have difficulty making sufficient progress at school. The minimum standards 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 the society.

The proficient standard in NAP–SL is the main reference point for monitoring science literacy in Australian primary schools over time. Every three years, a new Year 6 national science literacy assessment is conducted to gauge whether there has been changes in student proficiency.

Information about student performance against the Year 6 standard is reported in chapter 6 together with comparisons to performance in previous cycles.

How was 2015 NAP–SL different from previous cycles?

The 2015 NAP–SL assessment had a number of features that marked significant departures from past cycles. However, every care was taken to minimise the impact of these changes so that valid comparisons could be made.

Online delivery

In previous NAP–SL cycles, the assessment took the form of printed booklets containing both the items and space for students to answer them.

For the first time in 2015, NAP–SL was delivered online. This move to online test delivery is part of a general move towards online testing that has been implemented progressively in the National Assessment Program's sample assessments and follows the successful implementation of online delivery for NAP sample assessment – civics and citizenship in 2013.

Further details of how NAP–SL was adapted for online delivery, including the introduction of new item types, are described in chapter 3.

The inquiry tasks

In previous cycles, the assessment included two group work practical tasks. Each practical task required students to conduct an investigation in groups of three and then respond individually to a set of items about the investigation. The purpose of the practical tasks was to give students opportunities to demonstrate their science inquiry skills in context.

The move to online delivery for 2015 NAP–SL meant that the previous approach for assessing inquiry skills in context was not viable. Instead, students were presented with one of two online practical tasks that took them through the stages of a simulated science investigation using video and other stimuli. Students answered items as they progressed through the stages of the investigation. The 2015 inquiry tasks assessed similar knowledge, skills and understandings to the previous NAP–SL practical tasks, but used technology to provide the context.

The Australian Curriculum: Science

The Australian Curriculum: Science has introduced a greater degree of consistency in the science taught in Australian schools. States and territories began implementing the curriculum from 2011.

The Year 6 students who were assessed in 2015 NAP–SL were in Year 2 when the process of implementation began and have been only partly exposed to the new curriculum. The concepts and skills assessed by 2015 NAP–SL were chosen to enable a transition to the Australian Curriculum: Science as the definitive basis for future cycles of NAP–SL.

Who participated in the 2015 NAP–SL assessment?

A stratified random sample making up approximately five per cent of the total Australian Year 6 student population took part in this assessment. The sample was drawn from government, Catholic and independent schools in all states and territories.

Table 1.1 shows the number of schools and students in the final sample for which results were reported.

Table 1.1

Number of schools and students by state and territory in the final sample 2015

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	
NSW	93	91	97.8	2185	1911	87.5
Vic.	92	91	98.9	2162	1930	89.3
Qld	91	89	97.8	2177	1833	84.2
WA	92	88	95.7	2126	1878	88.3
SA	94	86	91.5	2178	1790	82.2
Tas.	62	60	96.8	1366	1198	87.7
ACT	55	55	100.0	1366	1221	89.4
NT	49	39	79.6	920	649	70.5
Aust.	**628**	**599**	**95.4**	**14480**	**12410**	**85.7**

A grade-based population of students enrolled at Australian primary schools was chosen. This is consistent with the other National Assessment Program sample assessments. 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 ages of students at the time of testing is presented in table 4.1 in chapter 4.

Appendix 2 provides a summary of the sample frame, with exclusions and response rates for participating schools and students by state and territory for the assessment. Further details about sampling procedures and computation of sampling weights are provided in the 2015 NAP–SL Technical Report.

What did NAP–SL participants have to do?

The online assessment comprised seven linked objective tests consisting of multiple choice and short response questions, two inquiry tasks and a survey.

The objective tests required students to work individually to respond to approximately 40 questions. Students were allowed 60 minutes to complete the assigned objective test.The objective tests were allocated randomly so that in each participating school, the students collectively completed all seven different objective tests.

Each student undertook one of two inquiry tasks. The inquiry tasks were assigned to students in every school in a way that ensured that each of the two tasks was attempted by approximately the same number of students. The inquiry tasks required students to respond to a set of approximately 10 items based on a simulation of a science investigation.

Merging the seven objective tests onto one scale was achieved by the use of common items between the tests. The inquiry tasks were then linked onto this scale using results from students doing the same objective test and inquiry task.

In addition, students were asked to respond to a 43-item survey. The student survey sought to obtain information about students' perceptions of and attitudes to science, and their experiences of science-learning within and outside school. Students were given approximately 10 minutes to complete the survey.

The students' regular classroom teachers administered the NAP–SL assessment. The online delivery of the assessment made it necessary to allocate windows of time for students at different schools to complete the task. These were set in the following periods:

- Monday 12 October – Friday 30 October 2015 in NSW, NT, Qld and Vic.
- Monday 19 October – Friday 6 November 2015 in the ACT, SA, Tas. and WA.

How are the NAP–SL results reported?

The results of NAP–SL are reported as mean scores and distributions of scores across proficiency levels. They are also described in terms of the understandings and skills that students demonstrated in the assessment.

Five levels of proficiency are defined and described for science literacy. Further details of the proficiency levels, including items exemplifying these levels, are presented in chapter 5. Chapter 6 includes results in relation to the proficiency levels by state and territory.

Results for groups such as male and female students, Indigenous students, students from different geographic locations, and students from language backgrounds other than English are presented in chapter 7.

Chapter 2

Developing the Science Literacy Assessment

Chapter 2. Developing the Science Literacy Assessment

Introduction

This chapter provides a description of the steps that were followed to develop the science literacy assessment.

In previous cycles, the process of developing the assessment included the following steps:

1. clarifying the assessment domain for science literacy
2. constructing assessments that comprised items and tasks, which defined the assessment strands operationally
3. trialling the assessments in a sample of schools
4. constructing the final assessments based on the results of the trial.

For 2015 NAP–SL, a similar overall structure was followed but extra steps were required to accommodate the transition to online testing and the implementation of the Australian Curriculum: Science.

Assessment framework development

To ensure that historical comparisons are valid, it is important that the underlying construct of the NAP–SL assessment is maintained. At the same time, the new Australian Curriculum: Science has provided an opportunity to consider aspects of science literacy that have not been assessed in previous cycles. As a consequence, the assessment framework was reviewed in consultation with ACARA's science curriculum experts and the NAP Science Literacy Working Group (SLWG). A bridging document was developed to describe and augment the existing NAP–SL framework in terms of the Australian Curriculum: Science to reflect developments in the curriculum.

The historic NAP–SL assessment domain

The historic NAP–SL assessment domain comprised two key components: the progress map and the major scientific concept areas.

The progress map describes the development of science literacy across three strands:

- Strand A: formulating or identifying investigable questions and hypotheses; planning investigations; and collecting evidence.
- Strand B: 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: using science understandings for describing and explaining natural phenomena; and for interpreting reports about phenomena.

In previous cycles, the major scientific concept areas provided an indication of the pool of concepts from which item content related to science understanding was drawn. The science understanding strand of the Australian Curriculum: Science has since superseded the state and territory curricula from which the major scientific concept areas and examples were drawn. This strand was primarily used in the 2015 NAP–SL cycle as the basis for item development related to understanding of science concepts as it provided more specific content descriptions.

Transitioning to the Australian Curriculum: Science

In 2010, the federal, state and territory education ministers of Australia endorsed the release of the Australian Curriculum: Science.

This curriculum requires students to develop an understanding of important science concepts and processes; the practices used to develop scientific knowledge; and science's contribution to our culture and society and its applications in our lives.

Accordingly, the Australian Curriculum: Science has three interrelated strands – science understanding, science as a human endeavour and science inquiry skills – which are designed to be taught in an integrated way. Together, these three strands provide students with understanding, knowledge and skills through which they can develop a scientific view of the world. Students are challenged to explore the nature of science, its concepts and uses through clearly described inquiry processes. Table 2.1 lists the strands of the curriculum and the sub-strands within each strand.

Table 2.1

Strands and sub-strands in the Australian Curriculum: Science

Strands	Sub-strands
Science understanding	Biological sciences
	Chemical sciences
	Earth and space sciences
	Physical sciences
Science as a human endeavour	Nature and development of science
	Use and influence of science
Science inquiry skills	Questioning and predicting
	Planning and conducting
	Processing and analysing data and information
	Evaluating

Previous cycles of NAP–SL were developed in the absence of any common science curriculum across the states and territories. With the implementation of the Australian Curriculum: Science in all states and territories, it is important that the NAP–SL construct and progress map can also be described in terms of the new curriculum.

Table 2.2

NAP–SL progress map strands mapped onto the strands/sub-strands of the Australian Curriculum: Science

The NAP–SL progress map strands	Australian Curriculum: Science strands/sub-strands
Strand A: formulating or identifying investigable questions and hypotheses; planning investigations; and collecting evidence	Science inquiry skills – questioning and predicting Science inquiry skills – planning and conducting Science as a human endeavour
Strand B: interpreting evidence and drawing conclusions from students' own or others' data; critiquing the trustworthiness of evidence and claims made by others; and communicating findings	Science inquiry skills – processing and analysing data and information Science inquiry skills – evaluating Science inquiry skills – communicating Science as a human endeavour
Strand C: using science understandings for describing and explaining natural phenomena; and for interpreting reports about phenomena	Science understanding Science as a human endeavour

Table 2.2 shows a high degree of alignment between the NAP–SL Strand A: experimental design and data gathering and Strand B: interpreting experimental data and the science inquiry skills strand of the Australian Curriculum: Science.

The NAP–SL progress map Strand C: applying conceptual understanding provides an abstract representation of progression in students' use of science concepts for describing and explaining natural phenomena and interpreting reports about phenomena that makes no reference to particular science concepts. In previous NAP–SL cycles, the progression articulated in Strand C provided guidance for the development of items that reflect levels of increasing complexity and abstraction in students' understanding of science concepts while (in the absence of a common science curriculum across states and territories) the major scientific concept areas document provided the contexts and specific concepts used to assess science understanding.

Table 2.3 shows how the NAP–SL major scientific concept areas map onto the Australian Curriculum: Science strand of science understanding. The science understanding strand of the Australian Curriculum: Science provides guidance about the specific concepts to be assessed in the NAP–SL tests. Appendix 1 shows the mapping between the major scientific concept areas and the science understanding strand at a finer level. There is no explicit equivalent of the abstracted progression articulated in Strand C in the Australian Curriculum: Science.

Table 2.3

Relation between NAP–SL major scientific concept areas and curriculum sub-strands

The NAP–SL major scientific concept areas	Australian Curriculum: Science – science understanding
Earth and space	Earth and space sciences
Energy and force	Physical sciences
Living things	Biological sciences
Matter	Chemical sciences

These four major scientific concept areas that guided development for the 2003 and 2006 cycles were updated for the 2009 cycle. This updated version of the concept areas guided test development for the 2012 and 2015 cycle and is included as table A1.7 in appendix 1.

The Australian Curriculum includes seven general capabilities. The capabilities, identified as being most relevant and appropriate to the assessment of science, and hence reflected in NAP–SL, included the following:

> **Literacy:** aspects of the literacy capability are found within the reading comprehension demands of both the stimuli and the items of the NAP–SL.

Numeracy: aspects of the numeracy capability are found within NAP–SL, including reading and construction of graphs and tables, calculations and measurement, as well as some elements of spatial reasoning.

Information and communication technology (ICT): aspects of the ICT capability will arise from the online delivery of the test.

Critical and creative thinking: aspects of the critical and creative thinking capability arise from important cognitive skills inherent in scientific inquiry.

Items and stimuli also drew on aspects of the personal and social capability, the ethical understanding capability, and the intercultural understanding capability, when appropriate. The following sections describe in more detail how the relevant capabilities were reflected in the 2015 NAP–SL assessment. It should be noted that the focus of the NAP–SL is the assessment of science literacy and not of general capabilities.

The 2015 NAP–SL Assessment Domain

The progress map was the key reference for test development for the 2003–2012 cycles of testing and was retained for the 2015 cycle to provide a direct connection with earlier assessments. Table A1.6 in appendix 1 includes the version of the progress map that informed test development for the most recent cycles.

In the 2015 NAP–SL cycle, the science as a human endeavour strand of the curriculum (nature and development of science; use and influence of science) informed both stimuli context in the assessment and attitudinal aspects of the student survey. Where appropriate, items were also classified against content descriptions related to this strand.

The focus of the 2015 assessment was on concepts and skills from the Australian Curriculum: Science Years 4–6. However, as the Australian Curriculum represents a continuum, concepts and skills from Foundation through to Year 6 were also considered. This is consistent with the approach taken in previous NAP–SL cycles, in which the progress map articulates a progression in development of understanding and skills.

Assessment construction and delivery

Expert item writers with past experience with NAP–SL were recruited for initial item-writing. All item writers were given an extensive briefing on the NAP–SL construct and on the changes that were being implemented for the 2015 cycle.

In the initial stages of item development, focus was placed on identifying suitable science contexts for the development of engaging stimulus. Stimulus and related items underwent an extensive initial review with feedback to item writers.

Transition to online delivery

After initial development of items by item writers, items and stimuli were authored into ACARA's online Item Authoring and Review System (IARS) platform. During this phase, graphics were developed for online delivery and items were tailored to make the most effective use of the capabilities of the online platform.

New item types

Previous cycles of NAP–SL used three basic item types:

- multiple choice that required students to choose from a set of given responses
- short constructed response that required a short response (a word, a number or a short phrase)
- long constructed response that required a more substantive response (one or several sentences).

These item types were still present in 2015 NAP–SL, but online delivery permitted a greater range of item types to be used. The list of item types is given below. Each item type is defined in terms of an 'interaction' to describe the mechanism used by the test delivery system to capture a student's response.

- Extended text: a text box is presented in which students can type text that is then typically marked by expert markers. This was used for items equivalent to the long constructed response and some items equivalent to the short constructed response.
- Text entry: a text box is presented in which students can type text. The answer is typically shorter than the answer for extended text, and this item type is often designed for machine marking. This was used for items where students were required to give a numerical response.
- Multiple choice: a set of options is presented, preceded by a 'radio button'; that is, a small circle students can click to select their response. Only one response can be selected.
- Multiple choices: similar in layout to multiple choice but with a square box in front of each response. Clicking on the box displays a 'tick' to show that a response has been selected. Students can select multiple responses. Clicking on a response the second time deselects it.
- Hotspot: a graphic divided into regions is displayed. Students choose their response by selecting a region. The interaction can be programmed to accept either a single correct answer (equivalent to multiple choice) or multiple answers (equivalent to multiple choices). This was used to present items equivalent to multiple choice or multiple choices with more complex graphical options and layouts.
- Interactive gap match: a set of words is presented, which can be dragged into marked gaps, such as a space in a sentence, a table or a diagram.
- Interactive graphic gap match: a set of pictures is presented, which can be dragged onto a larger graphic that has pre-defined regions or 'gaps'. This was used for completing diagrams and graphs.
- Select point: students can click on any point on a graphic. An invisible zone defines a correct response. This type was used for plotting points on graphs or diagrams.

Other item types, which combined features of the above, were used for special purposes. The type of interaction used for an item was carefully chosen to match the content of the item and the underlying skills being assessed. For example, graphic skills were tested using interactions such as select point or interactive graphic gap match as these item types were most effective at targeting the underlying skills.

To maintain comparability with past NAP–SL cycles, extended text and multiple choice were used more commonly than other types.

Adapting historical paper-based items to online delivery

Historical items were adapted for online display. Necessary adjustments were made to the way some stimulus material was presented and to the way students were required to respond to some items. For example, items where students had to circle a word in the paper test were changed to the multiple choice item type format in the online test. For further details regarding item adaptation for online delivery, refer to the 2015 NAP–SL Technical Report.

Inquiry task development

The 2015 NAP–SL online assessment also included a change in the way science inquiry skills were assessed. Previous NAP–SL cycles included a practical component in which students completed a practical task in groups of three and then answered items individually.

In the 2015 cycle, the assessment included an inquiry task that achieved similar objectives as the previous NAP–SL practical, but was delivered online and the stimulus material was presented via a video. Each task used the model derived from the science inquiry skills strand of the Australian Curriculum: Science in a simulated science investigation:

1. questioning and predicting
2. planning
3. conducting
4. processing and analysing.

Four online inquiry tasks were developed using this outline and piloted in a number of schools (see the ‘Inquiry task pilots’ section below). Feedback from the pilots guided modification of the tasks, which were included in the NAP–SL trial. The two tasks with the best overall psychometric performance and best balance of skills were then selected for the final study.

Use of multimedia

The proposed online test delivery platform was capable of providing multimedia stimulus material to students in the test.

In developing multimedia stimulus, the following factors were considered:

- Information was presented via audio, via text and visually.
- The videos for the assessment were custom-made rather than using stock footage.
- School-age students were featured and shown to be independently engaging in science.
- Students shown in the videos were of different genders.
- Audio was professionally recorded for maximum clarity.

Video was recorded and edited at high resolution and then compressed and resized for delivery via the online test delivery system. Because of the variety of equipment in schools and differences in available bandwidth, the videos were re-edited to improve text legibility on smaller screens.

While the availability of multimedia stimulus offered an improvement over past cycles, this had to be balanced against the technical limitations of school equipment and the unknown impact of multimedia use on test performance. Consequently, it was decided to limit the use of multimedia to one section of the test – the inquiry task. As the inquiry task was a new section with no historical link with past NAP–SL cycles, the use of multimedia in this section would have no direct impact on any historical comparison.

The inquiry task was placed after the objective test in the sequence of tasks that students had to complete. This ensured that any technical issues a school might encounter as a consequence of the greater demands of video files would only occur after students have completed most of the test.

As the video stimulus had an audio component, students were required to wear headphones when completing the inquiry task. Schools were expected to provide headphones for students who did not bring their own to school. In most cases this was not a problem but in a small number of schools, headphones were not normal equipment.

Some schools had specific technical problems accessing videos in the main study phase of the assessment. These technical problems were due to a combination of factors (see the ‘Challenges for schools’ section in chapter 3).

Review processes

All items and stimulus underwent several review stages by EAA internal review panels, ACARA curriculum specialists and SLWG members (see the 2015 NAP–SL Technical Report for more information).

At each stage of review, items and stimulus were examined against multiple criteria:

- Language demand: science stimulus may require some complex language, but it is important that the language is kept as simple as feasible.

- Scientific accuracy: the science presented needed to be correct. In some cases, complex scientific ideas were explained in a simplified way suitable for the age of the audience.
- Free from bias: items and stimulus were examined to ensure they were free from cultural or gender bias.
- Appropriate skills: the items were considered against the skills and content listed in the assessment framework.
- Metadata: the classifications of the items against multiple criteria were examined
- Item structure: the items were examined in terms of how well they were likely to perform in a psychometrically validated test.

At each review stage, comments on items were collated and tracked against subsequent edits.

Piloting and trialling items

To ensure that the assessment would be a valid and engaging experience for students and that would target the student population appropriately, a number of studies were undertaken in schools prior to the main assessment. These studies included:

- a pilot of the new inquiry task section of the assessment, designed to evaluate the effectiveness of the items and to gain early insights into running the assessment online
- a mode-effect study designed to evaluate the impact of adaptation to online delivery on historical paper-based link items
- a large-scale trial designed to evaluate the psychometric properties of the items.

Each study made use of the online delivery system and test developers were actively involved in visiting schools and observing students as they engaged with the system.

Inquiry task pilots

EAA conducted pilots of the 2015 NAP–SL inquiry tasks in November 2014. Four tasks were piloted:

1. Bouncing balls: Ellen investigates how high a ball will bounce when dropped from different heights. She plans how to collect data and record her results.
2. Mustard seeds: Jake investigates growing mustard seeds in different conditions. He looks at ways of controlling the variables in his investigation.
3. Pendulums: Ellen investigates how long it takes for a pendulum to swing from one side to another. She looks at how changing the length of her pendulum effects the time it takes to make a full swing.
4. Sunscreen: Jake investigates the effectiveness of different sunscreens. He uses photosensitive paper to set up an investigation.

A total of five schools (two government, two Catholic and one independent) participated in the pilots, with multiple sessions in each school.

The pilots allowed EAA to study several factors: the student engagement with the tasks; the practicalities of running the tasks online; and the interaction of students with the test platform.

Observations from the pilots informed:

- modifications and edits to items and stimulus in the tasks
- advice to schools involved in the NAP–SL trials and main study on how to conduct the tests
- advice to ACARA and EAA on issues relating to the test delivery platform.

Large-scale trial

The NAP–SL trial was administered online in March 2015. It had two purposes:

- to obtain item and test level data in order to inform the final item pool for the main study
- to trial the administration procedures and technology.

Students from approximately 50 schools selected from New South Wales, Queensland and Victoria participated in the trial. The trial schools were selected to reflect the range of educational contexts around the country. This included school type (government, Catholic and independent), location (metropolitan and regional), size (large and small), socioeconomic status (low and high socioeconomic areas) and language background.

Each student completed one of the eight trial objective tests and one of the four trial inquiry tasks. Tests and inquiry tasks were allocated randomly prior to the trial period so that all eight tests and all four inquiry tasks were delivered to students at each participating school.

As classroom teachers were required to administer the national sample assessment in October 2015, it was important that the trial be conducted in the same way. Classroom teachers were designated as test administrators and provided with an administration manual before the trial to allow them to familiarise themselves with the test procedures. At the completion of each session, the test administrator completed a session report form to provide feedback about various aspects of the trial. This feedback, in conjunction with a range of other sources of feedback, informed refinements to the administration manual.

Mode-effect study

Because of the necessary changes to the paper-based historical link items, it was decided that:

- the historical link items should be included in the trial of the items to ensure that these items remained effective test items after modification; and
- the historical link items would also be subject to a mode-effect study to ascertain whether there had been any general shift in difficulty of the items after adaptation to an online test delivery format.

A total of 499 students from 10 schools, selected at random from the general pool of trial schools, participated in the mode-effect study. At each of these schools students were randomly allocated to one of two test conditions:

- one group of students was given the online test consisting of the historical link items adapted for online delivery
- the other group took the paper test consisting of the original version of the historical link items.

The results showed that the overall difficulty of the online version of the link items was higher than that of the paper version of the link items. Owing to the fact that the item longitudinal equating in NAP–SL is conducted using the information about the item performance and difficulty of these link items the observed results indicated that an adjustment would be required to place the 2015 NAP–SL results onto the existing paper-based NAP–SL scale. For further detail about the study and equating methods, refer to the 2015 NAP–SL Technical Report.

It is important to note that these differences apply only to the set of adapted online test items and are therefore only applicable to the 2015 NAP–SL cycle. In addition, and equally important, the observed difference in mode of testing are not expected to occur in other NAP assessments including NAPLAN tests as they have different items and test design compared with NAP–SL tests. Further, NAPLAN online tests have been developed to maximise the comparability of test items across the online and paper tests.

The assessment

As in previous cycles, 2015 NAP–SL involved the use of seven linked tests for the final objective assessment. A cluster rotation design similar to that used in other sample-based international assessments was implemented. In the rotation design, each test is linked through common clusters to other tests. In this way a broader range of assessment items can be completed by students and linked to other items.

To achieve the cluster rotation design for NAP–SL, the items were first written in contextual units. Each unit contained one or more items that were developed around a single theme or stimulus. Clusters were then constructed by grouping three to five units. Each cluster contained approximately 13 items.

From there, tests were compiled by arranging three clusters in every test following a Balanced Incomplete Block rotation design, which reduces the possibility that an item's position in a test has an impact on its difficulty and discrimination.

In addition to an objective test, each student was allocated one of two inquiry tasks. Each inquiry task contained 10 or 11 test items.

Cluster development

Items were organised into clusters using several criteria. Each cluster covered a range of item types and included items from each of the historic NAP–SL strands as well as each of the three main Australian Curriculum: Science strands.

Three clusters contained only adapted historical link items. The other four clusters contained only new items.

Inquiry task selection

Four inquiry tasks were piloted and trialled (Bouncing balls, Mustard seeds, Pendulums, and Sunscreen). As each task had been developed as a complete sequence of steps in an investigation, only a limited number of items could be removed post-trial from any given task.

The performance of each item in a task was considered against psychometric criteria. The overall difficulty of each task was considered as well as the spread of difficulty across the items in each task. In addition to these criteria the tasks were also judged in terms of the variety of skills tested and the content areas they covered. After considering all of these factors, Pendulums and Sunscreen were selected as the two inquiry tasks for the main assessment.

Distribution of assessment item types

Items were classified by equivalence to past paper-based item type and by interaction.

The item types equivalent to paper-based were:

- multiple choice
- short constructed response
- long constructed response
- other.

'Other' was used for items with no equivalent paper-based type.

Online interaction types were classified as:

- extended text
- text entry
- multiple choice
- multiple choices
- hotspot
- interactive gap match
- interactive graphic gap match
- select point
- position object
- composite
- match.

‘Match’ item type was only used in the survey. ‘Position object’ items were developed for the trial but none were included in the main study. ‘Composite’ were used to combine multiple interactions in a single item.

This table shows the 2015 NAP–SL assessment items classified against both schemes.

Table 2.4

Interaction types and paper-based equivalents

Interaction type	Item type: paper-based equivalent				
	Long constructed response	Short constructed response	Multiple choice	Other	Total
Extended text	33	3			36
Multiple choice			46		46
Multiple choices			5		5
Hotspot			3		3
Interactive gap match				2	2
Interactive graphic gap match				6	6
Select point				2	2
Composite	6	3			9
Total	**39**	**6**	**54**	**10**	**109**

Coverage of science literacy

Items were classified against multiple criteria that connected items with classifications used in previous cycles and also with the Australian Curriculum: Science.

Coverage of skills and content using the historical framework

Items were classified against four historical concept areas. In general these concept areas were used to describe the wider context of the item including the context provided by the stimulus.

Table 2.5

Coverage of concept areas

Section	Concept area	Total
Objective test	Earth and space	22
	Energy and force	25
	Living things	24
	Matter	17
Inquiry tasks	Energy and force	12
	Living things	9
Total		**109**

Items were also classified against the historical A, B, and C strands.

Table 2.6

Coverage of A, B, C strands

Section	Strand	Total
Objective test	A	12
	B	19
	C	57
Inquiry tasks	A	7
	B	9
	C	5
Total		**109**

This table shows the overall spread of items across both sections.

Table 2.7

Coverage by concept area and A, B, C strands

Content area	Strand			Total
	A	B	C	
Earth and space	3	4	15	22
Energy and force	7	10	20	37
Living things	6	9	18	33
Matter	3	5	9	17
Total	**19**	**28**	**62**	**109**

Coverage of skills and content using the Australian Curriculum: Science

Every item developed was mapped against an Australian Curriculum code. These codes were then reviewed by ACARA's science curriculum specialists and adjusted as required.

This table shows the items by the main curriculum strands.

Table 2.8

Coverage by curriculum strands

Section	Curriculum strand	Total
Objective test	Science as a human endeavour	9
	Science inquiry skills	38
	Science understanding	41
Inquiry tasks	Science as a human endeavour	1
	Science inquiry skills	19
	Science understanding	1
Total		**109**

This table shows how the items in those strands were distributed across the Australian Curriculum: Science year levels.

Table 2.9

Coverage by curriculum strands and curriculum year level

Paper	Curriculum strand	Year 2	Year 3	Year 4	Year 5	Year 6	Total
Objective test	Science as a human endeavour		3	2	3	1	9
	Science inquiry skills	1	5	7	5	20	38
	Science understanding	1	3	11	14	12	41
Inquiry tasks	Science as a human endeavour				1		1
	Science inquiry skills	1		2	11	5	19
	Science understanding					1	1
Total		**3**	**11**	**22**	**34**	**39**	**109**

When an item was classified against a curriculum code at a year level below Year 5, it was done so as the best fit between the content of the curriculum statement and the content of the item. In some cases, an item may include content from an earlier year level but test a more sophisticated understanding of that content in a cognitively complex way.

This table shows the items by sub-strand across both sections of the test combined.

Table 2.10

Coverage by curriculum sub-strand

Curriculum strand	Sub-strand	Total
Science as a human endeavour	Nature and development of science	3
	Use and influence of science	7
Science inquiry skills	Questioning and predicting	4
	Planning and conducting	17
	Processing and analysing data and information	26
	Evaluating	6
	Communicating	4
Science understanding	Biological sciences	16
	Chemical sciences	7
	Earth and space sciences	7
	Physical sciences	12
Total		**109**

The four sub-strands used in science understanding cover four content areas broadly equivalent to those used in the historical classification. However, with the Australian Curriculum: Science classification, these four sub-strands indicate the specific content assessed in the item rather than the broader context of the item. For example, an item may have stimulus relating to biology / living things but the skill assessed may be a statement from science inquiry skills.

Intersection with general capabilities

As well as specific content domains, the Australian Curriculum also includes a set of general capabilities:

- literacy
- numeracy
- information and communication technology (ICT) capability
- critical and creative thinking
- personal and social capability
- ethical understanding
- intercultural understanding.

For the development of the 2015 NAP–SL assessment, literacy and ICT capability were regarded as background capabilities. The medium of the assessment was in English in an online environment; hence, to access the test, students needed a degree of literacy and ICT competence. However, in neither case were these capabilities intended to present a significant source of difficulty for specific items.

Personal and social capability, ethical understanding, and intercultural understanding were not specifically tracked as capabilities within the assessment. However, the stimulus associated with the items included some elements of these capabilities. In addition, there was some overlap between these capabilities and the science as a human endeavour strand of the Australian Curriculum: Science. This strand was used as the basis for the development of a small number of items in the objective test and to help frame aspects of the student survey.

Numeracy was an important related skill in the assessment. Key numeracy skills that appeared in the assessment included numerical reasoning (although complex calculations were avoided), and reading and constructing tables, graphs and diagrams, which contained numerical information.

The critical and creative thinking (CCT) capability was overtly investigated to identify connections between the CCT framework and the NAP–SL assessment. Each item was compared against the CCT framework and, when there was a strong match between the cognitive skills employed in the item and the framework, the connection was noted. In many cases there was no exact match as a given item might include a variety of cognitive skills or address critical thinking skills in a way that was not easily summarised by a single reference.

Therefore it should not be inferred from the items that were not directly matched that either they required no critical and creative thinking skills or that the skills they did employ were not present in the framework.

Table 2.11

Coverage by critical and creative thinking organising strands

CCT organising strand	Total
Inquiring – identifying, exploring and organising information and ideas	5
Generating ideas, possibilities and actions	6
Reflecting on thinking and processes	21
Analysing, synthesising and evaluating reasoning and procedures	15
General – no single match	62
Total	**109**

Survey development

In 2009, a student survey was introduced into NAP–SL. It served to gather information about students' attitudes to, and interests in, science and their science experiences in school. The students completed the survey after they finished the inquiry task. This addition to the assessment program was continued in 2012.

It was decided that the survey would continue to be included in the 2015 cycle and would be conducted online after the inquiry task. Further details related to the development of the survey can be found in chapter 8.

Historical survey items

All of the items from the 2012 survey were adapted for online delivery. The previous survey had items grouped together thematically and this grouping was repeated in the online version.

The 'Match' interaction was used for most survey items as this interaction allowed several survey items to be displayed at one time with a common set of responses in a grid.

The connection between the science as a human endeavour strand and the survey

The science as a human endeavour strand of the Australian Curriculum: Science covers some aspects of science that are best described as beliefs about science and attitudes towards science. These aspects were felt to be more appropriately covered by the survey than by the main assessment.

The survey items included in the previous cycle were compared against the science as a human endeavour strand. It was found that some aspects of the nature and development of science sub-strand were not adequately covered by the survey. To remediate this, two additional clusters of survey items were developed; these assessed the following Australian Curriculum: Science content descriptions more directly:

- Science involves testing predictions by gathering data and using evidence to develop explanations of events and phenomena.
- Science involves making predictions and describing patterns and relationships.
- Important contributions to the advancement of science have been made by people from a range of cultures.

Chapter 3

Delivering the Science Literacy Assessment

Chapter 3. Delivering the Science Literacy Assessment

The sampling procedures

As in previous cycles, the sample design for NAP–SL was a two-stage stratified cluster sample. Stratification involves ordering and grouping schools according to state, sector, size and school location. This helps ensure that all desired school types are represented in the sample.

- Stage 1 consisted of selecting schools that had Year 6 students. Within this process, the list of schools was explicitly stratified and schools were then selected with the probability of their being elected proportional to the estimated Year 6 enrolments relative to their stratum.
- Stage 2 involved the random selection of 26 students from the sampled schools selected in stage 1.

The number of students sampled in each jurisdiction was determined with two major considerations in mind:

First, it was important that results for each jurisdiction should be of similar precision. However, it was recognised that smaller sample sizes would be needed for the smaller jurisdictions (Australian Capital Territory, Northern Territory and Tasmania).

Second, the nation-wide 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 appendix 2 of this report and in the 2015 NAP–SL Technical Report.

Pre-assessment preparation

NAP–SL is a complex assessment that presents a number of logistical and organisational challenges for schools. Providing support for schools so that they could effectively participate in the assessment was a high priority.

Contacting schools

After schools had been selected in the sampling process, they were contacted in consultation with Australian Curriculum, Assessment and Reporting Authority (ACARA) and jurisdictions. Letters were sent to participating schools, advising them of their participation in NAP–SL and of the processes involved.

A website was made available to schools as an online portal. This portal provided general information including an overview of the assessment program and details of what a school would need to do to participate in the assessment.

The portal included online forms for schools to provide data on students. Schools were also asked about the numbers of computers (or equivalent devices) that were available for the assessment.

Feedback was sought from schools on their capacity to take part in the assessment. When a school indicated that there may be difficulties with them participating, this was discussed with the jurisdictional representative and ways of resolving the difficulties were investigated.

Technical readiness test and pre-practice test

An online technical readiness test was devised. This test would allow schools to individually check that a given computer (or equivalent device) could access the online assessment. The technical readiness test checked whether the device and the web browser used were compatible with the test delivery system and also provided feedback to the school on its internet access speed.

Observer feedback from the 2014 pilot study indicated that schools tended to underestimate the time it would take to get students settled at computers, logged on and ready to take the assessment. To help schools prepare for the assessment, a pre-practice test was included with the technical readiness test.

The pre-practice test included examples of different styles of interactions, video stimulus and every test navigation element that students would experience in the main assessment. The items included were intended to be easy and were designed to be both user-friendly and accessible for students.

Schools could access the technical readiness test and the pre-practice test well in advance of the main assessment. Schools were actively encouraged to use these tools to ensure that students would feel familiar and comfortable with the online environment and to confirm that technical issues would be minimised on the day of the assessment.

Helpdesk provision and online support

A frontline helpdesk was established to allow schools to speak directly with people who could assist them with NAP–SL related issues. The helpdesk was staffed during school hours for each state. A second level of technical support was provided by the agencies managing the online test delivery platform. In the event of the helpdesk being unable to assist a school with a technical issue, the issue was referred to the second level of technical support.

In addition to addressing calls from schools, helpdesk staff monitored completion of the assessments online. Schools that appeared to be having difficulties completing the assessment were contacted proactively.

The helpdesk also provided support to schools making use of the online portal and liaised with schools that had difficulties providing student background data.

Educational Assessment Australia (EAA) provided a toll-free telephone number and an email address to ensure all queries were dealt with promptly by the helpdesk.

Assessment administration procedures

The assessment was administered by the regular class teacher to minimise disruption to the normal class environment.

Standardised administration procedures were developed and published in an administration manual. Teachers and school administrators in all schools participating in NAP–SL were provided with the manual. Detailed instructions were also given in relation to the exclusion of students with a disability and students from language backgrounds other than English.

Teachers were able to review the administration manual before the assessment date and raise questions with the NAP–SL coordinator in their jurisdiction.

School observations

A quality-monitoring program was established to gauge the extent to which class teachers followed the specified administration procedures. This involved trained observers monitoring the administration of the 2015 assessment in a random sample of classes in 33 of the 628 schools involved.

Each observer attended a training session that included a background briefing on the project and details on the procedures to follow. A test observer manual was produced that contained details of test administration procedures and an extensive form to guide the writing of a test observation report.

The observers reported conformity with the administration procedures.

Online delivery of NAP in schools

Of the three areas assessed in the sample studies of the National Assessment Program, NAP–ICTL was the first to be delivered in a computer-based format in 2005. NAP – civics and citizenship was delivered online for the first time in 2013. For NAP–SL, the 2015 cycle was the first time it was delivered online.

Platforms used in schools

Data from the technical readiness test provided information on the operating systems and web browsers that were tested in schools. The platforms reported in the technical readiness test were not necessarily the same as those used in the actual assessment, but the information did provide a broad indication of the platforms used by schools.

This table summarises this information. It indicates the percentage of schools that used a given operating system on their computers.

Table 3.1

Operating systems used by schools in the technical readiness test

Operating system	Total (per cent)
Chrome OS	5.0
iPad	11.3
Mac OS	8.6
Windows	0.2
Windows 7	52.8
Windows 8	1.4
Windows 8.1	20.6
Windows XP	0.2

Note: percentages may not add up to 100 per cent due to rounding.

This table shows what percentage of schools used a given browser during the technical readiness test.

Table 3.2

Browsers used by schools in the technical readiness test

Browser	Total (per cent)
Chrome	34.2
Firefox	4.6
IE	46.7
Safari	14.5

Note: percentages may not add up to 100 per cent due to rounding.

There was some regional variation in operating systems used as shown in the following table.

Table 3.3

Operating systems used across states and territories

Operating system	State or territory (per cent)							
	NSW	Vic.	Qld	WA	SA	Tas.	ACT	NT
Chrome OS	4.2	5.9	8.5	0.2	0.2	0.0	27.0	0.0
iPad	5.0	35.3	5.6	10.6	7.2	24.9	27.4	17.1
Mac OS	18.1	9.9	5.3	13.1	13.2	38.8	0.3	0.2
Windows 7	72.7	48.8	80.6	76.2	79.4	36.3	45.3	82.8

Note: percentages may not add up to 100 per cent due to rounding.

Also there was regional variation in the browsers used.

Table 3.4

Browsers used across states and territories

Browser	State or territory (per cent)							
	NSW	Vic.	Qld	WA	SA	Tas.	ACT	NT
Chrome	36.2	29.2	25.9	42.7	30.8	31.5	52.3	28.1
Safari	8.2	25.2	6.9	14.1	14.5	12.6	26.9	10.3
Firefox	5.4	2.0	1.1	17.9	0.3	0.8	0.4	5.2
IE	50.2	43.5	66.1	25.3	54.4	55.0	20.4	56.4

Note: percentages may not add up to 100 per cent due to rounding.

Common classroom arrangements

Observations made during the pilot, the trial and the main study showed that schools accommodated the test in various ways:

- Some schools had a dedicated IT/computer room. This room was typically arranged either with computers in rows, with students facing a board, or with computers around the edge of the room, sometimes with a centre island.
- Some schools had at least one class-set of devices that could be moved from one classroom to another. In this case, the assessment was held either in a normal classroom or in a larger space (such as the school hall or library).
- Some schools had a mix of portable devices and desktop computers held in a library or learning centre. In this case, students were seated in different areas of a larger space.

- Some schools operated a 'bring your own device' policy, in which students brought their own device. Not all schools operating such a policy used student devices for the assessment but rather adopted one of the other methods.
- Some schools used a mix of methods due to an insufficient number of devices of any one kind. Students were split into two or more groups and supervised separately (for example, one group used desktop machines in a library and a separate group used notebooks in a classroom).
- In a small number of schools, there were either an insufficient number of devices for all students to take the assessment simultaneously or the schools' bandwidth was insufficient for students to access the internet simultaneously. In this case, the school ran split sessions.

Based on test-observer feedback, most schools used either a regular classroom or a dedicated IT room. The below figures are based on a small number of test observation reports; they may not be an accurate representation of the distribution across the whole sample.

Table 3.5

Room type used during NAP–SL

Type of room	Percentage of schools
Standard classroom	39
IT room	36
Library	18
Other	6

Note: percentages may not add up to 100 per cent due to rounding.

Challenges for schools

Schools faced a number of challenges in completing the assessment. The technology demands of the assessment were new for the schools and required them to ensure that they could organise sufficient devices in a space suitable for a formal assessment.

In many schools students had to go through multiple stages to log on to the assessment. These included:

1. logging onto the school network
2. logging onto an internet gateway to allow access to the web
3. opening a suitable web browser
4. navigating to the assessment platform website
5. logging onto the assessment platform.

Most students were quite adept at steps 1 and 2, but inevitably some students forgot their password or confused log-on details for one step with the details for another. In some cases, observers found that a supervising teacher had to spend several minutes with a student to log on, which delayed the start of the assessment for all other students.

Technical issues faced by schools

The technical issues experienced by schools were of two main types:

1. Students were locked out of the assessment because they had not logged off correctly during the break between the objective test and the inquiry task. A security measure designed to prevent two students logging into the same account and a separate measure that automatically logged students out after a defined period of inactivity resulted in some students being unable to log back on to the assessment if they did not follow the recommended steps. This issue was quickly resolved by a phone call to the helpdesk as helpdesk staff had direct access to the system that managed student logins. Schools were informed about the steps required to avoid or resolve this issue.

2. Some schools experienced difficulty loading videos for the inquiry task. During the trial, this problem was caused by students at a school simultaneously attempting to access the inquiry task. Steps were taken to alleviate this situation for the main study. However, during the main study, a small number of schools experienced a difficulty accessing the videos despite successfully doing so in the technical readiness test and pre-practice test.

There were other issues that were primarily self-managed by schools:

- ensuring all devices used had up-to-date web browsers
- ensuring that peripheral devices such as keyboards and mice were functional for all machines
- ensuring that headphones were available and that headphone jack ports on devices were operational.

USB delivery

For a small number of schools, the level of internet access available was not adequate for the test delivery system to work fully online.

To accommodate these schools, a version of the test delivery system was developed that operated from a USB flash memory device, which could be plugged into a computer. If a low-bandwidth internet connection was available, then the local USB version of the test would upload student responses. This was feasible because the bandwidth required to upload student responses was less than that needed to access the test online.

If no internet connection was available at all, then student responses were captured directly onto the USB device. These USB devices were then returned securely and student data were incorporated into the overall set of student responses.

A total of six schools completed the assessment using USB devices.

Student experience of the test

Once logged on to the test delivery platform, students were presented with four icons that represented each stage of the assessment.

Figure 3.1

Icons shown to students

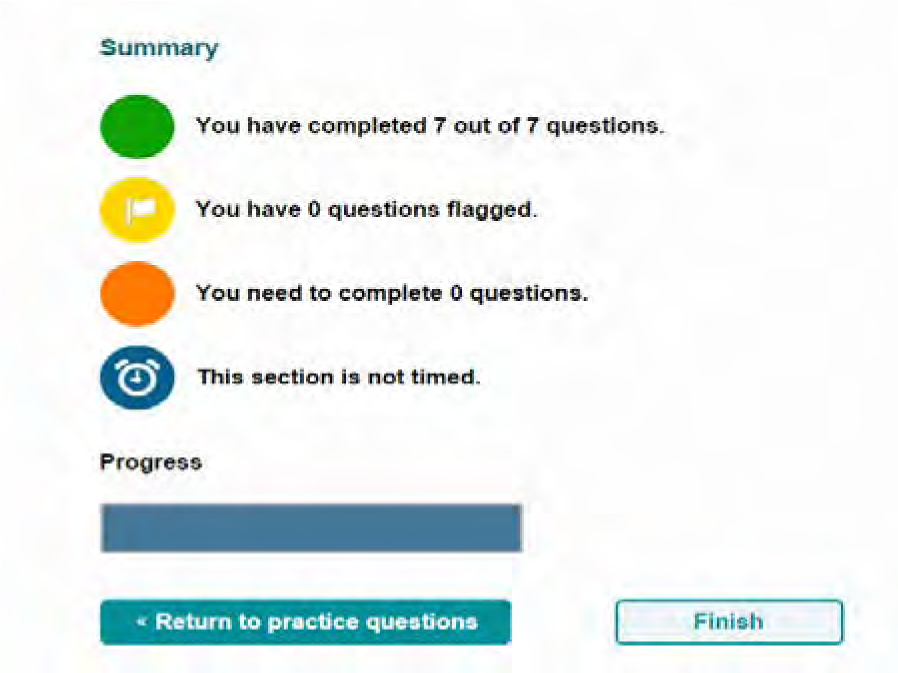
Students had to take each step in order:

1. Practice items: students completed a set of simple practice items that gave them an opportunity to familiarise themselves with the online test delivery system. Students were given approximately 10 minutes to complete this step.
2. Objective test: students completed one of the seven tests. Students had 60 minutes to complete this step. This step was timed automatically by the test delivery system.
3. Inquiry task: students completed one of the two inquiry tasks. Students had 35 minutes to complete this step, which included time to watch the accompanying video stimulus. This step was timed automatically by the test delivery system.
4. Student survey: students had approximately 10 minutes to complete the survey.

When students completed a step, the system showed them a summary of what they had done. This image shows the summary that appeared at the end of the practice items.

Figure 3.2

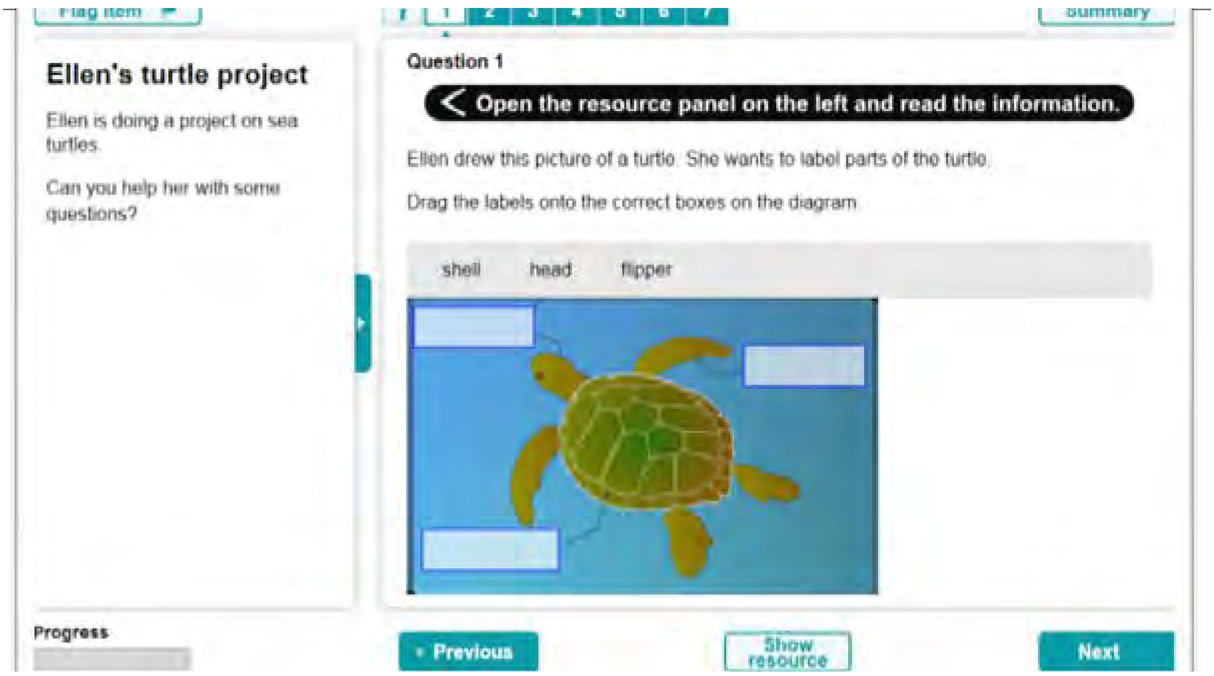
Summary screen shown to students



The test delivery system also included navigation features, the capacity to flag items that they wished to review, a progress bar and, for timed steps, a timer. This image shows the typical screen layout using the first of the practice items.

Figure 3.3

Sample screen layout of a practice question in the test delivery system



Student feedback collected by observers in the pilot, trial and main study was largely positive. The main problems reported by students regarding the assessment were technical issues such as slow loading of videos.

Marking of responses to open-ended items

Over a half of the items were open-ended and required marking by trained markers. Marking guides were prepared by EAA and refined during the trialling process.

Development of marking guides

Marking guides were developed in an incremental process. For constructed response items, item writers were required to indicate what they regarded as correct responses and likely incorrect responses. This information was reviewed along with the item during the review processes.

Before the start of trial marking, actual student responses from the pilot were reviewed. Responses that illustrated key distinctions between correct and incorrect responses (or between different levels of response) were identified. These examples were then incorporated into the trial marking guide.

The trial marking guide was further refined using student responses during the trial. Amendments, which reflected the decisions taken during the trial marking, were incorporated into a final marking guide for the main study.

Online marking

The marking team included experienced markers employed by EAA. Most markers had marked NAP–SL assessments in previous cycles.

As a part of the changes resulting from the online delivery of NAP–SL in the 2015 cycle, the marking process was also completed online. Marking was conducted using an online marking system designed to work in tandem with the test delivery system. The marking system allowed for student responses to be double-marked and for sampling of responses.

In simple terms, the markers were presented with the stimulus and stem of each item together with the student response on their computer screen. They then simply clicked on a button to award a mark and move on to the next student's response.

The training of the markers took place in stages. All the items that required expert marking were divided into five groups. The items associated with the two inquiry tasks comprised two of the groups with the objective test items split into three further groups.

The team leaders underwent a half-day training day presented by the marking professional leader who had a major role in developing the marking guides. During this training, the usual procedure was followed, whereby the marking guide for each item was explicated and the professional leader responded to any questions from the team leaders. The session involved a brief formal presentation followed by hands-on practice with pre-marked sample student responses. At the end of the session, the team leaders marked the same set of student answer responses. The scores were compared with those agreed to by expert scorers (the project director, the test development manager and the professional leader). This process gave greater scope for discussion and also resulted in two team leaders being proficient in marking those items even though only one would actually be marking them. This provided extra support for team leaders during the marking operation.

On the first day of marking, the team leaders trained the markers one item at a time. This was effective in building ownership of the marking guides among the team leaders. Training followed the same format as that of the team leader training.

Team leaders discussed with markers agreements and disagreements between their scores and the scores given by expert scorers. Additional practice was provided to markers for items where consistency and/or accuracy were low.

Markers were monitored for reliability by having samples of their marking check-marked by team leaders. In cases where there were differences between markers' and team leaders' markings, the scoring was reconciled in consultation with the professional leader. In addition, a part of the way through marking each item, all markers were asked to mark the same set of student answer responses. The scores were compared with the scores agreed to by expert scorers and any differences were discussed and reconciled.

In addition, approximately five per cent of the 2012 NAP–SL historical item responses were also marked by the 2015 markers to ensure the reliability of marking. These procedures, coupled with the intensive training at the beginning of the marking exercise, ensured that markers applied the scoring criteria consistently and accurately.

Data quality procedures

Student response data and marks were exported from the test-delivery / test-marking system. These data were then supplied to EAA, and EAA then reformatted the data for use in psychometric analysis software and school reporting software. Checks were undertaken at this stage to ensure that the data had been correctly exported.

School summary reports

Schools that participated in NAP–SL were provided with feedback about the performance of their students on the assessment before the close of the 2015 school year. The reports showed the results for each student on an item-by-item basis with comparative data showing the percentage of the school and the national sample of students responding correctly to the item. In the case of items that were worth more than one mark, the percentage of students achieving the maximum score on the item was provided.

For the first time, the Australian Curriculum: Science reference codes were provided in the school reports to help schools identify the key curriculum aspects that had been assessed. Item descriptors were written using the content of the Australian Curriculum: Science as a starting point so as to more clearly identify the relevant aspect of the curriculum assessed.

NAP–SL School Release Materials

In past paper-based cycles of NAP–SL, a set of items that is representative of the whole objective test has been identified and made publicly available in the form of an assessment. This has typically been in the form of a downloadable document that can be printed as a test booklet. In addition, a guide for administering the test, marking the student responses and converting student scores into levels on the NAP–SL scale is also provided. These post-sample school release materials can be found on the NAP website.

For this cycle, an online demonstration test will be made available to all schools via ACARA's NAP website.

The main purpose of this test is to familiarise students, teachers and parents with NAP–SL online item types and the functionalities of the student test player. Although the test will include a selection of items covering the range of science literacy strands, concept areas and item difficulties, it will not reflect all components of the test itself, as the platform is being refined and improved in preparation for future online assessments.

Chapter 4

Student Performance in Science Literacy in 2015

Chapter 4. Student Performance in Science Literacy in 2015

Introduction

In this chapter, summary statistics for the 2015 NAP sample assessment – science literacy (NAP–SL) are shown in terms of student mean scores and distributions of scores by state and territory. In addition, an overview of the methodology used to construct the science literacy scale for reporting the results of NAP–SL is provided. This chapter also contains a comparison of the performance of Year 6 students over the 2006, 2009, 2012 and 2015 assessment cycles.

Science literacy scale

A science literacy scale was constructed in 2003, using the Rasch measurement model. The Rasch analysis produced information about the relative difficulty of the assessment items, as well as information about students' abilities. These data were located on a continuum to form the science literacy scale, and a national mean was set at 400 with a standard deviation of 100.

After the second NAP–SL cycle in 2006, it was decided to use the results of the 2006 assessment to reconstruct the science literacy scale. The reasons for reconstructing the scale were:

- The 2006 test design was more robust than the 2003 test design.
- There were considerably more items in 2006 than in 2003, resulting in a better coverage of the assessment domain in 2006.
- The 2006 items generally discriminated better than the 2003 items.
- The 2006 sampling was more comprehensive, as remote schools were also included in the sample (see the 2006 NAP–SL Technical Report for more information).

The Rasch measurement model was used to estimate item difficulties and student abilities. Parameters used in the analysis included test form and a set of relevant student characteristics (for example, gender, jurisdiction and school location). The plausible values methodology was utilised to obtain population estimates of student abilities (for detailed information, see the 2015 NAP–SL Technical Report). These results were then mathematically transformed to construct the science literacy scale that has a mean of 400 and a standard deviation of 100. In the remainder of this report, all references to the science literacy scale are to the 2006 reconstructed scale.

Establishing proficiency levels

One of the main objectives of NAP–SL is to monitor trends in science literacy performance. One convenient and informative way of doing so is to reference students' results to the proficiency levels. Typically, students whose results are located within a particular proficiency level can demonstrate the understandings and skills associated with that level and possess the understandings and skills of lower proficiency levels.

As described in chapter 1, NAP–SL covers a range of five proficiency levels: Level 2, Level 3.1, Level 3.2, Level 3.3 and Level 4. In 2006, proficiency levels were assigned corresponding to cut-points on the science literacy scale.

The proficient standard in science literacy was set at the boundary between Level 3.1 and Level 3.2. This means that students who obtain a score equal to, or above, the Level 3.2 cut-point of 393 are deemed to have attained the proficient standard in science literacy. This cut-point is used for each assessment cycle.

An overview of 2015 results relative to the distribution of student scores in proficiency levels, as well as information about the proportion of students who attained the proficient standard, are presented in chapter 5 and chapter 6 of this report.

Analysing the 2015 results

The measurement model for analysing student responses in 2015 was the same as that of the previous cycles. The common item equating methodology was used to place the 2015 results on the science literacy scale. Additionally, an adjustment has been made to account for the item difficulty changes for historical NAP–SL link items due to moving from a paper-based test to an online test.

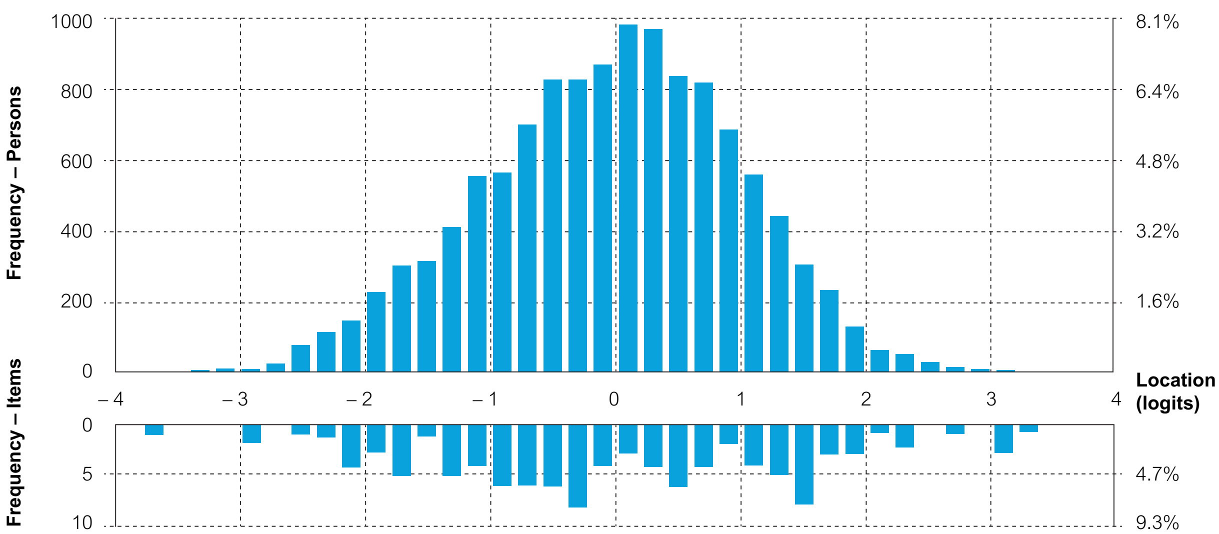
In 2015, a total of 38 link items were used to equate the 2015 results to the existing NAP–SL scale. This included 12 link items from the 2006 assessment, seven link items from the 2009 assessment, and 19 link items from the 2012 assessment. These link items covered a range of science literacy strands, concept areas and item difficulties. This set of 38 potential link items was evaluated and used to equate the 2015 item and student parameters to the science literacy scale (for detailed information about the link items, see the 2015 NAP–SL Technical Report).

Before presenting data for the application of the Rasch measurement model, it is important to ensure that the test has appropriately targeted the student population.

As can be seen from figure 4.1, the 2015 assessment achieved a good spread of item difficulties and was appropriately matched to the Year 6 cohort. This demonstrates that the items targeted the population well and were able to discriminate between achievements at the highest level, while still catering for less able students.

Figure 4.1

Distribution of items and students



Achievement by state and territory in 2015

Age of students

The average age of students who participated in NAP–SL in each state or territory is not the same.

Table 4.1

Average ages of students in the sample by state and territory 2015 and 2012

State/territory	Average age at time of testing 2015 NAP–SL	Average age at time of testing 2012 NAP–SL
NSW	12 years 1 month	12 years 0 months
Vic.	12 years 3 months	12 years 3 months
Qld	11 years 11 months	11 years 5 months
WA	11 years 10 months	11 years 9 months
SA	12 years 1 month	12 years 0 months
Tas.	12 years 4 months	12 years 3 months
ACT	12 years 1 month	12 years 0 months
NT	11 years 11 months	11 years 10 months

From the table, it can be seen that the average age of students varies between states and territories, with Western Australia having the youngest students on average. However, this range is smaller than was seen in previous cycles.

Reading the column graphs

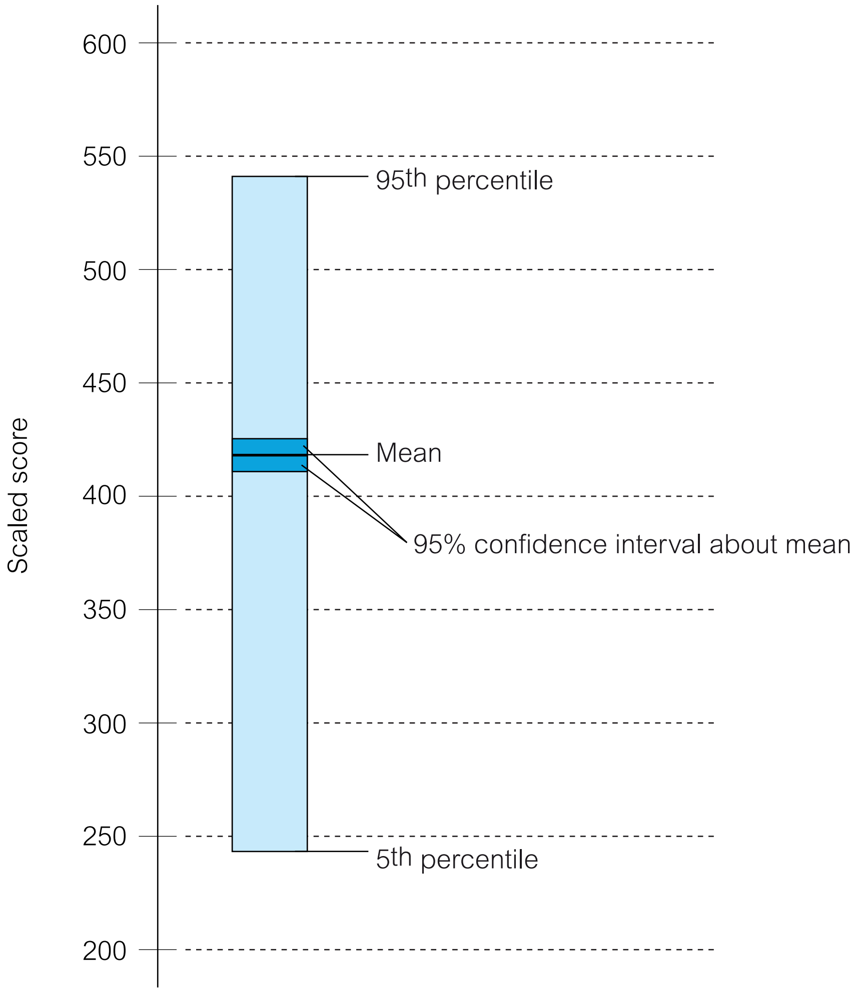
Figure 4.2 is an example of a column graph used to display the scaled mean scores and distributions for states and territories. The vertical bar shows the range of student performance.

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

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

Figure 4.2

Sample column graph



Located in the middle region of the bar is a darker blue band that contains a thin horizontal black line. This black line denotes the mean score, while the blue regions on either side represent a confidence interval, which gives an indication, through the width of the band, of the level of accuracy with which the mean was measured.

Given that NAP–SL is a sample-based assessment, the reported means are estimates of a true population mean (the mean that would be measured if the complete population of Year 6 students in Australia could be assessed). The confidence interval reported for the mean gives the range that is likely to contain the value of the true population mean.

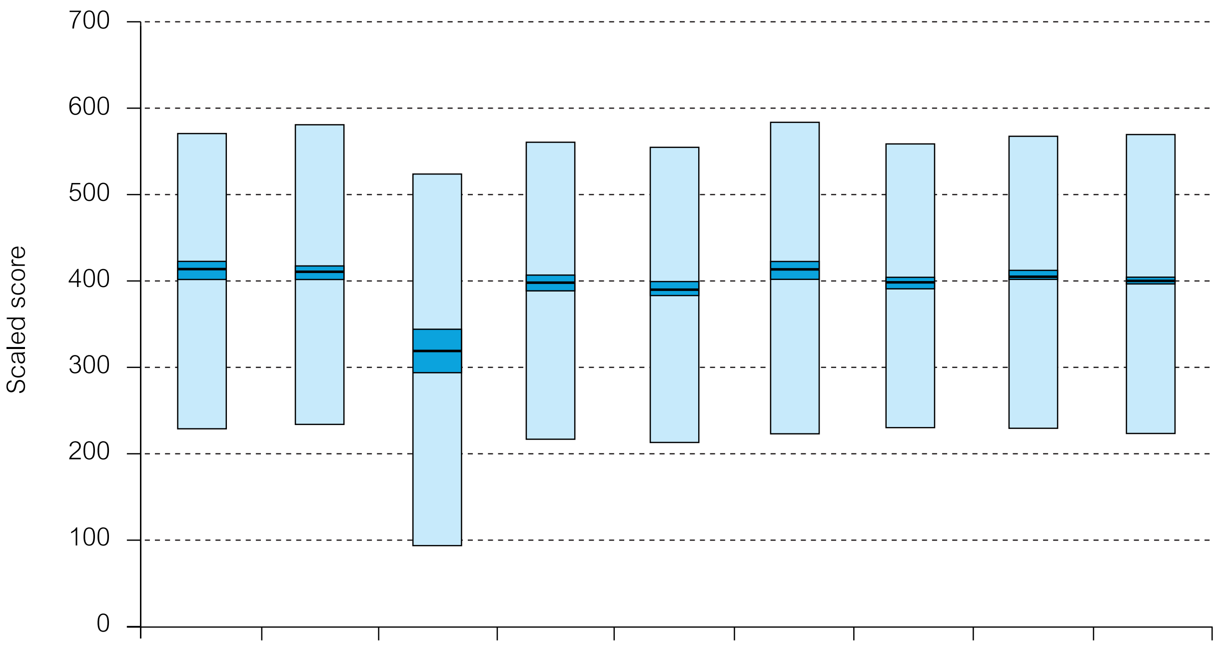
Confidence intervals in this report were constructed with an estimate of statistical precision. This estimate of precision is such that a constructed confidence interval will contain the true population mean 95 per cent of the time, for a set of random repeated measurements. Such a confidence interval is referred to as the 95 per cent confidence interval.

Mean and range of student scores in 2015

Figure 4.3 shows student performance in science literacy for each state and territory in 2015. The bars show the range of scores for each state and territory, which were achieved by the middle 90 per cent of the population. Shaded bands within each bar mark the confidence interval around the corresponding mean. Any interpretation of results needs to be made by considering the relative precision (as indicated by the size of the confidence intervals) of the reported estimates of the student mean achievement.

Figure 4.3

2015 distribution of Year 6 performance by state and territory



	ACT	NSW	NT	Qld	SA	Tas.	Vic.	WA	AUST
Mean	414	411	320	398	392	414	399	408	403
95% CI	±12.1	±8.6	±25.6	±10.6	±8.8	±11.7	±8.9	±7.5	±4.3

It can be seen that the Northern Territory had the widest spread of scores achieved by the middle 90 per cent of students, the largest confidence interval around the mean score, and the lowest mean score of all the states and territories. All other jurisdictions have relatively similar widths of score range and confidence interval. This suggests a more heterogeneous sample in the Northern Territory compared with other jurisdictions. This is likely due to the particular demographic features of the Northern Territory.

Extra information about the range of student scores in 2015 is provided by listing the scaled scores corresponding to the standard range of percentile values by each state and territory in table 4.2. The table shows that the Northern Territory has lower percentile scores than all other jurisdictions.

Table 4.2

Distribution of percentile scores by state and territory in 2015

State/ territory	Mean score	95 per cent confidence interval	Percentile						
			5th	10th	25th	50th	75th	90th	95th
NSW	411	±8.6	237	273	338	412	484	546	582
Qld	398	±10.6	217	256	329	405	472	531	563
Vic.	399	±8.9	230	267	330	403	471	525	559
WA	408	±7.5	230	267	338	413	480	536	569
SA	392	±8.8	219	257	321	395	463	523	556
Tas.	414	±11.7	222	265	338	419	494	552	585
ACT	414	±12.1	232	276	347	420	487	540	572
NT	320	±25.6	97	140	224	331	419	488	525
Aust.	**403**	**±4.3**	**225**	**264**	**332**	**407**	**476**	**535**	**569**

Comparisons of means by state and territory in 2015

Table 4.3 shows comparisons of each state and territory against each other. For each pair, the difference in mean scores is tested for statistical significance using the Bonferroni adjustment. By reading across the rows, it is possible to draw a comparison between any two jurisdictions. Comparisons that are statistically significant are shown by an upward or downward symbol.

The Bonferroni adjustment to statistical significance testing is conducted in order to account for the possibility that a difference can be deemed to be statistically significant by chance when multiple comparisons are conducted using the same data. The Bonferroni adjustment increases the strictness of the criterion for establishing statistical significance relative to a pair-wise comparison, hence making it harder to claim that a difference is statistically significant.

It can be seen in table 4.4 that when the Bonferroni adjustment was implemented, students from the Northern Territory achieved a significantly lower mean score than students in all other states and territories.

These results are in contrast with the 2012 cycle pair-wise comparison analysis which highlighted a greater number of statistically significant differences between jurisdicitons.

Table 4.3

Multiple comparisons of science literacy results by state and territory for 2015 (with the Bonferroni adjustment)

	Mean score	95% CI	ACT	NSW	NT	Qld	SA	Tas.	Vic.	WA
ACT	414	±12.1		●	↑	●	●	●	●	●
NSW	411	±8.6	●		↑	●	●	●	●	●
NT	320	±25.6	↓	↓		↓	↓	↓	↓	↓
Qld	398	±10.6	●	●	↑		●	●	●	●
SA	392	±8.8	●	●	↑	●		●	●	●
Tas.	414	±11.7	●	●	↑	●	●		●	●
Vic.	399	±8.9	●	●	↑	●	●	●		●
WA	408	±7.5	●	●	↑	●	●	●	●	

↑	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

Comparisons of student results in 2006, 2009, 2012 and 2015

The 2015 NAP–SL was the fifth time the science domain had been assessed in the national assessment programs, with the first assessment carried out in 2003.

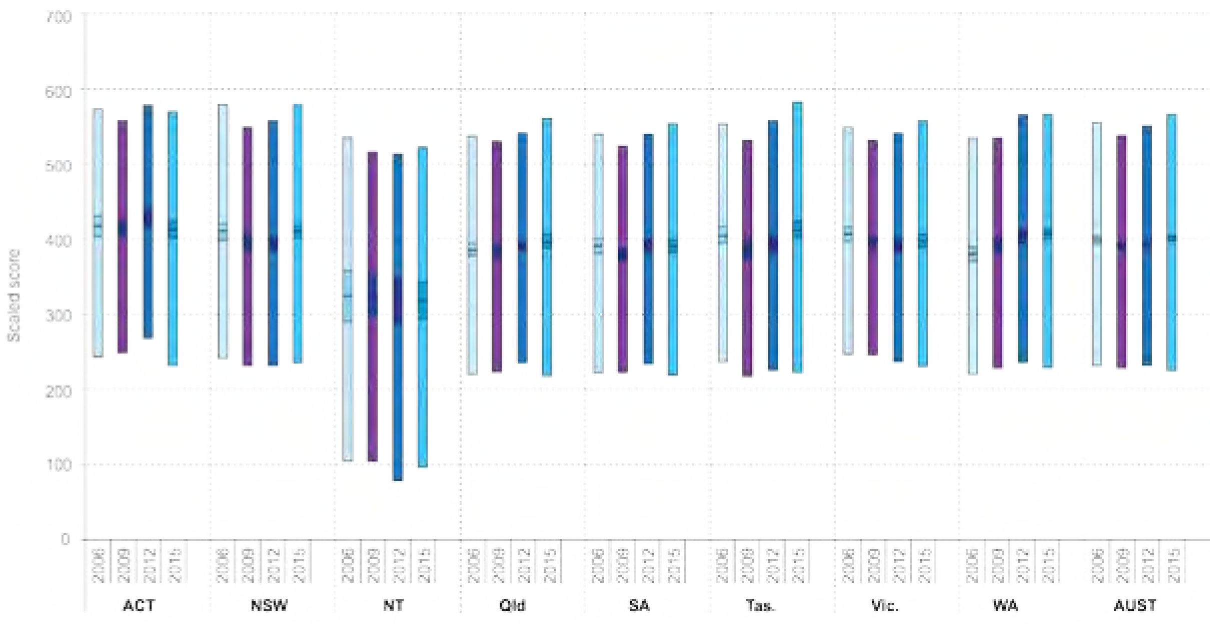
The 2003 assessment differed from the 2006 assessment in terms of the item booklet design, the sampling plan and the number of items. Because of these differences, it was decided to use the 2006 results to construct the science literacy scale. Tests of statistical difference, therefore, are only conducted between the 2006 and later assessment cycles.

To test whether the 2015 results differ from those of previous assessment cycles, link errors were added to the standard error estimate in a simple pair-wise test of statistical significance (for further information regarding the link error, refer to the 2015 NAP–SL Technical Report).

Figure 4.4 contains graphical comparisons of the student results in the middle 90 per cent of the distribution. Shaded bands around the mean within each bar mark the 95 per cent confidence interval.

Figure 4.4

Comparison of distributions of student scores by state and territory in 2006, 2009, 2012 and 2015



Note: Owing to the fact that the NAP-SL scale is designed to have mean of 400 (and standard deviation of 100) and that NT have students who perform at more than five units of standard deviation below such a mean, the fifth percentile point for NT 2006 is a negative score. For the graphical presentation purposes the tenth percentile (a scaled score of 104) is used as the lowest data point for NT.

As depicted in figure 4.4, overall the 2015 means at the national level and at the state and territory level were about the same as those of 2006. The exception was Western Australia, which showed an increase of 27 points in 2015 compared with the mean student achievement in 2006.

Table 4.4 shows the 2006, 2009, 2012 and 2015 mean scores by state and territory and indicates whether the differences in means between 2012 and 2015, between 2009 and 2015, and between 2006 and 2015 are statistically significant.

Table 4.4

Comparison of 2006, 2009, 2012 and 2015 jurisdiction mean scores

State/ territory	Mean score				Change and statistical significance					
	2006	2009	2012	2015	2012–2015		2009–2015		2006– 2015	
NSW	411 (±12.5)	396 (±12.1)	395 (±9.9)	411 (±8.6)	16	NO	15	NO	0	NO
Vic.	408 (±10.2)	398 (±9.2)	393 (±9.7)	399 (±8.9)	6	NO	1	NO	–9	NO
Qld	387 (±8.6)	385 (±8.9)	392 (±6.4)	398 (±10.6)	6	NO	13	NO	11	NO
WA	381 (±10.0)	393 (±9.6)	406 (±9.5)	408 (±7.5)	2	NO	15	NO	27	YES
SA	392 (±10.0)	380 (±10.4)	392 (±7.9)	392 (±8.8)	0	NO	12	NO	0	NO
Tas.	406 (±12.1)	386 (±13.5)	395 (±12.3)	414 (±11.7)	19	NO	28	YES	8	NO
ACT	418 (±14.3)	415 (±10.6)	429 (±13.2)	414 (±12.1)	–15	NO	–1	NO	–4	NO
NT	325 (±33.7)	326 (±28.6)	319 (±31.1)	320 (±25.6)	1	NO	–6	NO	–5	NO
Aust.	**400 (±5.4)**	**392 (±5.1)**	**394 (±4.4)**	**403 (±4.3)**	**9**	**NO**	**11**	**NO**	**3**	**NO**

Note: Figures in parentheses refer to 95 per cent confidence intervals. Mean scores have been rounded.

As can be seen in table 4.4, there is no statistically significant difference between 2012 and 2015 results at the national level or across the states or territories.

Similarly, the comparison between 2006 and 2015 results shows that there is no statistically significant difference at the national level. The only exception at the jurisdiction level is Western Australia for which the increase of 27 points in student mean achievement from 2006 to 2015 is statistically significant due to cumulative gains achieved in 2009 and 2012.

There is no significant difference in the national level results when comparing 2015 results to 2009. However, at the jurisdiction level, Tasmania’s results have steadily improved since then. The 2015 mean for Tasmania is significantly higher than that observed in 2009.

Ranking of jurisdictions by mean scores

Table 4.5 shows a jurisdiction-by-jurisdiction comparison of the mean scores in rank order for 2006, 2009, 2012 and 2015.

Table 4.5

State and territory mean score rankings in 2006, 2009, 2012 and 2015

Rank	2006		2009		2012		2015	
	State/territory mean score		State/territory mean score		State/territory mean score		State/territory mean score	
1	ACT	418 (±14.3)	ACT	415 (±10.6)	ACT	429 (±13.2)	ACT	414 (±12.1)
2	NSW	411 (±12.5)	Vic.	398 (±9.2)	WA	406 (±9.5)	Tas.	414 (±11.7)
3	Vic.	408 (±10.2)	NSW	396 (±12.1)	Tas.	395 (±12.3)	NSW	411 (±8.6)
4	Tas.	406 (±12.1)	WA	393 (±9.6)	NSW	395 (±9.9)	WA	408 (±7.5)
5	SA	392 (±10.0)	Tas.	386 (±13.5)	Vic.	393 (±9.7)	Vic.	399 (±8.9)
6	Qld	387 (±8.6)	Qld	385 (±8.9)	SA	392 (±7.9)	Qld	398 (±10.6)
7	WA	381 (±10.0)	SA	380 (±10.4)	Qld	392 (±6.4)	SA	392 (±8.8)
8	NT	325 (±33.7)	NT	326 (±28.6)	NT	319 (±31.1)	NT	320 (±25.6)

Notes: Figures in parentheses refer to 95 per cent confidence intervals. Mean scores have been rounded. Ranking is based on mean scores before rounding. Nevertheless, the change in ranking order for the states and territories should be regarded as indicative only given that the differences in mean achievement between states and territories in 2015 were generally not statistically significant.

It can be seen from table 4.5 that based on the mean score for students, Tasmania changed from fifth rank in 2009 to third rank in 2012 and to second rank in 2015. Mean performance for students from Western Australia is comparable from 2012 to 2015 but significantly improved since 2006. The table shows that Western Australia changed from seventh rank in 2006 to fourth rank in 2015. Although changes in mean performance for all other states and territories were not significant, the table shows that New South Wales moved up one place (from fourth to third rank), Queensland also moved up one place (from seventh to sixth rank), South Australia moved down one place (from sixth to seventh rank) and Victoria stayed in the same place (fifth) between the 2012 and 2015 tests. The Australian Capital Territory and the Northern Territory have maintained the same ranking since 2006. **It is important to note that any changes in ranking order for states and territories be regarded as indicative only given that the differences in mean achievement between states and territories in 2015 were generally not statistically significant—claim of significant shifts in student achievement among States and Territories is neither intended nor implied.**

Trends in mean achievement in science literacy

An overview of the trends in science literacy at the national level for 2003, 2006, 2009, 2012 and 2015 is provided below in table 4.6. However, as tests of statistical significance between 2003 and 2015 results were deemed not to be sound owing to the reasons detailed at the beginning of this chapter, the 2003 results are indicative only.

Table 4.6

Trends in mean scores in science literacy in 2003, 2006, 2009, 2012 and 2015

Aust.	Mean score
2003	409 (±3.7)
2006	400 (±5.4)
2009	392 (±5.1)
2012	394 (±4.4)
2015	403 (±4.3)

Notes: Figures in parentheses refer to 95 per cent confidence intervals. 2003 results are indicative only.

Summary

The 2015 results are similar to those of earlier years both in terms of student mean achievement and the distribution of student scores. The analysis shows that there are no statistically significant differences between the 2012 and 2015 results at jurisdictional levels and at the national level.

The analysis also shows that there are no statistically significant differences between the 2006 and 2015 results for most jurisdictions and at the national level. The only exception is Western Australia, where the mean student achievement obtained in 2015 represented a significant increase from 2006 levels, despite results remaining relatively stable between 2012 and 2015. In addition, Tasmania has shown significant cumulative improvement from 2009 to 2015.

Chapter

5

Interpreting the Science Literacy Results

Chapter 5. Interpreting the Science Literacy Results

Introduction

Chapter 4 showed students' score distributions on the science literacy scale. The results can also be referenced directly to the assessment domain, through the items comprising the tests, to reveal the understandings and skills demonstrated by students.

For the purposes of this report, the science literacy scale has been split into levels called 'proficiency levels'.

This chapter discusses the establishment of the proficiency levels and the cut-off scores for each of the levels, and provides examples of items that illustrate the skills and knowledge for each level.

Establishing proficiency levels

One of the main objectives of the NAP sample assessment – science literacy (NAP–SL) is to monitor trends in science 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 can demonstrate the understandings and skills associated with that level as well as the understandings and skills of lower proficiency levels.

Initially, 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 the 2003 assessment, three further proficiency levels within Level 3 were created, providing five levels for reporting student performance in the assessment.

The cut-off points, which denote the boundaries between the levels, were established using a combination of experts' knowledge of the skills required to answer each science literacy item and information from the analysis of student responses.

The difficulty range spanned by each level is such that students whose scores are at the top of a level have 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 in that level correctly. On average, these students would be expected to answer correctly about 76 per cent of the items in that level.

Students who are at the bottom of a level have 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 correctly about 50 per cent of the items in that level.

The cut-off scores for each level are shown in table 5.1. The same cut-off scores have been used to determine the proficiency levels for past NAP–SL cycles.

Table 5.1

Proficiency level cut-off scores

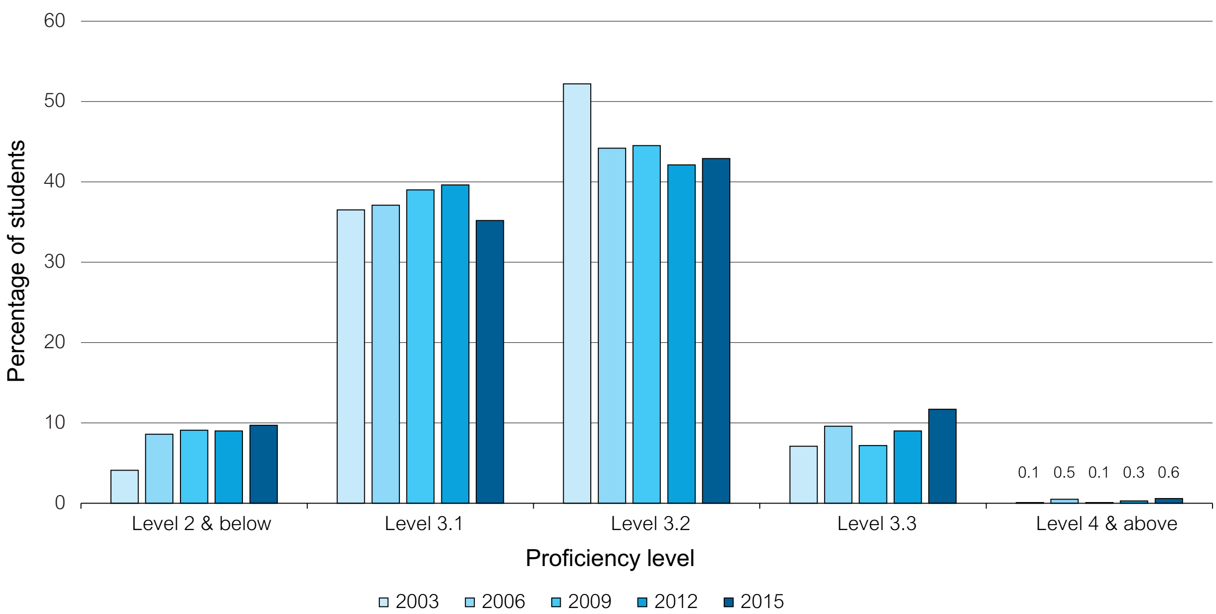
Level	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above
Cut-off score	262	393	523	653	
Percentage of students at that level	9.7	35.2	42.9	11.7	0.6%

A score greater than 653 locates students in proficiency level 4 and above. Scores in the range of 262 to 653 relate to proficiency level 3 on the assessment framework.

Figure 5.1 shows the distribution of students in these proficiency levels for each of the NAP–SL cycles from 2003 to 2015.

Figure 5.1

Distribution of students across proficiency levels 2003, 2006, 2009, 2012, 2015



Describing proficiency levels

Appendix 1 provides the descriptions of the understandings and skills required of students at each proficiency level. The descriptions come from the science literacy assessment domain presented in appendix 1, but Level 3 has been divided into sub-levels 3.1, 3.2 and 3.3.

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.

At each proficiency level, a wide range of items that varied in context, format and difficulty were used to give students the best opportunity to provide evidence of what they knew and could do.

Only a small number of items have been released in this report; other items have been retained as secure link items for scaling purposes in future cycles of NAP–SL.

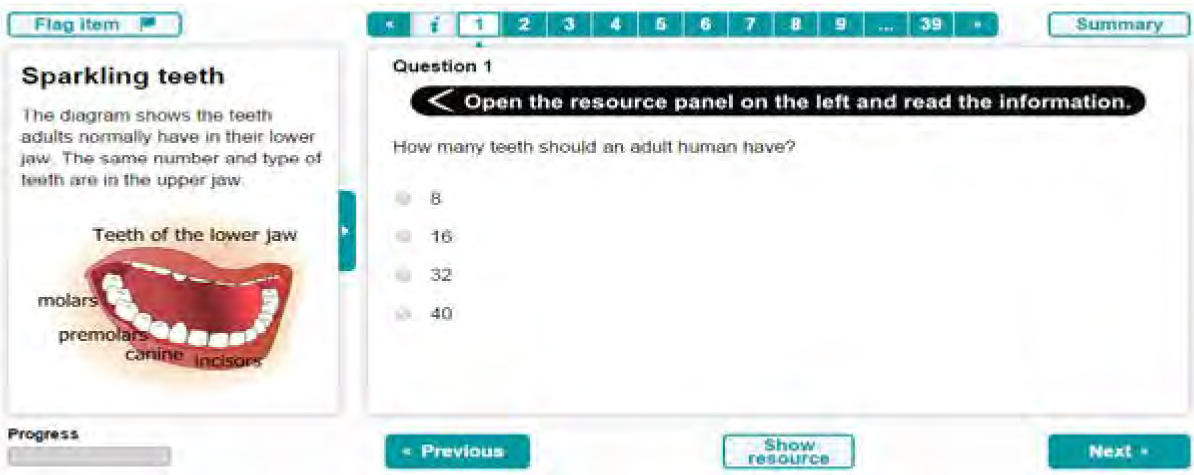
In the online test, students could expand the smaller preview version of the stimulus to see a larger version. In the examples given below, the full stimulus provided to students is not shown.

Sample items illustrating performance within proficiency level 2 and below

Item 1 in the item set Sparkling teeth (figure 5.2) illustrates performance at Level 2 or below.

Figure 5.2

Sparkling teeth Q1



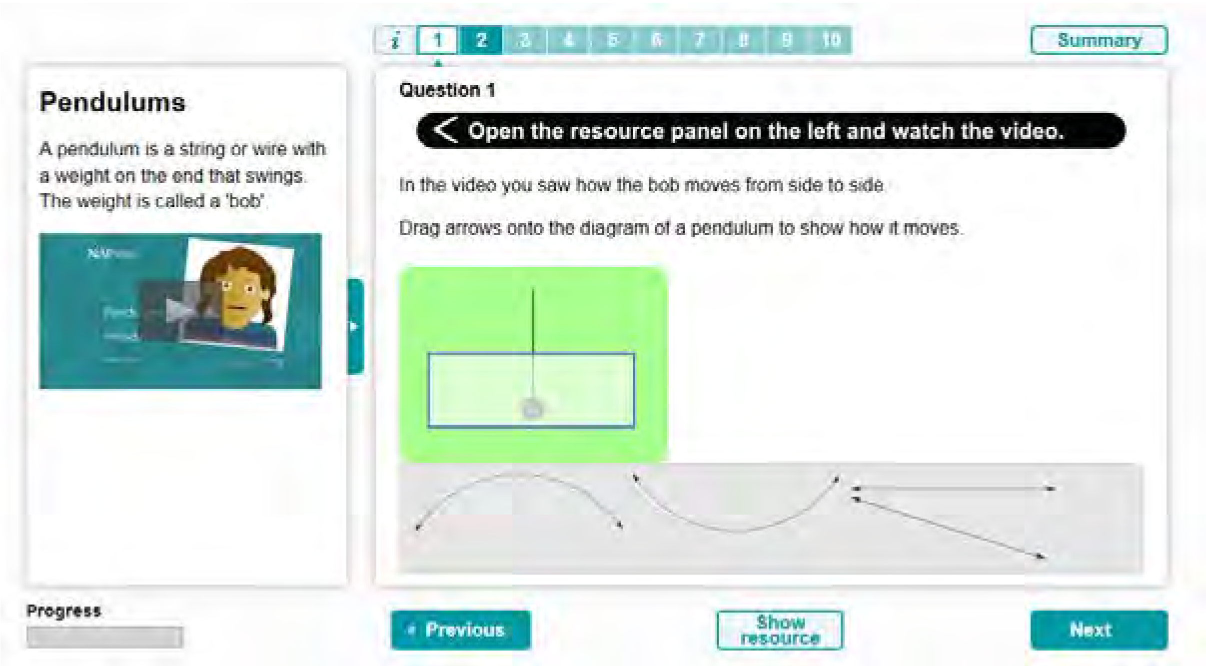
Type	Multiple choice		**Australian Curriculum**	Science understanding Biological sciences	
Strand	B		**Concept area**	Living things	
Level	2 and below	**Percentage correct**	84	**Science literacy scale location**	257

Students were provided with a diagram showing the number and types of teeth in an adult human lower jaw. Students were required to count the number of teeth in the bottom jaw and use that to calculate the total number of teeth an adult human has.

Item 1 in the inquiry task Pendulums (figure 5.3) is another illustration of an item at Level 2 or below.

Figure 5.3

Pendulums Q1



Type	Interactive graphic gap match		**Australian Curriculum**	Science inquiry skills Communicating	
Strand	C		**Concept area**	Energy and force	
Level	2 and below	**Percentage correct**	90	**Science literacy scale location**	193

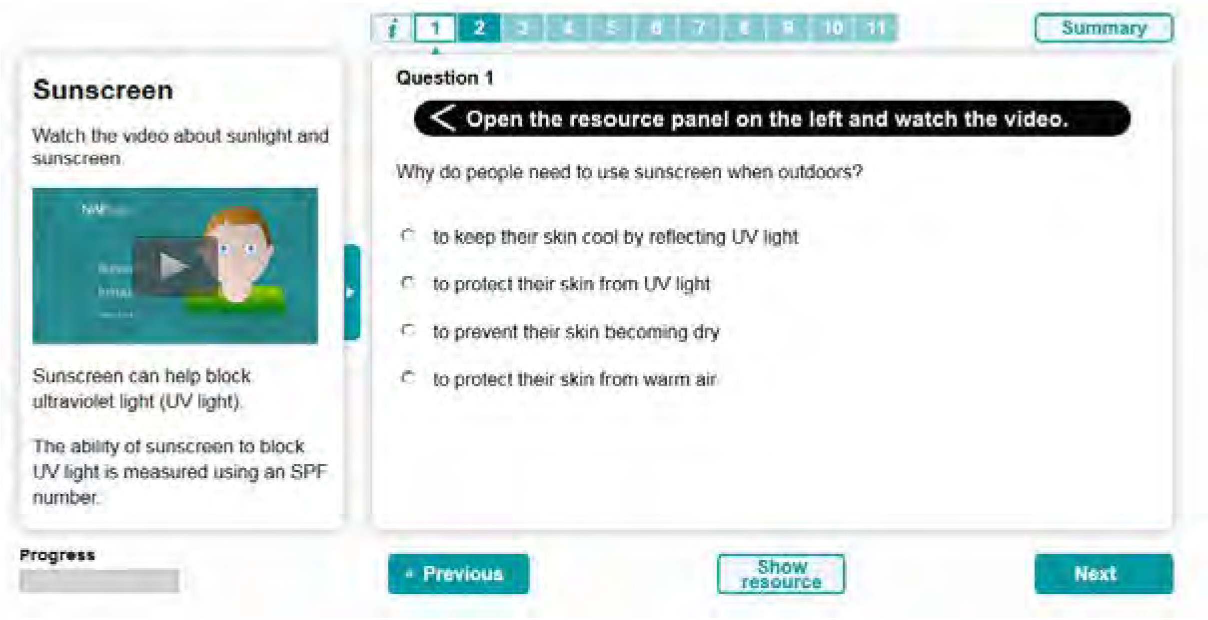
This ‘drag and drop’ style item assesses students’ ability to represent graphically the motion of a pendulum that a student had observed in a video clip.

The students had to choose from four double-headed arrows the one that best represented the motion of the pendulum’s bob.

Item 1 in the inquiry task Sunscreen (figure 5.4) also illustrates an item at Level 2 or below.

Figure 5.4

Sunscreen Q1



Type	Multiple choice		**Australian Curriculum**		Science as a human endeavour Use and influence of science
Strand	C		**Concept area**		Living things
Level	2 and below	**Percentage correct**	89	**Science literacy scale location**	214

This item assesses students' ability to give a reason for using sunscreen when outdoors.

The students were shown a video with information about the use of sunscreen lotions when outdoors. The students then had to inspect four statements and use their knowledge and personal experiences to select the one that gave a correct reason for using sunscreen when outdoors.

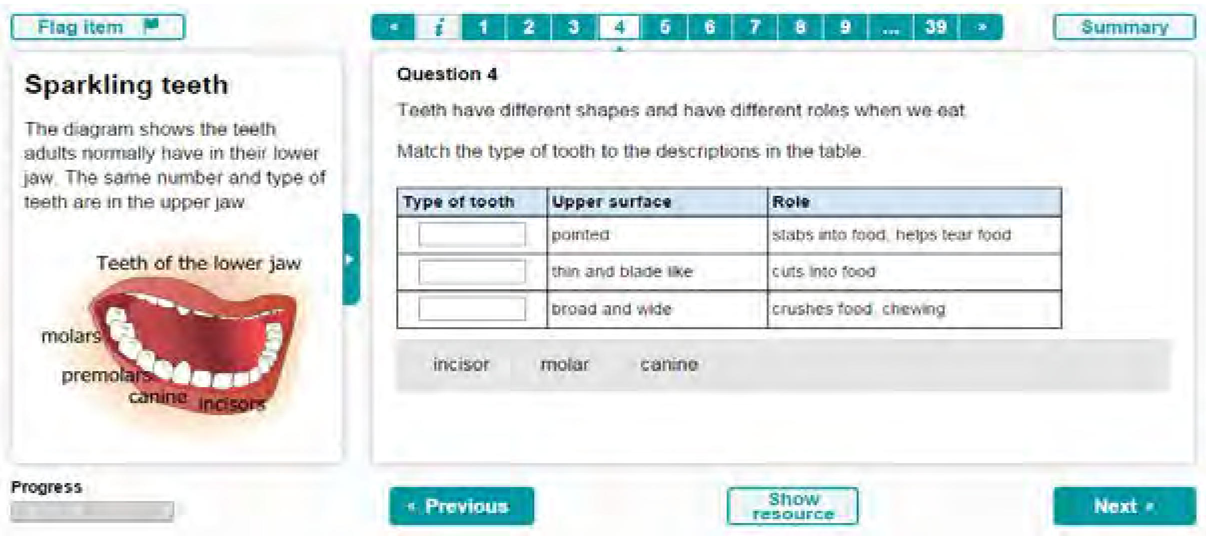
More generally, students who were able to complete items at this level of science literacy could also obtain data from tables and use them to make a comparison or to add labels to a diagram. They were able to use their knowledge and experience to describe events based on first-hand experiences.

Sample items illustrating performance within proficiency level 3.1

Item 3 in the item set Sparkling teeth (figure 5.5) illustrates performance at Level 3.1.

Figure 5.5

Sparkling teeth Q3



Type	Multiple choice		**Australian Curriculum**	Science understanding Biological sciences	
Strand	C		**Concept area**	Living things	
Level	3.1	**Percentage correct**	71	**Science literacy scale location**	352

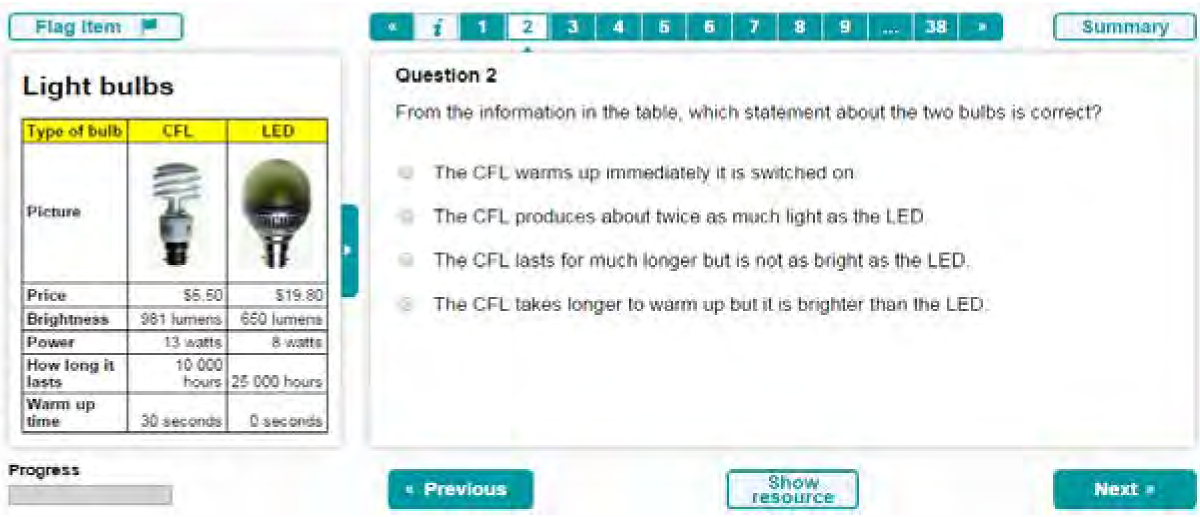
This item assesses the students' ability to match a description with a diagram.

Students were given a labelled diagram showing the different types of teeth adult humans have. They were also given a description of the appearance of the upper surface and the role of each tooth type. The students had to match three tooth types with these descriptions.

Item 2 in the item set Light bulbs (figure 5.6) is another illustration of an item at Level 3.1.

Figure 5.6

Light bulbs Q2



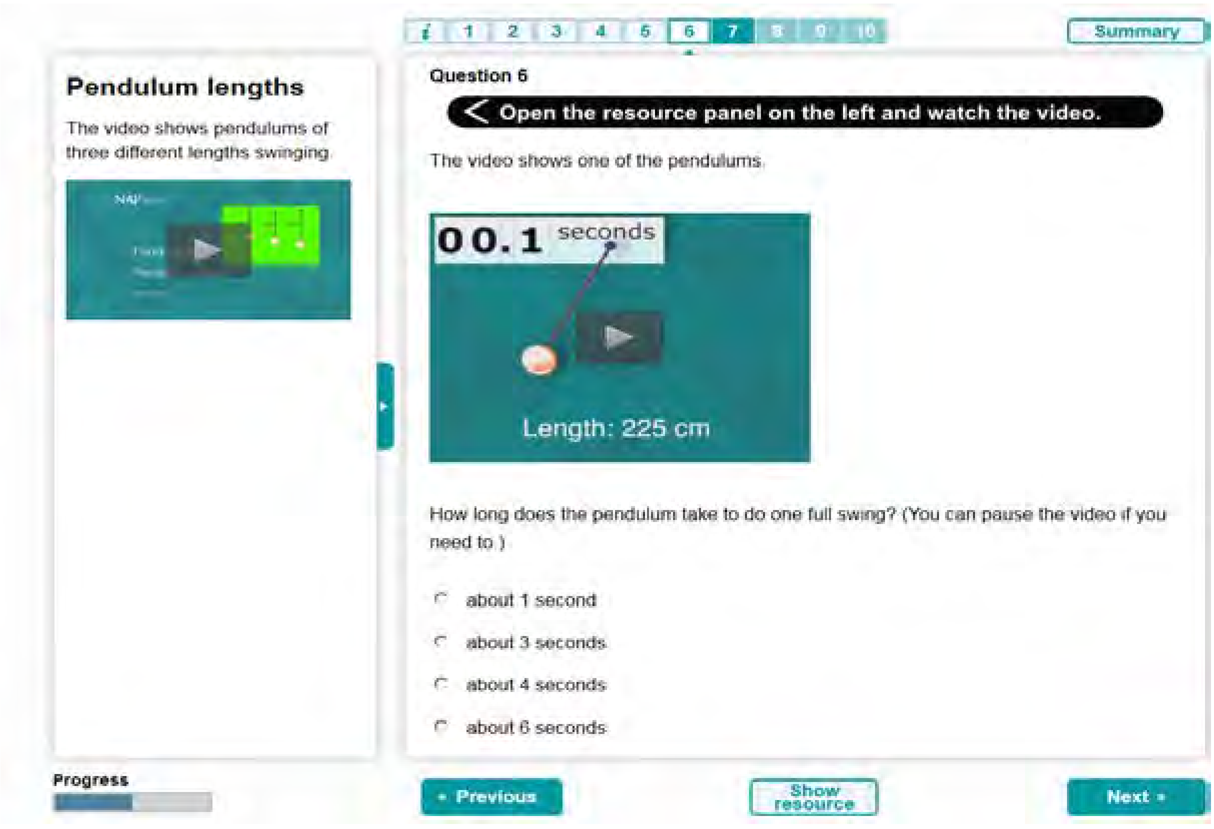
Type	Multiple choice		**Australian Curriculum**	Science inquiry skills Processing and analysing data and information	
Strand	B		**Concept area**	Energy and force	
Level	3.1	**Percentage correct**	77	**Science literacy scale location**	313

This item assesses students' ability to identify a statement supported by data in a table.

Students were provided with a table of information about two types of light bulbs. They were then asked to inspect four statements about the light bulbs and identify the one that was supported by the data in the table.Item 6 in the inquiry task Pendulums (figure 5.7) is another illustration of an item at Level 3.1.

Figure 5.7

Pendulums Q6



Type	Multiple choice		**Australian Curriculum**	Science inquiry skills Planning and conducting	
Strand	A		**Concept area**	Energy and force	
Level	3.1	**Percentage correct**	72	**Science literacy scale location**	357

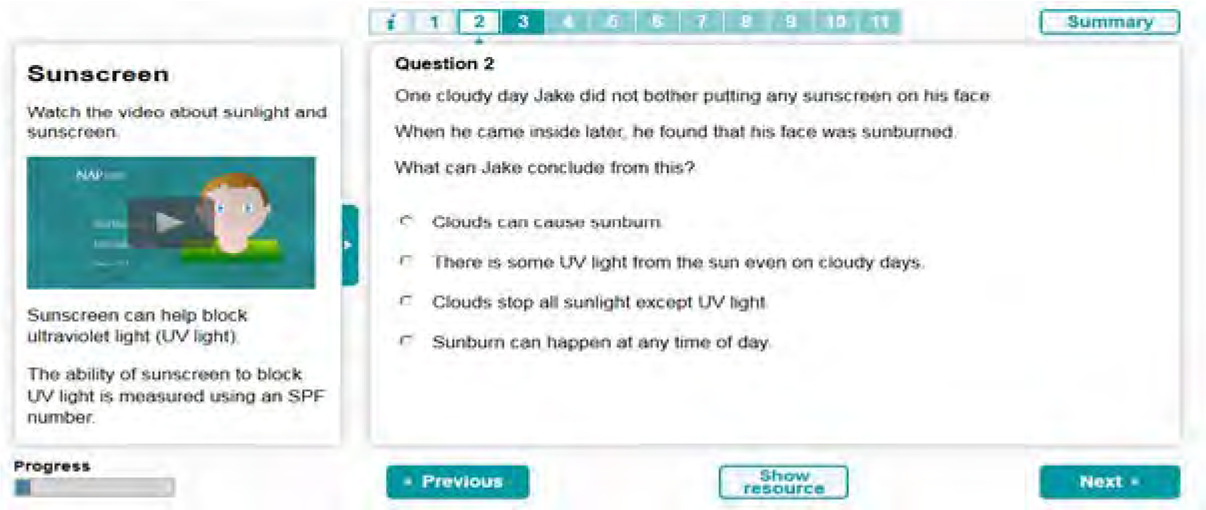
This item assesses students' ability to measure the time a pendulum takes for one full swing.

Students were shown a video that described a typical pendulum in terms of its construction and gave a definition of 'one full swing' of the pendulum. They were shown another video of a swinging pendulum together with an on-screen stopwatch that they had to use to measure the time of one full swing of the pendulum.

Item 2 in the inquiry task Sunscreen (figure 5.8) is another illustration of an item at Level 3.1.

Figure 5.8

Sunscreen Q2



Type	Multiple choice		**Australian Curriculum**	Science understanding Biological sciences	
Strand	C		**Concept area**	Energy and force	
Level	3.1	**Percentage correct**	70	**Science literacy scale location**	370

This item assesses students' ability to draw a conclusion based on information provided.

Students were given information about the use of sunscreen. They were then given a scenario where Jake, a character involved in a practical investigation, suffered sunburn on an overcast day. They were then asked to choose from four statements the one that contained a valid conclusion that Jake could draw from his being sunburnt.

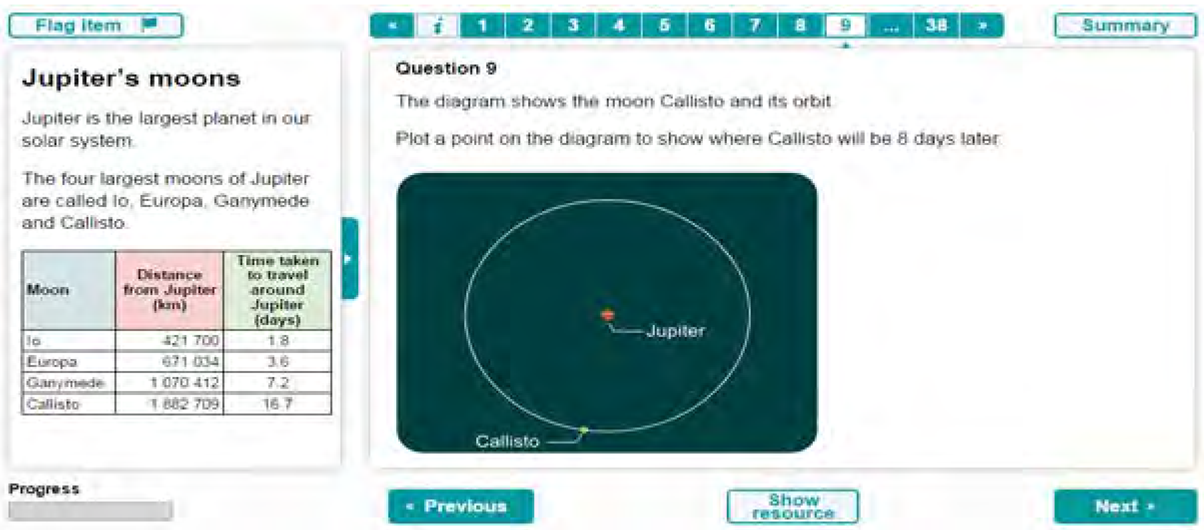
In a more general sense, students who could complete items at Level 3.1 were also able to recognise questions that could be answered using the results of an investigation. They were able to use their knowledge and experience to describe simple common phenomena. They were able to look up data in a table and use these to label a diagram and interpret more complex diagrams.

Sample items illustrating performance within proficiency level 3.2

Item 2 in the item set Jupiter's moons (figure 5.9) illustrates performance at Level 3.2.

Figure 5.9

Jupiter's moon Q2



Type	Select point		**Australian Curriculum**	Science inquiry skills Processing and analysing data and information	
Strand	A		**Concept area**	Earth and space	
Level	3.2	**Percentage correct**	44	**Science literacy scale location**	501

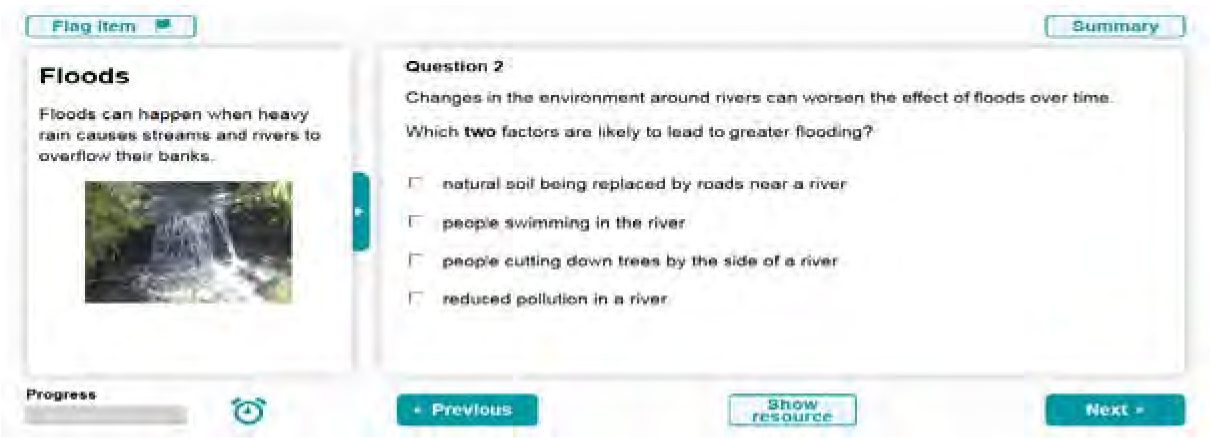
This item assesses students' ability to use data from a table to make a prediction and record it in a diagram.

Students were provided with a table containing data related to the time taken by four of Jupiter's moons to complete one orbit around the planet. The students were also shown a diagram representing the orbit of one of the moons. Using the data in the table, the students had to predict the location of the moon a given time later and mark this location on the diagram.

Item 2 in the item set Floods (figure 5.10) is another illustration of an item at Level 3.2.

Figure 5.10

Floods Q2



Type	Multiple choices		**Australian Curriculum**	Science as a human endeavour Use and influence of science	
Strand	C		**Concept area**	Earth and space	
Level	3.2	**Percentage correct**	43	**Science literacy scale location**	522

This item assesses students' ability to identify factors that affect the extent of flooding in an area.

Students were provided with a brief statement about flooding. They then had to inspect four statements describing changes in the environment surrounding rivers and use their knowledge and experience to select the two that describe factors that can worsen the effects of flooding.

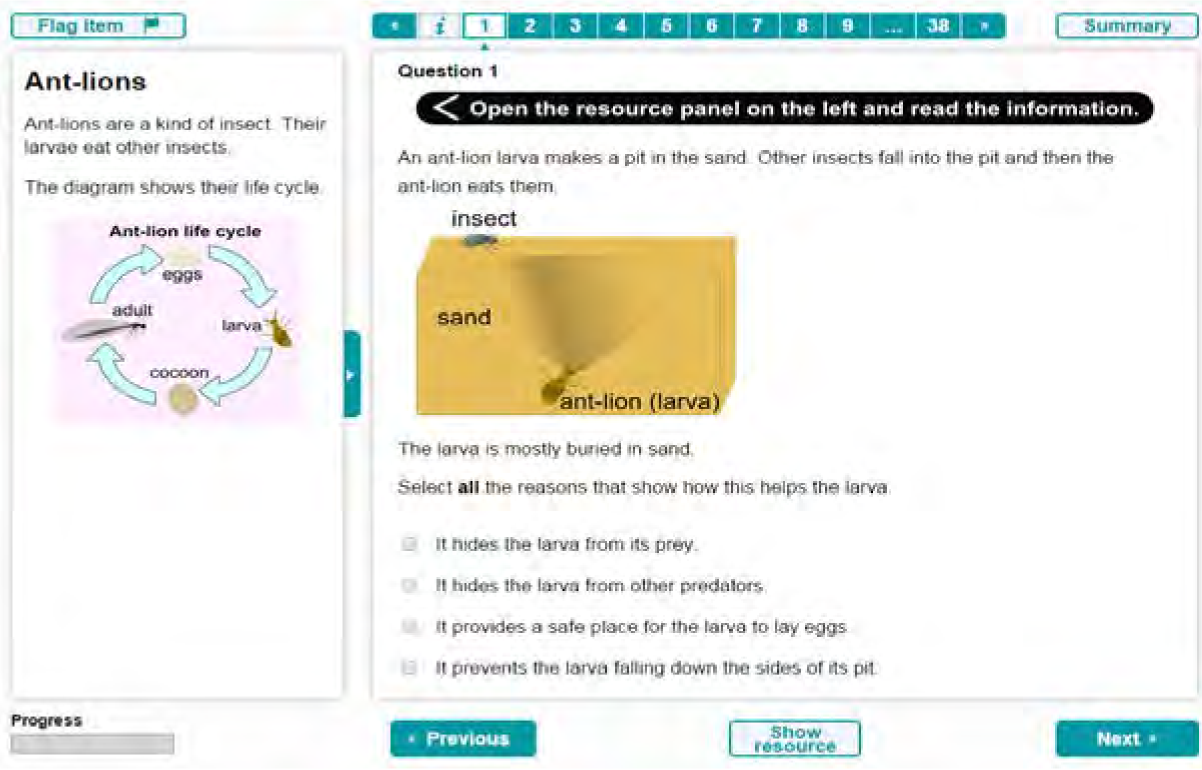
It could be expected that students who could complete items at Level 3.2 could also describe relationships between events or phenomena as well as make predictions based on information provided. They recognised the need to control variables so as to make an investigation valid. They could use their knowledge and experience to give an explanation for observations that they made and conclusions that they drew.

Sample items illustrating performance within proficiency level 3.3

Item 1 in the item set Ant-lions (figure 5.11) illustrates performance at Level 3.3.

Figure 5.11

Ant-lions Q1



Type	Multiple choices		**Australian Curriculum**	Science understanding Biological sciences	
Strand	C		**Concept area**	Living things	
Level	3.3	**Percentage correct**	23	**Science literacy scale location**	612

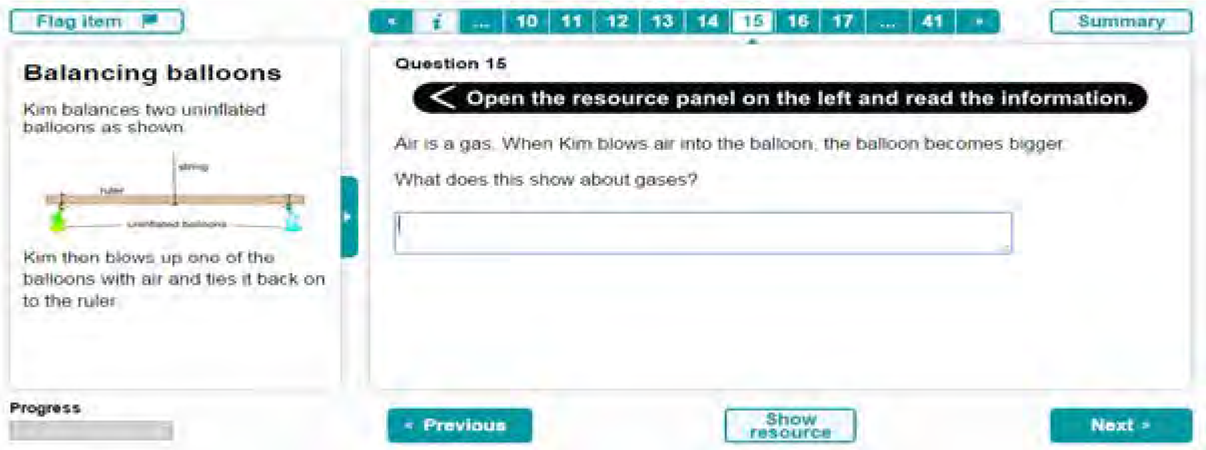
This item assesses students' ability to identify reasons for animal behaviour.

Students were presented with a diagram and text describing how an ant-lion larva catches its prey by digging a pit and then partly burying itself at the bottom of the pit, waiting for prey to fall in. The students had to inspect four statements and use their knowledge and insight to select the statements that explained how this action helps the ant-lion larva to catch its prey.

Item 1 in the item set Balancing balloons (figure 5.12) is another illustration of an item at Level 3.3.

Figure 5.12

Balancing balloons Q1



Type	Extended text		**Australian Curriculum**	Science understanding Chemical sciences	
Strand	C		**Concept area**	Matter	
Level	3.3	**Percentage correct**	21	**Science literacy scale location**	633

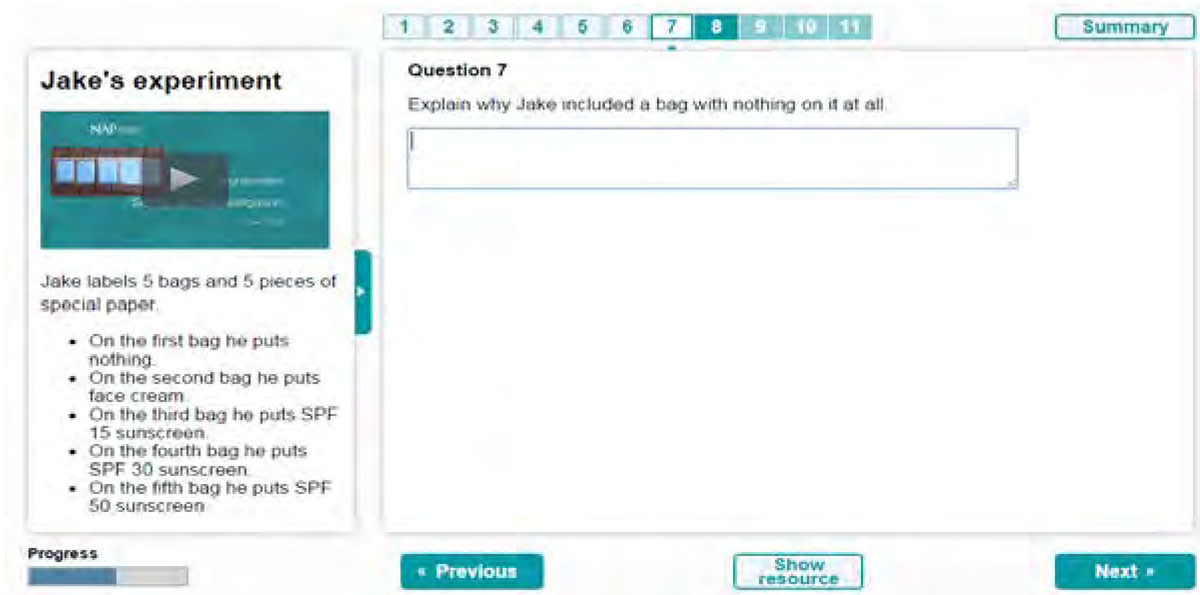
This assesses students' ability to identify a property of gases based on the description of observations in an investigation.

Students were presented with the experimental setup devised by a fictitious student, Kim, who suspended a metre ruler using a piece of string and balanced two uninflated balloons hanging from the ruler. The students were presented with the statements that air is a gas and that the balloon became bigger as Kim blew it up. The students then had to build on these statements and name the property of gases illustrated by this observation.

Item 7 in the inquiry task Sunscreen (figure 5.13) is another illustration of an item at Level 3.3.

Figure 5.13

Sunscreen Q7



Type	Extended text		**Australian Curriculum**	Science inquiry skills Planning and conducting	
Strand	A		**Concept area**	Living things	
Level	3.3	**Percentage correct**	23	**Science literacy scale location**	623

This item assesses students' ability to explain the need for a control in an investigation.

Students were presented with a video outlining the experimental method planned by a fictional character, Jake, to compare the effectiveness of different sunscreen lotions. In his method he used plastic bags coated with the different sunscreens but he also used one bag that did not have any coating at all. The students had to use their understanding of a control to explain why Jake used the bag that had no coating at all.

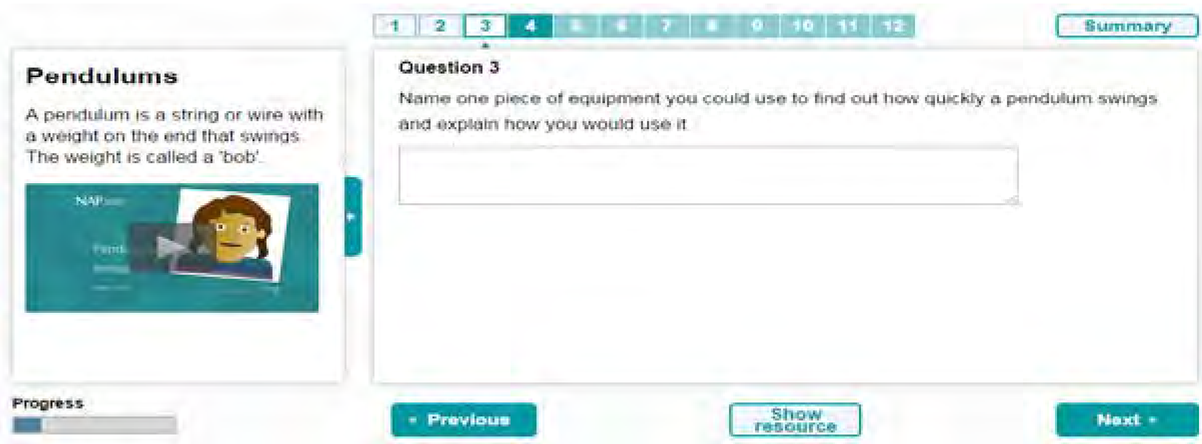
More generally, students who could complete items at this level of science literacy could also recognise the purpose of a control in an experimental design. They can explain the purpose of different steps in an investigation as well as the need to take a number of measurements to ensure the accuracy of their observations. They can describe and generate a pattern in data and use their knowledge and experience to explain more complex processes.

Sample items illustrating performance at proficiency level 4 and above

Item 3 in the inquiry task Pendulums (figure 5.14) illustrates performance at Level 4 or above.

Figure 5.14

Pendulums Q3



Type	Extended text		**Australian Curriculum**		Science inquiry skills Planning and conducting
Strand	A		**Concept area**		Energy and force
Level	4 and above	**Percentage correct**	12	**Science literacy scale location**	712

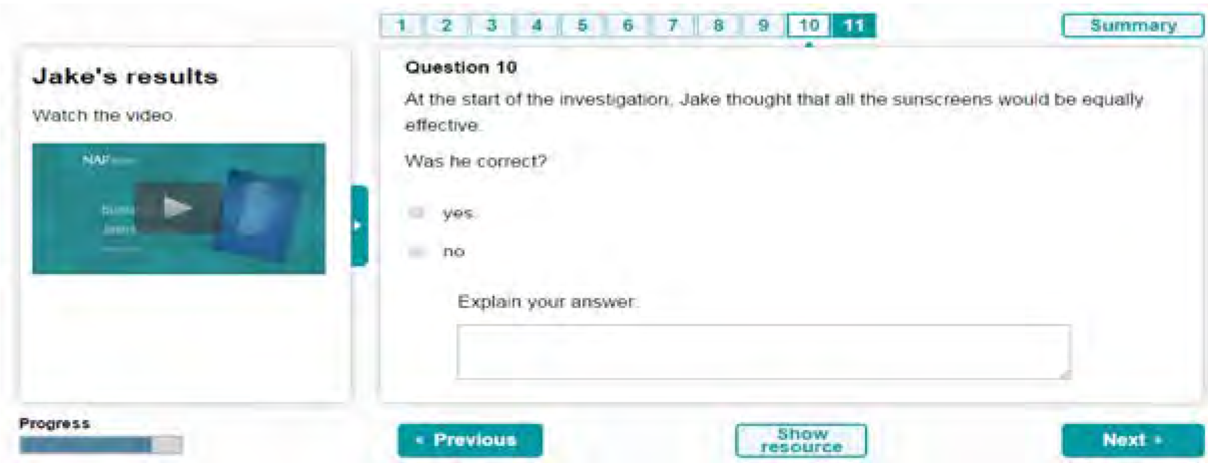
This item assesses students' ability to identify appropriate equipment to use in an investigation and describe how to use it.

Students were shown a video clip showing the parts of and motion of a swinging pendulum. The clip also defined the notion of 'one full swing' of a pendulum. After this, the students had to use their knowledge and understanding of science investigations to name one piece of equipment that they could use to find out the speed of a pendulum. They also had to explain how they would use this equipment.

Item 10 in the inquiry task Sunscreen (figure 5.15) is another illustration of an item at Level 4 or above.

Figure 5.15

Sunscreen Q10



Type	Composite: multiple choice & extended text		Australian Curriculum	Science inquiry skills Evaluating	
Strand	B		**Concept area**	Living things	
Level	4 and above	**Percentage correct**	9	**Science literacy scale location**	774

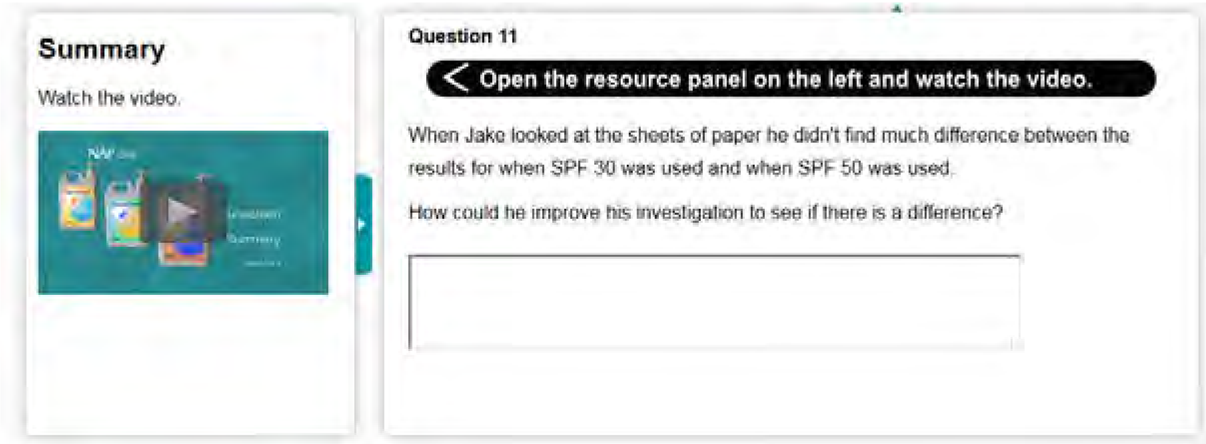
This item assesses students' ability to determine/evaluate whether the outcome of an investigation supports the prediction.

Students were shown a video clip showing a summary of Jake's investigation of sunscreen lotions and his results. They were also reminded of the prediction Jake made at the start of his investigation. They had to decide whether or not Jake's results supported his prediction and then use their reasoning skills to explain their decision, making reference to Jake's results.

Item 11 in the inquiry task Sunscreen (figure 5.16) is another illustration of an item at Level 4 or above.

Figure 5.16

Sunscreen Q11



Type	Extended text		**Australian Curriculum**		Science inquiry skills Evaluating
Strand	B		**Concept area**		Living things
Level	4 and above	**Percentage correct**	20 (1 mark) 3 (2 marks)	**Science literacy scale location**	774

This item assesses students' ability to modify an investigation to improve its level of accuracy and/or its ability to discriminate.

Following the video clip showing the summary of Jake's investigation and his results, the students were reminded that two of the results were so close as to be inconclusive. They had to suggest a way to modify the method of investigation so that the results would be more conclusive.

This is an example of a two-mark (polytomous) item. Responses that scored two marks provided two sensible improvements, while responses that scored only one mark provided only one. Examples of the type of responses that scored a mark include performing the test with a different light source, using a different and more sensitive detecting mechanism or comparing the SPF30 and SPF50 by varying the thickness of the layer applied.

It could be expected that students who could complete items at level 4 could also make suggestions for improving an investigation or recognise ways to get more accurate data. They were able to match a pattern in one set of data with one in another set and to integrate multiple ideas into one generalisation.

Item details

Table 5.2 shows a summary of the illustrative sample items found in this chapter. Further details about the performance of the items are provided in the 2015 NAP–SL Technical Report (available at www.nap.edu.au).

Table 5.2

Summary of sample items

Figure	Unit	Level	Scale score	Curriculum strand	Curriculum sub-strand	2012 strand	Area
5.2	Sparkling teeth	≤ 2	257	Science understanding	Biological sciences	B	Living things
5.3	Pendulums	≤ 2	193	Science inquiry skills	Communicating	C	Energy and force
5.4	Sunscreen	≤ 2	214	Science as a human endeavour	Use and influence of science	C	Living things
5.5	Sparkling teeth	3.1	352	Science understanding	Biological sciences	C	Living things
5.6	Light bulbs	3.1	313	Science inquiry skills	Processing and analysing data and information	B	Energy and force
5.7	Pendulums	3.1	357	Science inquiry skills	Planning and conducting	A	Energy and force
5.8	Sunscreen	3.1	370	Science understanding	Biological sciences	C	Energy and force
5.9	Jupiter's moons	3.2	501	Science inquiry skills	Processing and analysing data and information	A	Earth and space
5.10	Floods	3.2	522	Science as a human endeavour	Use and influence of science	C	Earth and space
5.11	Ant-lions	3.3	612	Science understanding	Biological sciences	C	Living things
5.12	Balancing balloons	3.3	633	Science understanding	Chemical sciences	C	Matter
5.13	Sunscreen	3.3	623	Science inquiry skills	Planning and conducting	A	Living things
5.14	Pendulums	≥ 4	712	Science inquiry skills	Planning and conducting	A	Energy and force
5.15	Sunscreen	≥ 4	774	Science inquiry skills	Evaluating	B	Living things
5.16	Sunscreen	≥ 4	733	Science inquiry skills	Evaluating	B	Living things

Performance levels of items by category

All items in the assessment are classified against a large number of categories. The performance of items by level can be looked at for each category. Some categories contain only a small number of items. Give consideration to this when reading the tables and interpreting the data.

Items excluded from the analysis for psychometric reasons are not included.

For more details of the categories used, see chapter 2.

Performance of items by item type

Items were classified by type in two different ways. Firstly items were classified by the nearest equivalent type used in paper-based testing.

Table 5.3

Levels of items by paper-based equivalent item type

Item type	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above	Total
Long constructed response		1	5	20	13	39
Short constructed response		2	2	1	1	6
Multiple choice	5	20	22	6		53
Other	3	2	3	1		9

As in previous cycles, the long constructed response item type was used to assess more complex and challenging skills. In particular, items assessing greater depth of understanding were often in this format and involved extended written answers.

Items were also classified by the type of online interaction used.

Table 5.4

Levels of items by type of online interaction

Interaction type	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above	Total
Composite		1	1	2	5	9
Extended text		2	6	19	9	36
Hotspot	1	1	1			3
Interactive gap match		1		1		2
Interactive graphic gap match	3	1	1			5
Multiple choice	4	19	17	5		45
Multiple choices			4	1		5
Select point			2			2

The composite item type involved two or more interactions combined together to make a single item.

The choice of interaction used for an item was based on the skills being tested. Consequently, not all interactions were used at every level of difficulty. It should not be concluded from this table that interactive graphic gap match items (a style of ‘drag and drop’ item) were inherently easier for students but rather that this style of interaction was often used to assess less complex skills.

Performance of items by the Australian Curriculum: Science categories

All items were classified against the Australian Curriculum: Science. Table 5.6 shows the spread of items in each of the major strands of the curriculum by proficiency level.

Table 5.5

Levels of items by curriculum strand

Curriculum strand	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above	Total
Science as a human endeavour	2	2	4	1	1	10
Science inquiry skills	2	13	15	16	9	55
Science understanding	4	10	13	11	4	42

Items from each curriculum strand appeared at every proficiency level in the assessment.

The spread of items across the curriculum can be further broken down into the number of items per sub-strand.

Table 5.6

Levels of items by curriculum sub-strand

Curriculum strand	Sub-strand	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above	Total
Science as a human endeavour	Nature and development of science		2	1			3
	Use and influence of science	2		3	1	1	7
Science inquiry skills	Communicating	1			2		3
	Evaluating			1	2	3	6
	Planning and conducting		2	2	10	3	17
	Processing and analysing data and information	1	9	11	2	2	25
	Questioning and predicting		2	1		1	4
Science understanding	Biological sciences	1	5	6	4		16
	Chemical sciences	1	1	1	3	1	7
	Earth and space sciences	1	1	2	3		7
	Physical sciences	1	3	4	1	3	12

Performance of items by 2012 NAP–SL categories

To ensure compatibility between the different cycles of the assessment, all items were also classified against the same categories used in 2012. This included the major scientific concept areas and Strands A, B and C (see chapter 2).

Table 5.7

Levels of items by concept area

Concept area	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above	Total
Earth and space	2	4	11	5		22
Energy and force	3	10	8	7	9	37
Living things	2	7	10	10	3	32
Matter	1	4	3	6	2	16

Table 5.8

Levels of items by strand

Strand	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above	Total
A		4	5	9	1	19
B	2	7	7	4	6	26
C	6	14	20	15	7	62

Performance of items by critical and creative thinking

To aid future investigation of the skills involved in assessing science literacy, items were categorised against the Australian Curriculum: critical and creative thinking (CCT) capability. In many of the items that were used in the final assessment, there was no single ideal category for the item within the critical and creative thinking learning continuum. In these cases the item was listed as ‘general’.

Table 5.9

Levels of items by critical and creative thinking strands

CCT strand	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above	Total
Inquiring – identifying, exploring and organising information and ideas			2	2	1	5
Generating ideas, possibilities and actions			2	2	2	6
Reflecting on thinking and processes		2	10	6	3	21
Analysing, synthesising and evaluating reasoning and procedures		3	1	7	4	15
General	8	20	17	11	4	60

Overall, items at Level 3.2 and above were easier to classify against the critical and creative thinking learning continuum than items below that level.

Chapter 6

Distribution of Students Within Proficiency Levels for 2015 With Comparisons to Previous Cycles

Chapter 6. Distribution of Students Within Proficiency Levels for 2015 With Comparisons to Previous Cycles

Introduction

Student achievement in science literacy is reported against three broad levels of proficiency. Level 3 is further segmented into three sub-levels represented by 3.1, 3.2 and 3.3. The proficient standard in science literacy is situated at the boundary between Levels 3.1 and 3.2.

Student performance by proficiency level

The 2015 distribution of students across these levels is shown in table 6.1. The NAP sample assessment – science literacy (NAP–SL) assessment was constructed with the expectation that most Year 6 students would demonstrate the understandings and skills described at Level 3. Table 6.1 shows that, at the national level, 9.7 per cent of students did not reach Level 3.

However, in the Northern Territory 32.8 per cent of students did not demonstrate science literacy corresponding to Level 3.

Table 6.1

Percentages of students at proficiency levels by state and territory in 2015

State/territory	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above
NSW	8.2 (±1.7)	34.5 (±3.3)	43.0 (±3.4)	13.3 (±2.9)	1.0 (±0.7)
Vic.	9.1 (±1.7)	37.3 (±3.3)	43.1 (±3.0)	10.1 (±2.4)	0.4 (±0.5)
Qld	11.1 (±3.2)	34.6 (±3.6)	42.8 (±4.1)	11.3 (±2.5)	0.3 (±0.3)
WA	9.3 (±2.1)	33.0 (±3.1)	44.7 (±3.0)	12.5 (±2.4)	0.4 (±0.4)
SA	11.0 (±2.4)	38.3 (±3.5)	40.8 (±3.3)	9.6 (±2.2)	0.3 (±0.3)
Tas.	9.6 (±2.6)	31.4 (±4.1)	42.5 (±4.1)	15.7 (±3.7)	0.8 (±1.0)
ACT	8.0 (±3.5)	31.5 (±3.9)	46.3 (±4.8)	13.8 (±2.5)	0.4 (±0.5)
NT	32.8 (±8.3)	35.4 (±6.5)	26.7 (±5.3)	5.0 (±2.2)	0.1 (±0.2)
Aust.	**9.7 (±1.0)**	**35.2 (±1.6)**	**42.9 (±1.5)**	**11.7 (±1.2)**	**0.6 (±0.3)**

Notes: Figures in parentheses refer to 95 per cent confidence intervals. Percentages may not add up to 100 per cent due to rounding.

Table 6.1 also shows that 55.1 per cent of students demonstrated skills and understandings that placed them at or above Level 3.2 (that is, at or above the proficient standard). Results further indicate that, at both the national and jurisdiction levels, the proportion of students performing at Level 4 and above accounts for only 0.1 to 1.0 per cent of the student population.

The proportion of students who demonstrated science literacy skills and understandings at or above the proficient standard is presented in table 6.2. The states and territories are listed according to the percentage of students operating at or above the proficient standard. Table 6.2 also contains the corresponding results and ranking for the 2006, 2009 and 2012 assessments.

In 2015, 55.1 per cent of students were found to be performing at or above the proficient standard at the national level. This compares to 54.3 per cent in 2006, 51.9 per cent in 2009 and 51.4 per cent in 2012. At the national level, the difference between 2015 and 2006 in the percentage of students achieving at or above the proficient standard is less than 1.0 per cent.

Table 6.2

Jurisdictions by percentage of students at or above the proficient standard in rank order for 2006, 2009, 2012 and 2015

Rank	2006		2009		2012		2015	
	State/ territory	Level 3.2 or above	State/ territory	Level 3.2 or above	State/ territory	Level 3.2 or above	State/ territory	Level 3.2 or above
1	ACT	62.0 (±5.6)	ACT	61.2 (±4.8)	ACT	65.3 (±5.3)	ACT	60.5 (±5.1)
2	Vic.	58.3 (±5.0)	Vic.	54.6 (±4.6)	WA	56.4 (±4.2)	Tas.	59.1 (±4.7)
3	NSW	57.4 (±4.3)	WA	53.3 (±4.5)	Tas.	51.3 (±5.4)	WA	57.7 (±3.3)
4	Tas.	57.4 (±5.5)	NSW	53.0 (±5.0)	Vic.	51.3 (±4.7)	NSW	57.2 (±3.6)
5	SA	51.6 (±4.7)	Tas.	49.8 (±6.0)	SA	51.1 (±3.9)	Qld	54.3 (±4.6)
6	Qld	49.2 (±3.8)	Qld	48.8 (±3.8)	NSW	50.9 (±4.3)	Vic.	53.6 (±3.8)
7	WA	46.6 (±4.7)	SA	46.5 (±5.0)	Qld	49.9 (±3.3)	SA	50.7 (±3.9)
8	NT	38.4 (±6.5)	NT	33.6 (±7.5)	NT	31.0 (±7.6)	NT	31.8 (±5.6)
	Aust.	**54.3 (±2.1)**	**Aust.**	**51.9 (±2.2)**	**Aust.**	**51.4 (±2.0)**	**Aust.**	**55.1 (±1.8)**

Note: Figures in parentheses refer to 95 per cent confidence intervals. Nevertheless, the change in ranking order for the states and territories should be regarded as indicative only as the differences are generally not significant.

Table 6.3 shows the 2015 percentage of students at or above the proficient standard against those in 2006, 2009 and 2012 and includes a column next to each comparison indicating change and whether or not the change is statistically significant.

Table 6.3

Percentage of students at or above the proficient standard in 2006, 2009, 2012 and 2015 with significance results

	2006	2009	2012	2015	2006-2015		2009-2015		2012-2015	
State/ territory	At or above the proficient standard	At or above the proficient standard	At or above the proficient standard	At or above the proficient standard	Change	Significant	Change	Significant	Change	Significant
NSW	57.4 (±4.3)	53.0 (±5.0)	50.9 (±4.3)	57.2 (±3.6)	-0.2	No	4.2	No	6.3	No
Vic.	58.3 (±5.0)	54.6 (±4.6)	51.3 (±4.7)	53.6 (±3.8)	-4.7	No	-1.0	No	2.3	No
Qld	49.2 (±3.8)	48.8 (±3.8)	49.9 (±3.3)	54.3 (±4.6)	5.1	No	5.5	No	4.4	No
WA	46.6 (±4.7)	53.3 (±4.5)	56.4 (±4.2)	57.7 (±3.3)	11.1	Yes	4.4	No	1.3	No
SA	51.6 (±4.7)	46.5 (±5.0)	51.1 (±3.9)	50.7 (±3.9)	-0.9	No	4.2	No	-0.4	No
Tas.	57.4 (±5.5)	49.8 (±6.0)	51.3 (±5.4)	59.1 (±4.7)	1.7	No	9.3	No	7.8	No
ACT	62.0 (±5.6)	61.2 (±4.8)	65.3 (±5.3)	60.5 (±5.1)	-1.5	No	-0.7	No	-4.8	No
NT	38.4 (±6.5)	33.6 (±7.5)	31.0 (±7.6)	31.8 (±5.6)	-6.6	No	-1.8	No	0.8	No
Aust.	**54.3 (±2.1)**	**51.9 (±2.2)**	**51.4 (±2.0)**	**55.1 (±1.8)**	**0.8**	**No**	**3.2**	**No**	**3.7**	**No**

Table 6.3 shows that the difference in the percentages of students in Western Australia achieving at or above the proficient standard in 2015 represents a significant increase from 2006 after recording steady gains over 2009 and 2012. However, the table also shows that differences between cycles for all other states and territories are not statistically significant.

Trends in proficiency levels

The trend in distribution of students at the national level across levels in all five NAP–SL test cycles is presented in Table 6.4.

Table 6.4

Percentage distribution of students across proficiency levels in 2003, 2006, 2009, 2012 and 2015

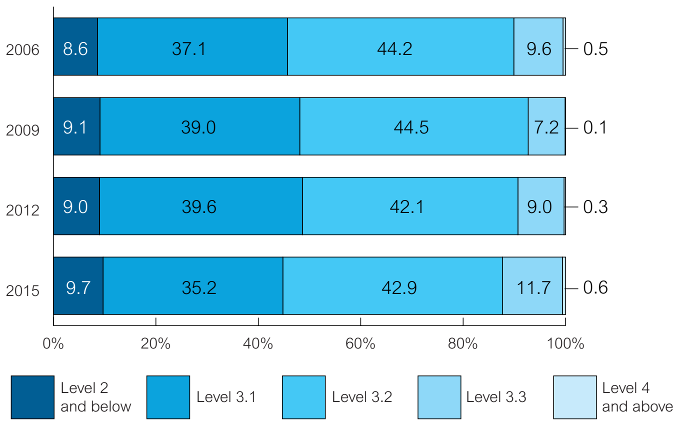
Aust.	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above
2003	4.1 (±0.7)	36.5 (±1.7)	52.2 (±1.7)	7.1 (±0.9)	0.1 (±0.1)
2006	8.6 (±1.1)	37.1 (±1.7)	44.2 (±1.8)	9.6 (±1.2)	0.5 (±0.4)
2009	9.1 (±1.2)	39.0 (±1.7)	44.5 (±1.8)	7.2 (±1.1)	0.1 (±0.1)
2012	9.0 (±1.0)	39.6 (±1.6)	42.1 (±1.7)	9.0 (±1.1)	0.3 (±0.2)
2015	9.7 (±1.0)	35.2 (±1.6)	42.9 (±1.5)	11.7 (±1.2)	0.6 (±0.3)

Notes: Figures in parentheses refer to 95 per cent confidence intervals. 2003 results are indicative only. Percentages may not add up to 100 per cent due to rounding.

Table 6.4 shows that the distribution of students across proficiency levels at the national level remained relatively stable across the assessments in 2006, 2009, 2012 and 2015. In 2015, 55.1 per cent of students were proficient at Level 3.2 and above. For 2006, 2009 and 2012, these percentages were 54.3, 51.9 and 51.4 respectively.

Figure 6.1

Percentage of students across proficiency levels for 2006, 2009, 2012 and 2015



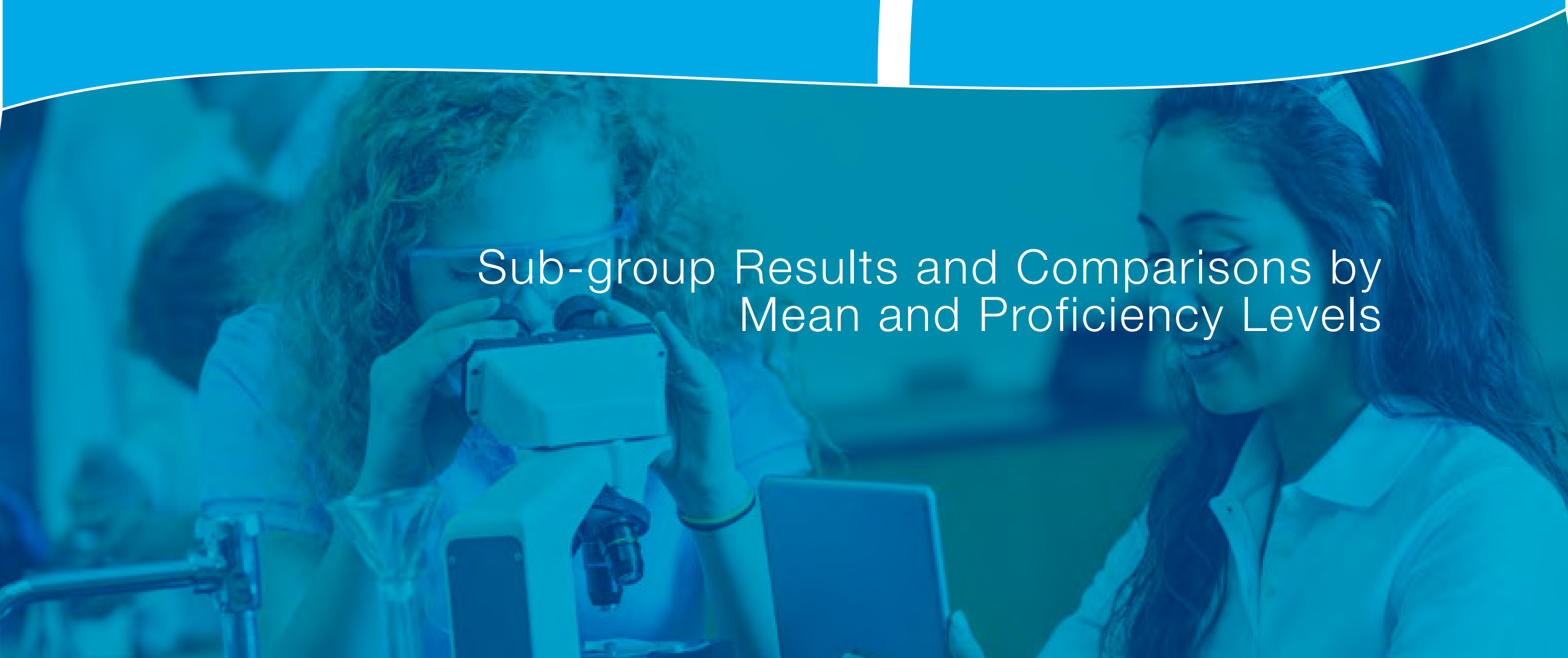
Note: Percentages may not add up to 100 per cent due to rounding.

This figure shows that the percentage of students who performed at or above Level 3.2 in 2015 has increased since 2012. Also, the distribution of students within Levels 3.1 and 3.3 are slightly different. In the 2012 assessment, 39.6 per cent of students performed at Level 3.1, whereas only 35.2 per cent of 2015 students performed at the same level.

The percentage of students in Level 3.3 has increased between 2012 and 2015 from 9.0 per cent in 2012 to 11.7 per cent in 2015. The percentage of students in the topmost level, Level 4, remains comparable to previous cycles.

Chapter 7

Sub-group Results and Comparisons by Mean and Proficiency Levels



Chapter 7. 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 across the states and territories are described for:

- male and female students
- Indigenous and non-Indigenous students
- students from diverse geographic locations
- students of diverse language backgrounds.

This chapter also contains the same information, where applicable, from the 2003, 2006, 2009 and 2012 assessments in order to allow trends in results for the NAP sample assessment – science literacy (NAP–SL) to be investigated.

Data were also collected on the parental education and parental occupation of students in the sample. However, as in previous cycles, the level of missing data for these categories was both high and variable across states and territories. Consequently, no comparisons have been made for these categories.

Mean scores by gender

Mean scores for male and female students across jurisdictions are presented in table 7.1. This table also provides information about the proportion of male and female students in the sample.

Table 7.1

Mean scores for male and female students by state and territory in 2015

State/territory	Percentage of male students in sample	Male students: mean score	Female students: mean score
NSW	52.9	405 (±10.4)	418 (±9.8)
Vic.	52.4	395 (±10.4)	403 (±11.1)
Qld	49.3	399 (±11.9)	398 (±12.5)
WA	47.8	393 (±9.9)	421 (±8.5)
SA	52.3	388 (±9.9)	397 (±12.2)
Tas.	49.5	403 (±14.0)	424 (±13.7)
ACT	51.6	410 (±11.8)	419 (±14.8)
NT	49.3	313 (±29.6)	327 (±30.9)
Aust.	**50.8**	**398 (±5.1)**	**408 (±5.1)**

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

It can be seen from table 7.1 that at the national level the mean score for female students is significantly higher than that for male students (scaled scores of 408 and 398 respectively). In every jurisdiction except Queensland, the mean score for female students was higher than male students. Western Australia had the largest difference between gender groups, with the mean score for female students 28 points higher than that for male students.

Proficiency levels by gender

Table 7.2 shows the distribution of results across the levels for male and female students and demonstrates that there were no significant differences in performance.

Table 7.2

Percentage distribution of male and female students across proficiency levels by state and territory in 2015

State/ territory	Gender of students	Level 2 and below	Level 3.1	Level 3.2	Level 3.3	Level 4 and above
NSW	Male	9.3 (±2.4)	35.6 (±3.9)	41.8 (±4.5)	12.5 (±3.8)	0.8 (±0.7)
	Female	7.0 (±2.4)	33.3 (±4.9)	44.3 (±4.3)	14.2 (±3.2)	1.2 (±1.1)
Vic.	Male	10.0 (±2.4)	36.9 (±4.1)	43.8 (±3.9)	8.9 (±2.9)	0.4 (±0.6)
	Female	8.1 (±2.2)	37.8 (±4.4)	42.3 (±4.6)	11.4 (±3.7)	0.5 (±0.8)
Qld	Male	11.5 (±3.5)	34.4 (±4.8)	41.7 (±4.8)	12.0 (±3.0)	0.3 (±0.5)
	Female	10.8 (±3.7)	34.7 (±4.4)	43.7 (±5.4)	10.5 (±2.9)	0.2 (±0.4)
WA	Male	12.2 (±3.4)	36.0 (±4.2)	41.4 (±3.8)	10.1 (±2.7)	0.4 (±0.6)
	Female	6.6 (±2.2)	30.3 (±3.7)	47.9 (±4.0)	14.7 (±3.3)	0.5 (±0.5)
SA	Male	12.2 (±2.9)	37.8 (±4.1)	40.5 (±4.4)	9.2 (±2.5)	0.3 (±0.4)
	Female	9.9 (±3.5)	37.6 (±4.7)	41.7 (±4.4)	10.5 (±3.1)	0.3 (±0.6)
Tas.	Male	11.5 (±4.2)	33.8 (±5.7)	39.8 (±5.4)	14.0 (±5.0)	1.0 (±1.3)
	Female	7.7 (±2.6)	29.0 (±4.6)	45.2 (±5.3)	17.5 (±4.4)	0.7 (±0.9)
ACT	Male	9.2 (±4.0)	31.9 (±6.2)	45.7 (±5.6)	13.2 (±3.2)	0.2 (±0.4)
	Female	6.8 (±3.7)	31.1 (±4.4)	47.0 (±6.4)	14.4 (±3.6)	0.6 (±0.9)
NT	Male	34.1 (±10.0)	36.8 (±8.5)	23.7 (±6.0)	5.4 (±2.2)	0.1 (±0.4)
	Female	31.5 (±11.5)	34.1 (±8.4)	29.7 (±8.5)	4.7 (±3.6)	0.1 (±0.3)
Aust.	**Male**	**10.7 (±1.2)**	**35.8 (±1.9)**	**42.0 (±2.0)**	**11.0 (±1.7)**	**0.5 (±0.3)**
	Female	**8.6 (±1.3)**	**34.5 (±2.2)**	**43.8 (±2.1)**	**12.5 (±1.5)**	**0.6 (±0.4)**

Notes: Figures in parentheses refer to 95 per cent confidence intervals. Percentages may not add up to 100 per cent due to rounding.

Trend analysis by gender

Table 7.3 shows the mean scores achieved by male and female students at the national level, as observed in the 2003, 2006, 2009, 2012 and 2015 NAP–SL.

Table 7.3

Mean scores for male and female students in 2003, 2006, 2009, 2012 and 2015

AUST	Percentage of male students in sample	Male students: mean score	Female students: mean score
2003	51.1	412 (±4.7)	405 (±4.0)
2006	50.8	402 (±6.4)	398 (±5.1)
2009	50.5	393 (±6.0)	391 (±5.2)
2012	50.6	394 (±5.6)	395 (±4.4)
2015	50.8	398 (±5.1)	408 (±5.1)

Notes: Figures in parentheses refer to 95 per cent confidence intervals. 2003 results are indicative only.

In 2015, the mean for female students was higher than the mean for male students by 10 points. In 2012, the mean for female students was only one point higher than the mean for male students and was not statistically significant.

Before 2012, male students achieved a slightly higher mean than female students but the difference was not statistically significant.

Table 7.4 shows the distribution of performance across all proficiency levels for male (M) and female (F) students.

Table 7.4

Percentage distribution across proficiency levels of male (M) and female (F) students in 2003, 2006, 2009, 2012 and 2015

AUST.	Level 2 and below		Level 3.1		Level 3.2		Level 3.3		Level 4 and above	
	M	F	M	F	M	F	M	F	M	F
2003	4.1 (±1.3)	4.2 (±0.8)	35.3 (±2.2)	37.6 (±2.1)	52.5 (±2.0)	52.0 (±2.1)	8.1 (±1.4)	6.1 (±1.0)	0.01 (±0.2)	0.1 (±0.1)
2006	9.0 (±1.4)	8.2 (±1.3)	36.1 (±2.2)	38.2 (±2.1)	43.6 (±2.2)	44.8 (±2.1)	10.6 (±1.7)	8.5 (±1.3)	0.7 (±0.6)	0.3 (±0.3)

2009	9.7 (±1.5)	8.3 (±1.3)	38.0 (±2.1)	40.2 (±2.4)	44.2 (±2.3)	45.1 (±2.2)	8.0 (±0.8)	6.5 (±0.6)	0.1 (±0.1)	0.1 (±0.1)
2012	9.9 (±1.4)	8.0 (±1.3)	38.4 (±2.2)	41.0 (±1.9)	41.9 (±2.3)	42.3 (±1.9)	9.4 (±1.5)	8.6 (±1.3)	0.4 (±0.2)	0.2 (±0.2)
2015	10.7 (±1.2)	8.6 (±1.3)	35.8 (±1.9)	34.5 (±2.2)	42.0 (±2.0)	43.8 (±2.1)	11.0 (±1.7)	12.5 (±1.5)	0.5 (±0.3)	0.6 (±0.4)

Notes: Figures in parentheses refer to 95 per cent confidence intervals. 2003 results are indicative only. Percentages may not add up to 100 per cent due to rounding.

The pattern of results remained relatively stable for the percentage of male and female students achieving various proficiency levels across the 2003 to 2015 assessments.

Table 7.5 shows the distribution of the percentage of male and female students at or above the proficient standard in 2006, 2009, 2012 and 2015.

Table 7.5

Percentage of male and female students at or above the proficient standard in 2006, 2009, 2012 and 2015

AUST.	Male students at or above the proficient standard	Female students at or above the proficient standard
2006	54.9 (±2.5)	53.7 (±2.3)
2009	52.3 (±2.6)	51.7 (±2.6)
2012	51.7 (±2.6)	51.1 (±2.2)
2015	53.5 (±2.1)	56.9 (±2.3)

Notes: Figures in parentheses refer to 95 per cent confidence intervals. The percentage of male and female students achieving at or above the proficient standard is not available for 2003 due to changes in scaling that occurred between 2003 and 2006.

Trend analysis by Indigenous status

Students from an Aboriginal and/or Torres Strait Islander background were classified as Indigenous students and other students were classified as non-Indigenous students.

The results for Indigenous students relative to non-Indigenous students are shown in table 7.6.

Table 7.6

Mean scores for Indigenous and non-Indigenous students in 2003, 2006, 2009, 2012 and 2015

AUST.	Indigenous students: mean score	Non-Indigenous students: mean score
2003	350 (±11.3)	412 (±3.7)
2006	311 (±29.4)	402 (±5.8)
2009	297 (±16.0)	397 (±5.0)
2012	303 (±15.1)	399 (±4.5)
2015	315 (±13.7)	408 (±4.2)

Notes: Figures in parentheses refer to 95 per cent confidence intervals. 2003 results are indicative only.

In 2015, the mean score for Indigenous students was 315, indicating that they did not perform as well as non-Indigenous students with a mean score of 408. This difference is statistically significant, as it also was in previous assessment cycles.

Table 7.7 contains a summary of the distribution across proficiency levels of Indigenous and non-Indigenous students for all the cycles of NAP–SL. Appendix 2 provides further details on the numbers of Indigenous students in the sample.

Table 7.7

Percentage distribution of Indigenous and non-Indigenous students across proficiency levels in 2003, 2006, 2009, 2012 and 2015

AUST	Level 2 and below		Level 3.1		Level 3.2		Level 3.3		Level 4 and above	
	Indigenous	Non-Indigenous	Indigenous	Non-Indigenous	Indigenous	Non-Indigenous	Indigenous	Non-Indigenous	Indigenous	Non-Indigenous
2003	15.9 (±1.3)	3.6 (±0.6)	51.6 (±6.3)	35.7 (±1.7)	30.9 (±6.7)	53.3 (±1.7)	1.7 (±2.0)	7.4 (±0.9)	0.0 (±0.0)	0.1 (±0.1)
2006	31.4 (±8.1)	8.1 (±1.1)	43.1 (±7.5)	37.3 (±1.8)	22.3 (±7.4)	44.3 (±1.9)	3.1 (±3.9)	9.8 (±1.4)	0.1 (±0.4)	0.6 (±0.5)
2009	38.6 (±6.8)	7.4 (±0.1)	41.8 (±5.8)	38.7 (±1.9)	18.6 (±6.0)	46.1 (±1.8)	1.0 (±1.5)	7.6 (±1.2)	0.0 (±0.0)	0.1 (±0.1)
2012	33.4 (±6.3)	7.9 (±1.0)	46.5 (±7.3)	39.3 (±1.6)	19.0 (±5.7)	43.1 (±1.8)	1.1 (±1.4)	9.4 (±1.2)	0.0 (±0.0)	0.3 (±0.2)
2015	32.1 (±5.9)	8.3 (±0.9)	44.6 (±6.0)	34.6 (±1.7)	21.0 (±4.7)	44.1 (±1.6)	2.0 (±1.7)	12.3 (±1.3)	0.4 (±1.1)	0.6 (±0.3)

Notes: Figures in parentheses refer to 95 per cent confidence intervals. 2003 results are indicative only. Percentages may not add up to 100 per cent due to rounding.

Table 7.7 shows that in 2015, approximately 32 per cent of Indigenous students were working at Level 2 and below, whereas approximately eight per cent of non-Indigenous students were working at that level. This pattern of results has been similar since 2006 with about one third of Indigenous students, and one twelfth of non-Indigenous students operating at Level 2 and below.

Table 7.8 shows the percentage of Indigenous and non-Indigenous students at or above the proficient standard in 2006, 2009, 2012 and 2015.

Table 7.8

Percentage of Indigenous and non-Indigenous students achieving at or above the proficient standard in 2006, 2009, 2012 and 2015

AUST.	Indigenous students at or above the proficient standard	Non-Indigenous students at or above the proficient standard
2006	25.5 (±10.0)	54.7 (±2.2)
2009	19.6 (±6.0)	53.9 (±2.3)
2012	20.1 (±5.8)	52.8 (±2.0)
2015	23.4 (±4.8)	57.0 (±1.8)

Notes: Figures in parentheses refer to 95 per cent confidence intervals. The equivalent data is not available for 2003 due to changes in scaling that occurred between 2003 and 2006.

Table 7.8 also shows that in 2015, 23.4 per cent of Indigenous students performed at or above the proficient standard. The typical pattern observed since 2006 shows that only one in four or one in five Indigenous students are operating at or above the proficient standard, whereas over half of the non-Indigenous students are operating at the same level.

Trend analysis by schools geographic location

Table 7.9 shows the mean scaled scores in 2009, 2012 and 2015 for students attending schools in different geographic locations. The pattern of results indicates students from metropolitan areas have significantly higher mean scores than students in provincial areas who in turn have higher mean scores than students in remote and very remote areas.

Table 7.9

Mean scores for students by school geographic location in 2009, 2012 and 2015

Geographic location category	2009		2012		2015	
	Percentage of students	Mean score	Percentage of students	Mean score	Percentage of students	Mean score
Metropolitan areas	72.3	395 (±6.2)	72.9	400 (±5.2)	72.1	410 (±5.1)
Provincial areas	24.7	389 (±7.9)	25.3	381 (±9.5)	25.3	389 (±8.3)
Remote and very remote areas	3.0	336 (±23.6)	1.9	349 (±31.0)	2.6	348 (±35.8)
AUST.	**100.0**	**392 (±5.1)**	**100.0**	**394 (±4.4)**	**100.0**	**403 (±4.3)**

Notes: Figures in parentheses refer to 95 per cent confidence intervals. The percentage of students in geographic location regions is weighted to reflect population percentages. They are not percentages of students in the sample.

Table 7.10 shows the percentage of students across proficiency levels by the geographic location of the sampled schools in 2009, 2012 and 2015. It can be seen that the percentage of students across proficiency levels in the 2012 and 2015 assessments was similar.

Table 7.10

Percentage distribution across proficiency levels by school geographic location in 2009, 2012 and 2015

		Geographic location category							
		Metropolitan areas		Provincial areas		Remote and very remote areas		AUST	
Level 2 and below	2009	8.4	(±1.5)	8.6	(±1.7)	28.2	(±8.8)	9.1	(±1.2)
	2012	7.8	(±1.2)	11.3	(±2.8)	23.2	(±9.5)	9.0	(±1.0)
	2015	8.5	(±1.0)	11.2	(±2.4)	27.9	(±11.1)	9.7	(±1.0)
Level 3.1	2009	38.1	(±2.0)	41.9	(±3.4)	37.9	(±8.4)	39.0	(±1.7)
	2012	39.0	(±2.0)	41.7	(±3.4)	35.1	(±7.4)	39.6	(±1.6)
	2015	34.3	(±1.8)	38.2	(±3.1)	31.7	(±8.3)	35.2	(±1.6)
Level 3.2	2009	45.5	(±2.1)	43.5	(±3.3)	29.6	(±7.2)	44.5	(±1.8)
	2012	42.9	(±1.9)	40.5	(±3.8)	35.5	(±9.2)	42.1	(±1.7)
	2015	43.5	(±1.9)	42.1	(±3.2)	32.8	(±10.2)	42.9	(±1.5)

Level 3.3	2009	7.8	(±1.4)	6.0	(±1.5)	4.1	(±3.7)	7.2	(±1.1)
	2012	10.0	(±1.5)	6.5	(±1.6)	6.2	(±3.5)	9.0	(±1.1)
	2015	13.1	(±1.6)	8.2	(±1.6)	7.4	(±4.3)	11.7	(±1.2)
Level 4 and above	2009	0.1	(±0.1)	0.1	(±0.1)	0.2	(±0.5)	0.1	(±0.1)
	2012	0.4	(±0.2)	0.1	(±0.1)	0.1	(±0.2)	0.3	(±0.2)
	2015	0.7	(±0.3)	0.3	(±0.4)	0.1	(±0.4)	0.6	(±0.3)

Notes: Figures in parentheses refer to 95 per cent confidence intervals. Percentages may not add up to 100 per cent due to rounding.

Table 7.11 shows the percentage of students achieving at or above the proficient standard in 2009, 2012 and 2015 by geographic location.

Table 7.11

Percentage of students achieving at or above the proficient standard in 2009, 2012 and 2015 by geographic location

AUST	At or above the proficient standard		
	Metropolitan areas	Provincial areas	Remote and very remote areas
2009	53.4 (±2.6)	49.5 (±4.1)	33.9 (±8.2)
2012	53.2 (±2.3)	47.0 (±4.4)	41.7 (±9.2)
2015	57.3 (±2.2)	50.6 (±3.6)	40.4 (±13.3)

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

Table 7.11 shows that 40.4 per cent of students who attended schools located in remote and very remote areas were at or above the proficient standard in science literacy in 2015. This is slightly less than the 41.7 per cent of 2012 students from the same geographic location at or above the proficient standard. However, this difference is not statistically significant.

Trend analysis by student language background

In 2006, an online system for collecting demographic information about students participating in NAP–SL was implemented. However, the system did not provide accurate and complete information; therefore, only 2003, 2009, 2012 and 2015 data are presented here.

In 2015, data were collected to understand the language background of students in Year 6. Results for students from a language background other than English (LBOTE) and students from an English speaking background (ESB) were compared. Table 7.12 provides a comparison of results between 2003, 2009, 2012 and 2015. It should be noted that a student's language background does not indicate the student's proficiency in English.

Table 7.12

Comparison of mean scores by student language background in 2003, 2009, 2012 and 2015

AUST.	Mean score	
	LBOTE	ESB
2003	374 (±10.7)	405 (±4.5)
2009	384 (±13.0)	396 (±4.7)
2012	389 (±13.7)	397 (±4.5)
2015	396 (±9.3)	405 (±4.6)

Notes: Figures in parentheses refer to 95 per cent confidence intervals. 2003 results are indicative only.

In 2015, LBOTE students had a mean score of 396, which was lower than students with an English speaking background (ESB) (mean=405), but not significantly lower. A similar trend was observed in previous assessments.

The distribution of students across the proficiency levels by language background is given in table 7.13.

Table 7.13

Percentage distribution across proficiency levels by student language background in 2003, 2009, 2012 and 2015

AUST	Level 2 and below		Level 3.1		Level 3.2		Level 3.3		Level 4 and above	
	LBOTE	ESB	LBOTE	ESB	LBOTE	ESB	LBOTE	ESB	LBOTE	ESB
2003	7.4 (±2.4)	3.5 (±0.6)	43.1 (±4.0)	35.3 (±1.8)	44.7 (±4.1)	53.5 (±1.7)	4.9 (±1.9)	7.5 (±0.9)	0.0 (±0.0)	0.1 (±0.1)
2009	12.4 (±3.2)	7.7 (±0.1)	38.7 (±3.9)	38.9 (±1.9)	40.0 (±3.5)	46.3 (±1.9)	8.8 (±13.2)	7.0 (±1.1)	0.2 (±0.3)	0.1 (±0.1)
2012	11.0 (±2.5)	8.2 (±1.1)	41.4 (±4.4)	39.2 (±1.6)	36.7 (±3.7)	43.6 (±1.8)	10.2 (±4.1)	8.9 (±1.1)	0.7 (±0.6)	0.2 (±0.1)
2015	12.1 (±2.3)	8.8 (±1.1)	35.6 (±3.1)	35.0 (±1.9)	39.9 (±3.0)	43.9 (±1.8)	11.7 (±2.3)	11.8 (±1.4)	0.7 (±0.7)	0.5 (±0.3)

Notes: Figures in parentheses refer to 95 per cent confidence intervals. 2003 results are indicative only. Percentages may not add up to 100 per cent due to rounding.

Table 7.13 shows that in 2015 12.1 per cent of LBOTE students were working at Level 2 and below, whereas only 8.8 per cent of ESB students were working at the same level. The other end of the scale shows that a similar percentage of LBOTE (12.4) and ESB (12.3) students achieved level 3.3 or above.

Table 7.14 shows the percentage of students achieving at or above the proficient standard by language background in 2009, 2012 and 2015. It shows that over time the percentage of students at or above the proficient standard is generally about four to five percentage points lower for LBOTE students when compared with ESB students.

Table 7.14

Percentage of students achieving at or above the proficient standard in 2009, 2012 and 2015 by student language background

AUST	At or above the proficient standard	
	LBOTE	ESB
2009	48.9 (±4.9)	53.4 (±2.3)
2012	47.6 (±5.4)	52.6 (±2.1)
2015	52.3 (±3.6)	56.2 (±2.0)

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

Chapter 8

Student Survey

Chapter 8. Student Survey

Introduction

In 2009, a student survey about their attitudes to, and interests in, science and science experiences in school was introduced into the NAP sample assessment – science literacy (NAP–SL). This addition to the testing program was continued in 2012. Students answered the survey after completing the objective test and the practical task.

For 2012 NAP–SL, a survey instrument consisting of 42 items was developed in consultation with stakeholders and then trialled. The instrument was guided by the items and results of the 2009 survey as well as by recommendations from state and territory representatives. Following analysis of the responses from the trial, 34 items were selected for inclusion in the final survey form.

The survey required Year 6 students to respond to items. which varied from simple yes/no questions to Likert scale and frequency rating scale formats.

Survey coverage of the Australian Curriculum: Science

For 2015, the survey was reviewed again to see how well it captured aspects of the science as a human endeavour strand of the Australian Curriculum: Science. Several existing items dealt with issues relevant to this strand but other aspects at Year 6 were not covered.

Two new sets of items were developed and trialled to improve the coverage of the survey with respect to the science as a human endeavour strand of the curriculum. One set of items looked at student understanding of the nature of science. A second set of items looked at what kind of people students perceived as being involved in science.

An extra nine items were added for the 2015 student survey, giving a total of 43 items. All items were delivered online.

Survey themes

Thematically, the student survey was divided into 12 groupings as shown in table 8.1.

Table 8.1

Survey item groupings

Grouping label	Items	Grouping
G01	1–4	Interest in science
G02	5–7	Self-concept of science ability
G03	8–11	Value of science
G04	12–16	The nature of science 1
G05	17–19	Science-related activities outside school
G06	20–21	Science-related activities at school
G07	22–25	Science teaching 1
G08	26–29	Science topics studied
G09	30	Time spent on science
G10	31–34	Science teaching 2
G11	35–39	The nature of science 2
G12	40–43	Who is involved in science

The items could be further organised into three broad categories:

- science as a human endeavour
- student engagement with science
- teaching and learning science.

Note that there are many ways of grouping these items thematically and there are naturally common themes running through all the items.

Distribution of students' responses to the student survey

The survey was completed by all Year 6 students in the NAP–SL sample assessment. For the 2015 assessment, the sample comprised 11,503 students. The results provide important insights into Australian Year 6 students' perceptions and attitudes toward science as they are based on a large sample of students.

As noted in the previous section, the student survey items can be divided into 12 groupings. A description of each grouping is provided below, organised by the three broad categories. The descriptions are followed by the results of the student survey presented as percentage frequencies. The percentages provided in each figure are derived from the responses received from all students for a particular response option. The response options are defined underneath each figure.

In all the following figures, percentages have been rounded and may not add up to 100.

Science as a human endeavour

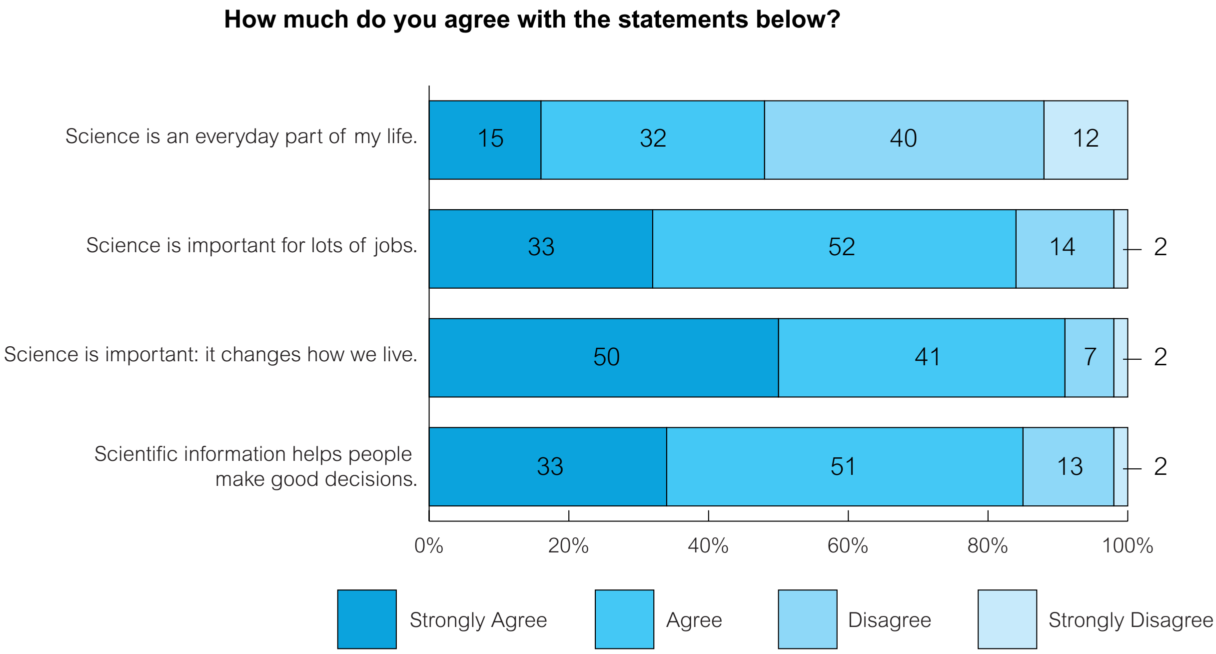
These items looked at aspects of student perception of science within the Science as a human endeavour strand.

G03 Value of science

This group included four statements designed to show student perceptions of the importance of science to society and themselves.

Figure 8.1

Value of science



Note: Percentages may not add up to 100 per cent due to rounding.

Very high percentages of student agreement were found for the last three statements that relate to the importance of science to society. A range of 41–52 per cent of students responded that they ‘agree’ and a range of 33–50 per cent of students responded that they ‘strongly agree’.

However, the pattern of responses was not the same for the statement ‘Science is an everyday part of my life’. This indicates that a large proportion of students shows a general appreciation for science but does not necessarily relate this to their own lives.

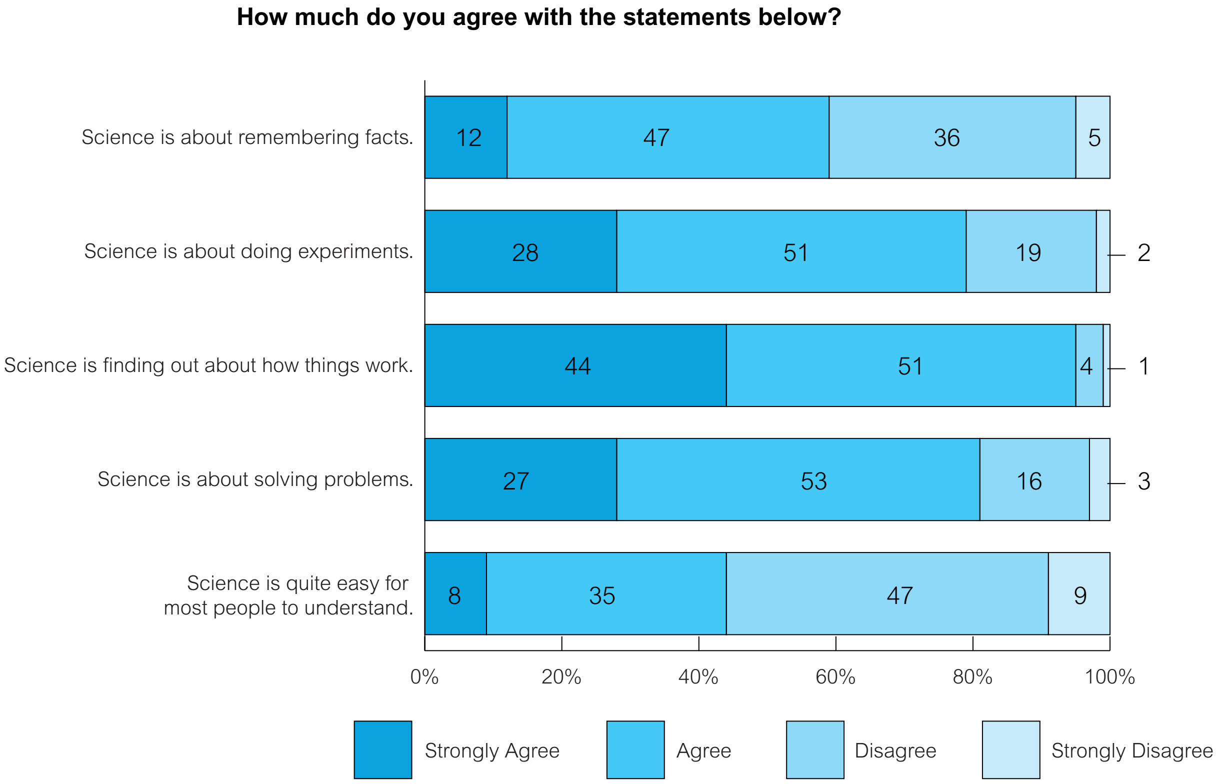
There were small but significant shifts towards agreement for each of the statements in 2015 in comparison with 2012.

G04 The nature of science 1

This group included four statements on what students consider to be science and one statement on whether science is easy for people to understand.

Figure 8.2

The nature of science 1



Note: Percentages may not add up to 100 per cent due to rounding.

Most students agreed or strongly agreed that science is about finding out how things work (95 per cent), doing experiments (79 per cent) and solving problems (80 per cent). Fewer students agreed that science is about remembering facts (59 per cent). These results were broadly similar to 2012.

More than half of students disagreed with the idea that science is easy for most people to understand.

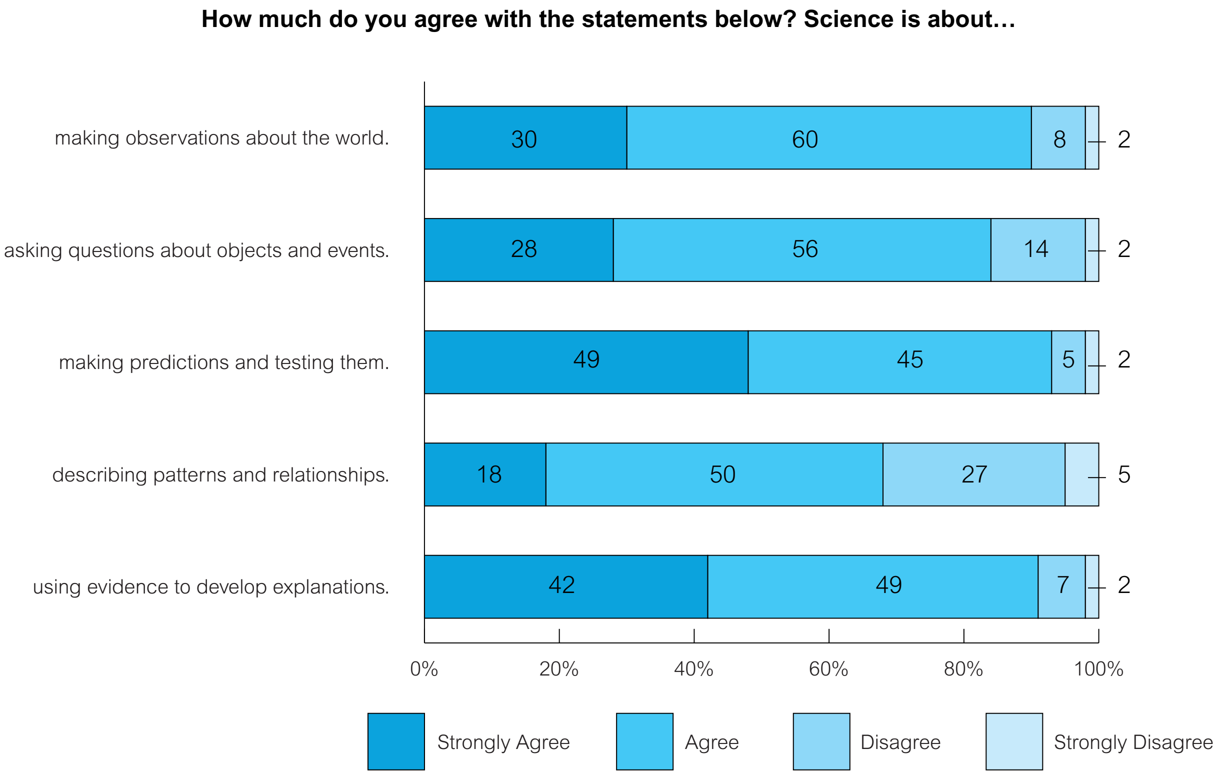
The statement 'Science is about solving problems' showed a small but significant increase in agreement in 2015 when compared with 2012 results. Other statements showed lower changes.

G11 The nature of science 2

This group was similar to G04 in that it included five statements on what students consider to be science. However, the statements were based on the language used in the Australian Curriculum and looked at science from a broader perspective. These statements were new for the 2015 survey.

Figure 8.3

The nature of science 2



Note: Percentages may not add up to 100 per cent due to rounding.

For four of the statements, the responses were broadly similar with between 84 and 94 per cent of students agreeing or strongly agreeing with the statements. In particular, nearly half of all students strongly agreed that science was about making predictions and testing them.

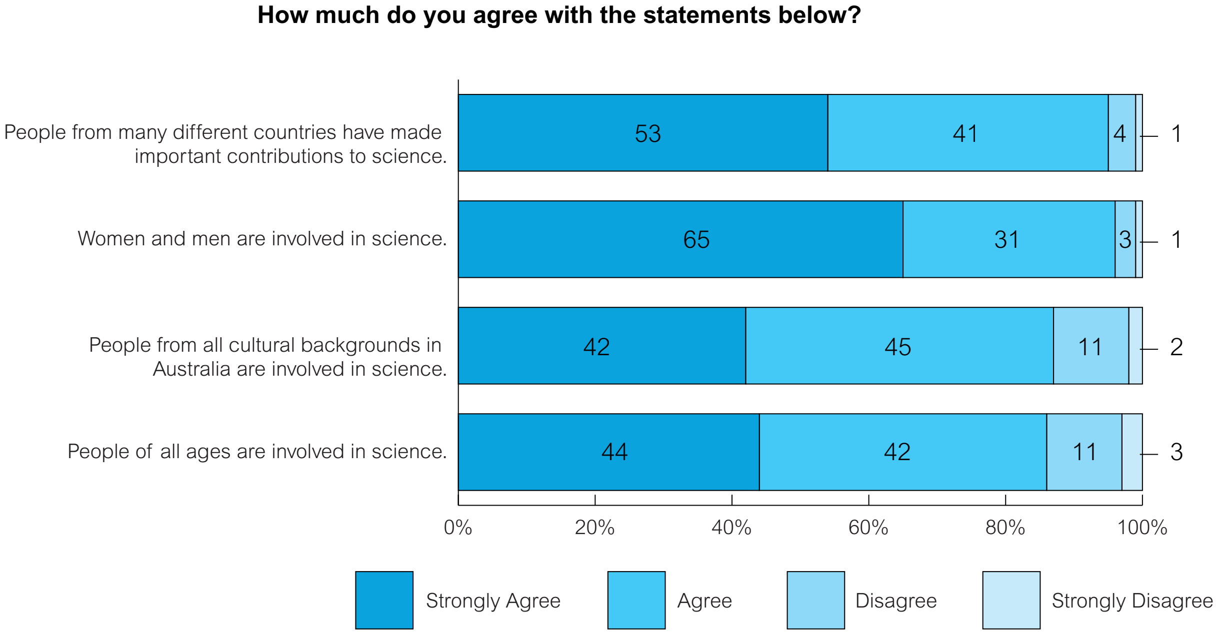
However, only 68 per cent of students agreed with the statement that science is about describing patterns and relationships. It is possible that some of these responses were shaped by the tasks that students had completed in the objective test and the inquiry task.

G12 Who is involved in science

The science as a human endeavour strand also covers the cultural and social aspect of science. This group of items was designed to gain a better understanding of how students perceive the people who may be involved in contributing to science.

Figure 8.4

Who is involved in science



Note: Percentages may not add up to 100 per cent due to rounding.

Students showed strong agreement with all of these statements. In particular, 65 per cent of students strongly agreed with the statement that women and men are involved in science.

Student engagement with science

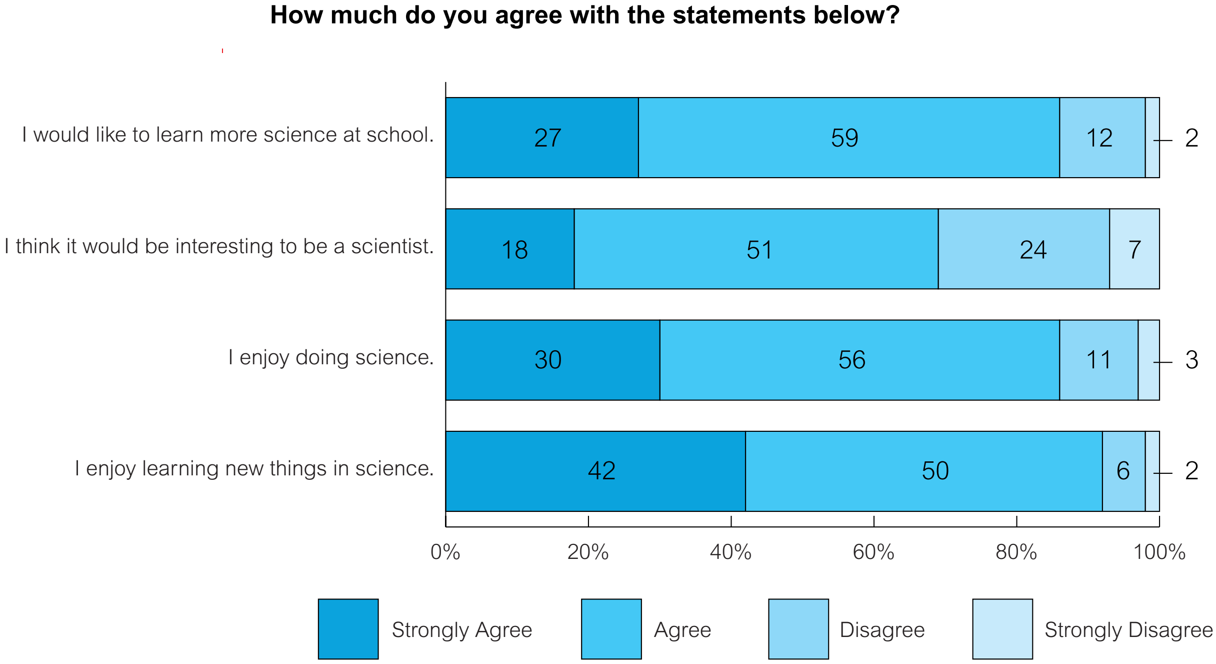
These items looked more specifically at how students engage with science.

G01 Interest in science

This group included four statements designed to show students' interest in learning and doing science as well as their interest in being a scientist.

Figure 8.5

Interest in science



Note: Percentages may not add up to 100 per cent due to rounding.

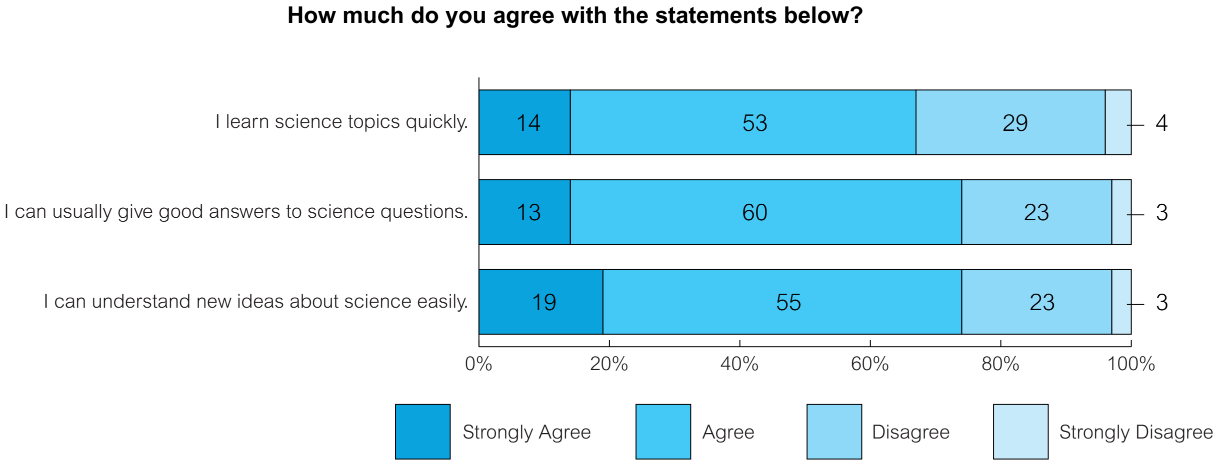
High percentages of student agreement were found for these statements. A range of 50–59 per cent of students responded 'agree' and a range of 18–42 per cent responded 'strongly agree'. While the great majority of students appear to be interested in learning about science and doing science, fewer students thought it would be interesting to be a scientist. The percentage of students who agreed or strongly agreed with the statement 'I think it would be interesting to be a scientist' was 69 per cent, which was a significant increase from 2012.

G02 Self-concept of science ability

This group included three statements to elicit students' own assessment of their abilities in science.

Figure 8.6

Self-concept of science ability



Note: Percentages may not add up to 100 per cent due to rounding.

High percentages of student agreement were found for these statements. A range of 53–60 per cent of students responded 'agree' and a range of 13–19 per cent responded 'strongly agree'. This indicates that a large proportion of students appear to be confident in learning science, reporting that they can understand and learn science ideas easily and quickly.

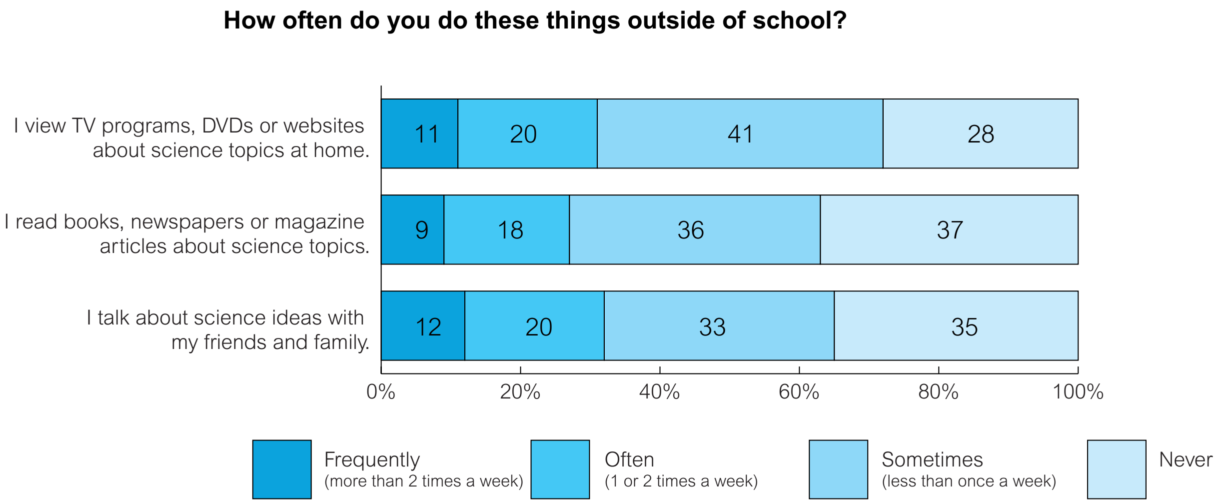
All of the statements in this group showed a significant shift towards greater agreement in 2015 when compared with 2012 results.

G05 Science-related activities outside school

The three statements in this group of the student survey sought to gather information about the frequency with which students engage in various science-related activities. These include watching television programs or DVDs, or viewing websites about science topics at home, reading books and newspaper or magazine articles about science topics, or talking about science ideas with friends and family.

Figure 8.7

Science-related activities outside school



Note: Percentages may not add up to 100 per cent due to rounding.

Responses to this group of items indicate some of the ways that students engage with science in their personal lives. About 72 per cent of students responded that they viewed science programs or science websites at home, with 31 per cent of students doing so 'often' or 'frequently'.

With more traditional print media, 63 per cent of students indicated that they read about science topics, with 27 per cent doing so 'often' or 'frequently'. A similar percentage of students indicated that they talked about science ideas with their friends and family at least sometimes.

Overall there was some shift towards students answering 'frequently' or 'often' to these statements in 2015 when compared with 2012 results.

Teaching and learning science

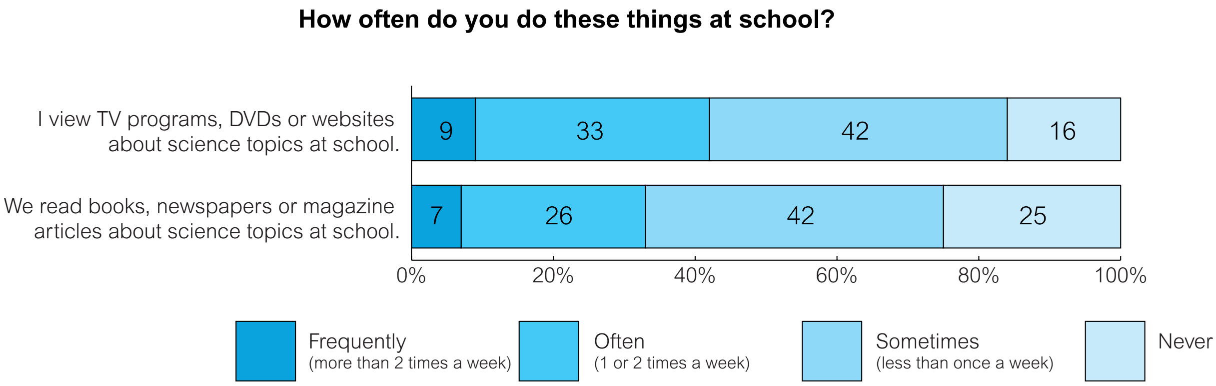
These items looked at student experience of science within school. Note that this is how students characterise their experience within schools. The data are not intended to necessarily show how science is actually being taught in the sampled schools.

G06 Science-related activities at school

These items were similar to two of the items used in G05 but with a change of context from outside of school to within school.

Figure 8.8

Science-related activities at school



Note: Percentages may not add up to 100 per cent due to rounding.

The proportions of students who responded ‘never’ to these two statements were noticeably smaller than to the similarly worded statements in G05. About 84 per cent of students responded that they viewed science programs or science websites at school, with 42 per cent of students doing so ‘often’ or ‘frequently’. This was a small but significant increase for ‘often’ and ‘frequently’ in 2015 when compared with 2012 results.

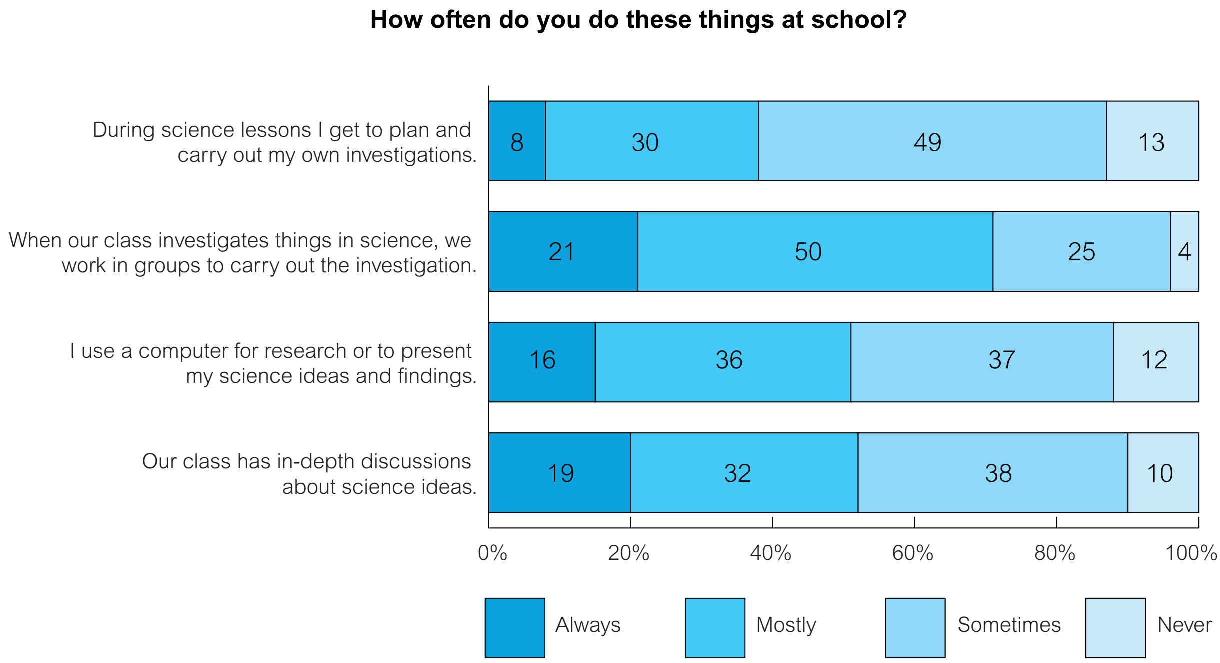
With more traditional print media, 75 per cent of students indicated that they read about science topics, with 33 per cent doing so ‘often’ or ‘frequently’.

G07 Science teaching

These items asked students to respond to statements about the kinds of things they do in school when learning science.

Figure 8.9

Science teaching



Note: Percentages may not add up to 100 per cent due to rounding.

The data show that 38 per cent of students reported that they 'always' or 'mostly' carried out their own self-directed investigations in science. However, 71 per cent reported that they 'always' or 'mostly' worked in groups to carry out investigations.

The over half (52 per cent) of students agreed that they use computers for research or to present science ideas 'mostly' or 'always'. A similar percentage said that they 'always' or 'mostly' have in-depth discussion in class about science ideas.

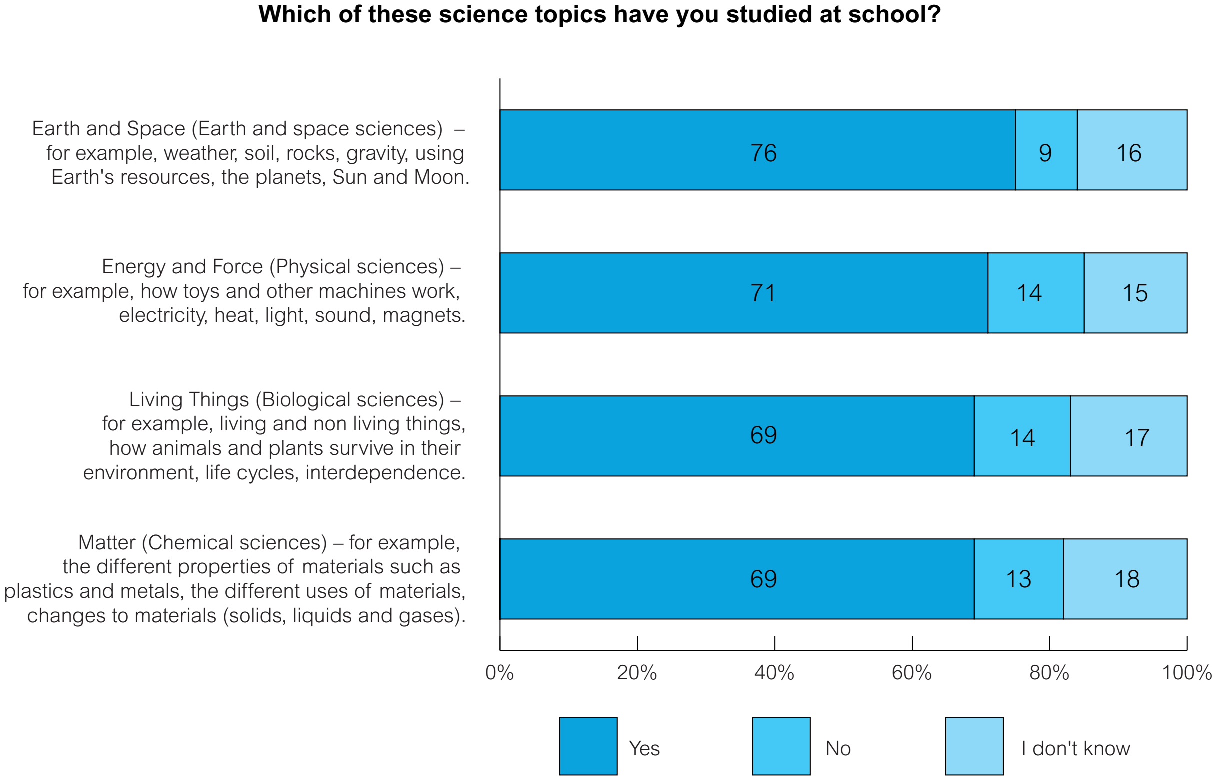
All statements in this group showed a shift towards 'always' and 'mostly' in 2015 when compared with 2012 results.

G08 Science topics studied

This set of items considered the four major scientific concept areas of science content in the curriculum. The statements had been amended from the 2012 versions by the inclusion in brackets of the names now used in the Australian Curriculum: Science. The original names and examples were developed in consultation with stakeholders in previous cycles of NAP–SL.

Figure 8.10

Science topics studied



Note: Percentages may not add up to 100 per cent due to rounding.

Figures for all four concept areas were very similar but when looked at together only 39 per cent of students responded yes to all four statements.

When compared with 2012 results there was no significant changes for the ‘Earth and space’ category. ‘Energy and force’ showed a small decrease in students responding ‘yes’ with a small increase in ‘I don’t know’. ‘Living things’ showed a decrease in ‘yes’ responses and an increase in ‘no’ responses. However, ‘Matter’ showed an increase in ‘yes’ responses and a decrease in ‘I don’t know’.

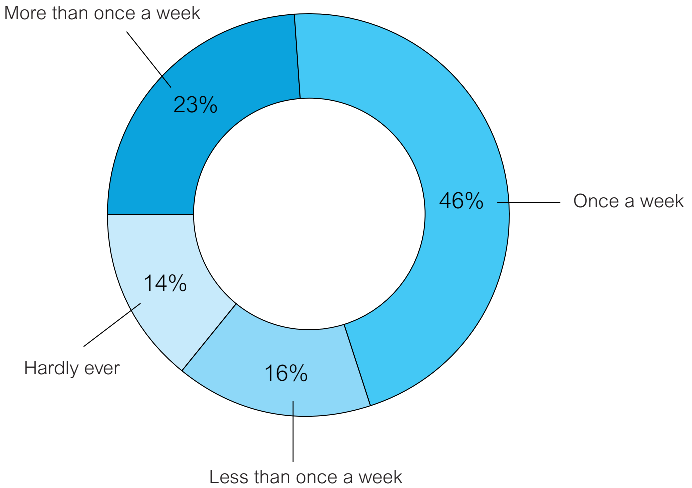
G09 Time spent on science

This item looked purely at the question of how often the students felt they had a science lesson at school. It is important to note that these figures are based purely on student opinion.

Figure 8.11

Time spent on science

How often do you have science lessons at school?



Note: Percentages may not add up to 100 per cent due to rounding.

The majority of students (69 per cent) reported having science lessons at least once a week.

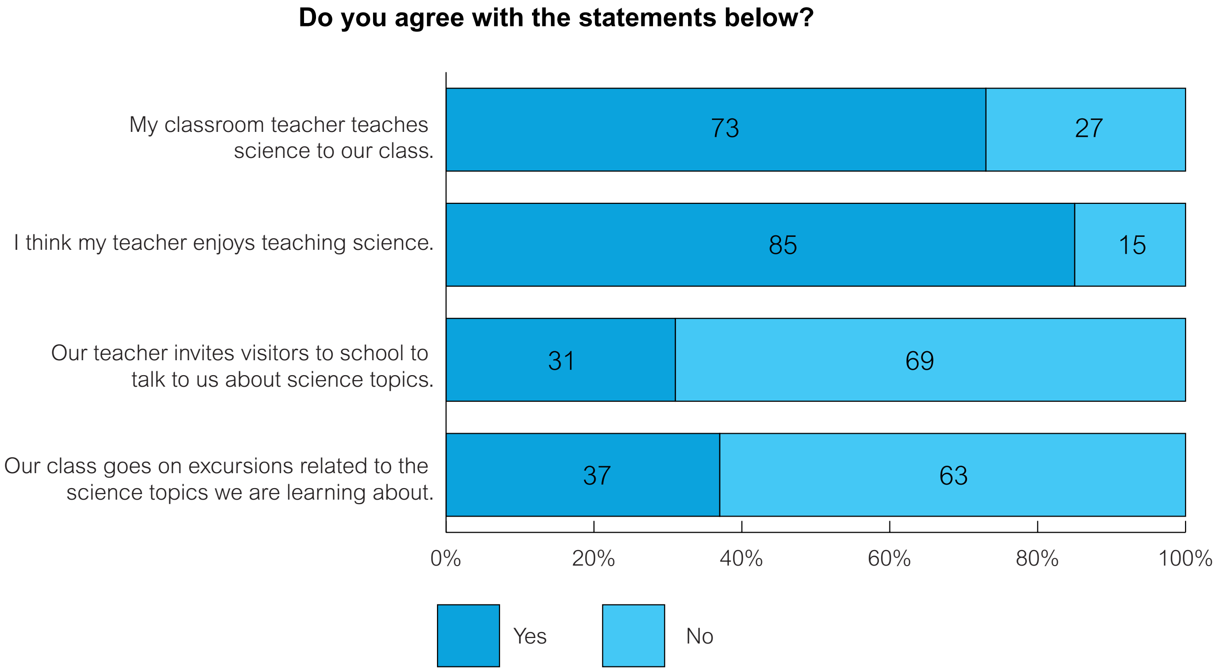
There was a decrease in students responding ‘hardly ever’ and an increase in students responding ‘once a week’ in 2015 when compared with 2012 results.

G10 Science teaching

The statements in this group of items were focused on the teacher and organised activities.

Figure 8.12

Science teaching



Note: Percentages may not add up to 100 per cent due to rounding.

Typically, students in Year 6 reported that they are taught science by their regular class teacher (73 per cent). Students also reported that they believe their teacher enjoys teaching science (85 per cent).

Most students reported that they do not have visitors come to speak to them about science topics (69 per cent) and that they do not go on science-related excursions (63 per cent).

There were few changes in results in this category in 2015 when compared with 2012 results. The exception was the statement ‘I think my teacher enjoys teaching science.’ This statement showed a small increase in students responding ‘yes’ when compared with 2012 results.

Longitudinal comparison of survey results 2012 to 2015

Appendix 4 contains a detailed comparison of the 2012 survey results compared with the 2015 survey results.

Examining the survey results as part of a scale

An examination of the 43 items of the survey suggested that there was a possible underlying construct of perception and engagement of science that was similar to the construct of science literacy that underpins the items of the NAP–SL assessment.

A new component of the 2015 NAP–SL cycle was to investigate whether there is some structure in the attitudes and beliefs covered by the survey that can be represented by a measurement scale. For example, is the item 'I enjoy doing science' something students are less likely to agree with than the item 'I learn science topics quickly'? A second question was how consistently students respond to all of the items in the survey.

A psychometric analysis of the data was undertaken to investigate these items and provide a measure for each student on a survey scale.

Students with high levels of agreement with the statements in the items are placed high on the scale. Students with low levels of agreement are placed low on the scale. However, being low on the scale is not intended to suggest low performance or lesser ability.

Items are placed on the scale differently. Items with less agreement from students would be high on the scale. Items with more agreement from students are placed low on the scale.

To more easily identify how the items reflect different positions on the scale; each item was then treated as a simple two-value response from students. For example, an item with the options 'strongly agree', 'agree', 'disagree' and 'strongly disagree' can be looked at by grouping 'strongly agree' and 'agree' together as a positive response to the statement, and 'disagree' and 'strongly disagree' together as a negative response to the statement.

It was found that an underlying construct has been measured with the 2015 NAP–SL survey items by fitting the data to the Rasch model. A similar analysis was performed on the survey data collected in 2012 confirming the existence of a measurable construct with the NAP–SL survey items. Further details are available in the 2015 NAP–SL Technical Report.

Relationship between student survey responses and science literacy in 2012

In 2012, the results from the NAP–SL survey were examined to see if there were systematic patterns between students' responses to the student survey. To do this, responses were subjected to a statistical technique called 'factor analysis'. This process found that a number of factors could be identified in the student survey results.

In order to investigate further the relationships between information collected in the student survey and student achievement in NAP–SL, a regression analysis was carried out. The regression analysis used items from the survey as independent variables, and student achievement as dependent variables.

The 2012 regression analysis showed that only 23 per cent of the variability in students' scores in science literacy could be predicted based on their responses to the items in the survey. Furthermore, only a few items demonstrated a correlation strong enough to be meaningful.

Of those items that showed some meaningful correlation, most were from the set of items relating to students' self-concepts in science. That is, there was some correlation between how capable students thought they were at learning science and their level of achievement in NAP–SL.

However, the approach taken in 2012 could only show some limited relationship between the survey results and student achievement in NAP–SL.

Relationship between student survey responses and science literacy in 2015

Because of the limitations found with the methods used in 2012, a different approach was taken with the results in 2015. This approach is called 'multilevel modelling'.

Multilevel modelling provides reliable tests of statistical significance between the survey data and the achievement data by modelling the structure of the data due to students attending different schools. It is more likely that students within a given school share similar characteristics than students enrolled in different schools. Ignoring clustering of students can be the source of technical problems that may lead to underestimating the relationship between the survey data and the student achievement data.

The aim of the multilevel modelling of 2015 NAP–SL data was to determine whether observed differences were statistically significant for the purpose of making inferences to the population from which the representative sample was drawn.

Variables considered

To examine the relationships between the data, the following variables were defined:

- student achievement in science as measured by the NAP–SL assessment
- student survey results broken into the groupings of items in the survey (G09 was not included as it represents only a single item)
- gender
- Indigenous status
- language background
- jurisdiction.

The multilevel modelling technique looked at the degree to which each of these variables relates to one another, while taking into account the multiple relationships between the variables.

Results comparing groups of survey items with achievement data

The first set of results compares each of the groupings of survey items (except for G09) against student achievement in NAP–SL.

Table 8.2

Correlations of survey groupings and achievement in NAP–SL by state

Group of items	Significantly positive correlation	No significant correlation	Significantly negative correlation
G01		AUST. and all states/ territories	
G02	AUST. and all states/ territories		
G03	AUST, NT, Qld, Tas., Vic. and WA	ACT, NSW and SA	
G04			AUST. and all states/ territories
G05		AUST. and all states/ territories	
G06		ACT, NT, Qld and Tas.	AUST., NSW, SA, Vic. and WA
G07		AUST. and all states/ territories	
G08	AUST., Qld, SA and WA	ACT, NSW, NT, Tas. and Vic.	
G10		ACT, SA, Tas., Vic. and WA	AUST., NSW, NT and Qld
G11	AUST and all states/ territories		
G12	AUST and all states/ territories		

Table 8.2 shows that the relationship between students' performance and the groupings of the survey questions. It should be noted that the relationships outlined in this analysis do not imply any causal connection with science achievement and that further investigation is required to better understand all the factors driving the successes in science literacy.

G01 Interest in science: Overall, interest in science does not correlate statistically significantly with achievement in science and this result is confirmed in the data of each state or territory. This does not mean, of course, that interest in science is irrelevant to student learning. However, it does show that there is a broad level of interest in science among students of all achievement levels.

G02 Self-concept of science ability: Overall, self-concept of science ability is positively correlated in a statistically significant way with achievement in science. This result is confirmed in each state/territory. This finding is similar to that found in past cycles. Students with a strong self-concept of their science ability tend to do better overall on the NAP–SL scale.

G03 Value of science: Overall, value of science is positively correlated with achievement in science. It is positively correlated in each state but the correlation is not statistically significant in ACT, NSW and SA.

G04 The nature of science 1: The nature of science (how) is negatively correlated with achievement in science overall and in each State.

G05 Science-related activities outside school: Overall, the variable science-related activities outside school is weakly correlated with science achievement. This result is confirmed in each state/territory. Such a result is similar to G01. Students of all abilities show some interest in science and science activities. This may contribute to overall levels of achievement but does not appear to contribute to individual differences.

G06 Science-related activities at school: Overall, the variable science-related activities at school is negatively correlated with science achievement. This result is confirmed in each state/territory but it is not statistically significant in ACT, Northern Territory, Queensland and Tasmania.

G07 Science teaching 1: Overall, the variable science teaching is weakly correlated with science achievement. This result is confirmed in each state and territory.

G08 Science topics studied: Overall, the variable science topics studied is positively correlated with science achievement. The same result has been observed in each state and territory but it is statistically significant only in Queensland, South Australia and Western Australia.

G10 Science teaching 2: Overall, science teaching 2 is negatively correlated with science achievement. There is a negative correlation in each State but it is statistically significant only in NSW, Northern Territory and Queensland.

G11 The nature of science 2: Overall, this variable is positively correlated with science achievement. This result is statistically significant in each state/territory. This group of items was a new group based on aspects of the science as a human endeavour strand. Notably G04 also included statements about the nature of science. However, G04 showed negative correlation with achievement, whereas the newer statements based on the curriculum in G11 showed positive correlation with achievement.

G12 Who is involved in science: Overall, this variable is positively correlated with science achievement. This result is statistically significant in each state/territory. Most students agreed with these statements about the kinds of people involved in science. However, it is notable that it showed some positive correlation with science achievement.

Conclusion

The NAP–SL survey provides rich insights into how students perceive and engage with science. Year 6 students show a high degree of agreement with the view of science outlined in the science as a human endeavour strand of the Australian Curriculum: Science. These views include a positive view of how women and men of all ages and from diverse backgrounds have a role within science.

Chapter 9

NAP–SL for teachers

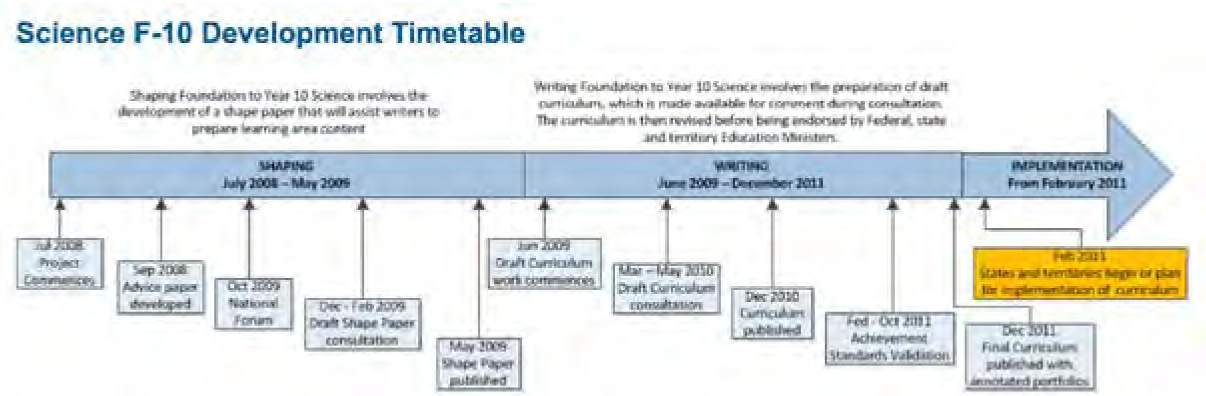
Chapter 9. NAP–SL for teachers

The purpose of this chapter is to provide information that is specifically intended for classroom teachers including an overview of:

- the development and implementation of the Australian Curriculum: Science
- the links between the Australian Curriculum and NAP–SL
- helpful suggestions to improve student science learning in the classroom.

Development and implementation of the Australian Curriculum:

The development of the Australian Curriculum: Science occurred in the period 2008–2011 and is summarised in this diagram:



Reference: the ACARA website, www.acara.edu.au/curriculum/learning-areas-subjects/science

While the curriculum was available for introduction from the beginning of 2012, the implementation differed across the jurisdictions with starting dates in 2012–2014. A detailed overview of the curriculum implementation schedule for each state and territory can be found on ACARA's Australian Curriculum website.

By 2015, all Australian Year 6 students had some exposure to learning science through the Australian Curriculum. However, given that the level of exposure to the curriculum across the states and territories is not consistent, alignment of NAP–SL to the Australian Curriculum will not occur until the next cycle in 2018. By this time, teachers and students will have had a minimum of four years' exposure to the Australian Curriculum: Science.

Special characteristics of the Australian Curriculum

The Australian Curriculum: Science has some characteristics that set it apart from previous curricula and align it well with other contemporary curricula in countries such as Singapore and Finland. The intent of these curricula in terms of the requirement for students to think and act scientifically, and in a reflective way is similar to the intent of the Australian Curriculum: Science. While thinking and acting scientifically is not new, the strengthening of the alignment of these long established ways of working in science to a socio-cultural context is new. The complete rationale for teaching science and the key ideas associated with this rationale can be found in the Australian Curriculum: Science version 8.2.

The science as a human endeavour strand provides a context for learning science for the other two strands – science understanding and science inquiry skills. As such, it is not simply a content strand and is therefore better considered as a context strand. This provides strong alignment with the definition for scientific literacy provided by PISA and adopted for the original NAP–SL in 2003.

Importantly, the definition currently used by PISA has evolved since then, and this change is reflected in the Australian Curriculum: Science most significantly through the science as a human endeavour strand.

In 2015, PISA defines scientific literacy[2] as the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen. A scientifically literate person is willing to engage in reasoned discourse about science and technology, which requires the competencies to:

- explain phenomena scientifically – recognise, offer and evaluate explanations for a range of natural and technological phenomena.
- evaluate and design scientific enquiry – describe and appraise scientific investigations and propose ways of addressing questions scientifically.
- interpret data and evidence scientifically – analyse and evaluate data, claims and arguments in a variety of representations and draw appropriate scientific conclusions.

The science understanding and science inquiry skills strands continue to develop the more traditionally taught aspects of the science curriculum (knowledge, conceptual understanding and skills). However, within the Australian Curriculum: Science, all three strands are intended to be delivered in an integrated manner. The structure of the units of questions developed for NAP–SL reflect this integrated approach when the context or stimulus for each unit of questions includes aspects pertaining to all three strands. Therefore, in order for students to perform well in science, it is critical that students are familiarised with this style of integrated thinking (*thinking and acting scientifically*).

---

2 Reference: OECD (2016), PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy, PISA, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264255425-en

The Australian Curriculum: Science can be described as a relatively low definition curriculum. This, in combination with the integrated nature of teaching the curriculum, provides teachers with time to develop the concepts in depth, rather than breadth, using an inquiry approach. This approach strongly supports the development of science literacy, and the way in which science literacy is assessed by NAP–SL.

The general capabilities in the Australian Curriculum also have relevance to this cycle of NAP–SL. While measurement of student performance across these dimensions is not a function of NAP–SL, future assessment cycles may provide opportunities to collect additional data and conduct analyses on aspects of the general capabilities, especially in the area of critical and creative thinking – an area where science provides many strong learning opportunities.

Science inquiry skills

The NAP–SL progress map (see table A1.6) can further assist teachers to understand the development of science inquiry skills. The progress map should be read in conjunction with the detailed description of the items presented in chapter 5, where the proficiency levels are discussed and specific examples of assessment items relevant to each level are provided. These item descriptions may be useful in providing ideas for teachers who want to construct criteria for student assessment performance in the science inquiry skills strand of the Australian curriculum as there is a high degree of alignment between the NAP–SL strands (A and B) and the Australian Curriculum: Science inquiry skills strand (see table A1.1).

School Release Materials - Assessment Items

School release materials are available to schools following each sample assessment.

As the 2015 NAP–SL is an online assessment, a stand-alone test comprising a selection of items will be available in the form of a public demonstration test on the NAP website. The mini-test will include:

- a range of item types
- a range of items across strands and sub-strands
- a range of difficulty.

Student survey – attitude and disposition

Chapter 8 examines student information on attitude and disposition collected by a survey that was completed by students at the end of the NAP–SL test. Some interesting results warrant attention.

Survey curriculum links

GO3 Value of science (see figure 8.1): results here support a positive view of science in our society but many students did not necessarily relate science to their own lives. The elements of the science as a human endeavour strand that encourage students to value science are:

- Science knowledge helps people to understand the effect of their actions; ACSHE051 and ASCHE062, for Years 3 and 4 respectively.
- Scientific knowledge is used to solve problems and inform personal and community decisions; ACSHE083 and ACSHE100, for Years 5 and 6 respectively.

The results of the survey indicate that students hear the messages relating to the usefulness of science within our society. However, it is possible that students do not perceive that they are part of the decision-making process. Therefore, units of work and associated investigations need to include contexts that directly involve the students. For example, a question for Year 2 students might be, 'What can my family do to reduce water consumption in our home?' This question focuses on the science understanding sub-strand, earth and space sciences ACSSU032, within the context of the science as a human endeavor sub-strand, use and influence of science, ACSHE035. Year 4 students could focus on the use or management of materials such as plastics within the context of the science understanding strand ACSSU074, and the science as a human endeavour strand ACSHE062, or perhaps investigate an erosion problem within their school grounds. The curriculum link for this investigation falls within the science understanding sub-strand, earth and space ACSSU075, and the science as a human endeavour sub-strand, use and influence of science ACSHE062.

GO11 The nature of science 2 (see figure 8.2) examines the nature of science itself. Interestingly, the overwhelming majority of students agreed that science was about making predictions and testing them. These elements are an essential part of the science as a human endeavour strand but are more evident in the science inquiry skills strand including:

- Year 4: ACSIS064 and ACSIS065
- Year 5: ACSIS231 and ACSIS086
- Year 6: ACSIS232 and ACSIS103.

While 68 per cent of students agreed with the statement that science is about describing patterns and relationships, the fact that 32 per cent disagree is somewhat disappointing as this refers directly to:

- ACSIS057 and ACSIS068 in Years 3 and 4 respectively: use a range of methods including tables and simple column graphs to represent data and to identify patterns and trends
- ACSIS090 and ACSIS107 in Years 5 and 6 respectively: construct and use a range of representations, including tables and graphs, to represent and describe observations, patterns or relationships in data using digital technologies as appropriate.

Importantly, the achievement standard specifies the knowledge, understanding and skills to be taught and learnt. For example, in Year 6: students collect, organise and interpret their data … describe and analyse relationships in data using appropriate representations. These processes all form part of scientific investigations, hence the importance of students regularly engaging in active investigations.

Scientific investigation is the fundamental methodology associated with thinking scientifically. Teachers can help students to develop their scientific thinking through the explicit teaching of the necessary science inquiry skills used in investigations, including:

- questioning and predicting
- planning and conducting
- processing and analysing data and information
- evaluating and communicating the results.

Multiple opportunities and appropriate scaffolding of learning is required for the development of these important skills.

Finally, critical thinking is an important and in-demand skill required by 21st century citizens. The link between critical thinking skills and the science inquiry strand can be further seen in the summary of the critical thinking skills in appendix 1, table A1.2 or in greater detail on the Australian Curriculum website.

In summary, teaching science inquiry skills strongly supports the development of important scientific, critical and creative thinking skills.

Student survey – student engagement

The survey findings about student engagement were encouraging in that students were generally very positive about science, in particular:

- interest in science
- self-concept in science ability
- science related activities at school and out of school
- how science was conducted in class (that is, pedagogical aspects) and topics covered.

However, a few other results are worthy of note:

- reading of books, newspapers and articles about science topics at school
  - 67 per cent of students responded that they read about science topics less than once a week or never

The use of printed material not only provides wonderful contexts for science learning, along with multimedia sources of information, but also provides opportunities to help students learn the literacy of science.

- During science lessons, I get to plan and carry out my own investigations:
  - 62 per cent of students responded sometimes or never.

The Australian Curriculum: Science requires students to do this from an early stage of schooling. For example:

Years 3 and 4: ACSIS053 and ACSIS065

» with guidance, plan and conduct scientific investigations to find answers to questions, considering the safe use of appropriate materials and equipment

Years 5 and 6: ACSIS086 and ACSIS103

» Identify, plan and apply the elements of scientific investigations to answer questions and solve problems using equipment and materials safely and identifying potential risks

- I use a computer to research or to present my science ideas and findings:
  - 49 per cent of students responded sometimes or never.

ICT is both a general capability and a learning area in its own right (Technologies curriculum) and is identified in the science inquiry skills strand. For example:

Year 6: ACSIS104 and ACSIS107

» Decide variables to be changed and measured in fair tests, and observe measure and record data with accuracy using digital technologies as appropriate

» Construct and use a range of representations, including tables and graphs, to represent and describe observations, patterns or relationships in data using digital technologies as appropriate

- How often do you have science lessons?
  - 30 per cent of students responded less than once a week or never.

Even though different implementation and timetabling models are used in schools, it would be difficult to see how meaningful science could be encouraged or learnt without regular access to sufficient time to do so.

- Our teacher invites visitors to school to talk about science topics:
  - 31 per cent of students responded affirmatively.

There is a growing body of research that clearly points to the benefits of contextualisation of science, which can be achieved quite simply by the inclusion of outside specialists. These could be parents in the school community who work in a science-based field, or a business in the community that uses aspects of science that would be of interest to students. CSIRO and many STEM (Science, Technology, Engineering and Maths) related industry groups, along with universities have developed outreach programs to help integrate these specialist areas with students' learning in schools. Exposure to science-based professionals at school can have a very important and positive effect on student perceptions of science-based careers. Currently, STEM is a major area of interest for the Australian Government and this is one of the strategies widely canvassed to promote improved learning in many skill areas, but especially critical and creative thinking, working collaboratively and problem-solving. The involvement of external specialists and people working in the area can be highly influential on student perceptions and dispositions.

- I think it would be interesting to be a scientist:
  - 69 per cent of students responded it would be interesting.

This is encouraging, and, interestingly, represents an increase since 2012.

- Our class goes on excursions related to the science topics we are learning about:
  - only 27 per cent of students responded 'yes'.

The importance of science in the real world cannot be overstated. Students enjoy completing science tasks outside of the classroom – the opportunities are varied and rich and can help to contextualise learning so that it relates to the real world. The involvement of outside expertise as discussed above, adds credibility and validity to out-of-classroom experiences and an opportunity to gather real data in the real world. Opportunities to combine science-learning with other learning areas; for example, mathematics, geography, history and physical education, also help to provide transference of skills and contextualisation. Science-based learning could occur in many different ways and across year levels. For example:

Year 3: ACSSU044 and ACSHE050

- Living things can be grouped on the basis of observable features and can be distinguished from non-living things; and science involves making predictions and describing patterns and relationships.

Year 4: ACSSU073 and ACSHE062

- Living things depend on each other and the environment to survive; and science knowledge helps people to understand the effect of their actions.

Year 5: ACSSU043, ACSSU078 and ACSHE083

- Living things have structural features and adaptations that help them to survive in their environment; the Earth is part of a system of planets orbiting around a star (the sun); and within the context of scientific knowledge is used to solve problems and inform personal and community decisions.

Year 6: ACSSU097 and ACSHE098

- Electrical energy can be transferred and transformed in electrical circuits and can be generated from a range of sources; and science involves testing predictions by gathering data and using evidence to develop explanations of events and phenomena and reflects historical and cultural contributions.

Results comparing groups of survey items with achievement data

Chapter 8 concludes with an analysis of the correlation between the student survey and science literacy performance.

While results are not conclusive for some aspects of the survey, some are noteworthy:

- G02 Self-concept of science ability: this is a measure of a students' own assessment of their abilities in science. Overall, self-concept of science ability is positively correlated with achievement in science.
- G11 The nature of science 2: this is a measure of what students consider science to be. For example, science is about using evidence to provide explanations. The examples used in the survey were all statements drawn from the Australian Curriculum. A positive correlation is clear and, as discussed above, an essential element of science literacy is to develop the broader conceptual awareness of the nature of science itself, inherent in the key ideas for science. These key ideas are developed through each year level description and connect the key ideas to the content. This relationship between key ideas, year level descriptions and content descriptions is an important one, as detail is added to the intent of the Australian Curriculum: Science; that is, to better understand the nature of science itself.
- G12 Who is involved in science: these survey questions were designed to gain a better understanding of how students perceive people who may be involved in contributing to science. Overall, this is also positively correlated to science achievement.

This next section will look at individual items and students' constructed responses to provide an insight into how teachers might use the Australian Curriculum to help improve student learning.

Examples of student responses to constructed response items

The following example items assess performance of a variety of proficiencies and require students to respond with a written response. If students are to demonstrate thinking scientifically, one way is to use their own words to respond to a question.

One consideration, when looking at student responses, is the relationship between a student's general literacy capability and their science literacy (see appendix 1, Literacy general capability). The general literacy demands of science stimulus material and items used in the NAP–SL test are appropriate for students in Year 6. This level is defined within the Australian Curriculum literacy continuum as Level 4 for Years 5 and 6 students.

However, many of the example student responses provided below show that students have sometimes failed to read the question properly. In some cases, students may have provided scientific information but have failed to actually answer the question. This highlights a more general need to teach literacy contextually and through all learning areas, including science.

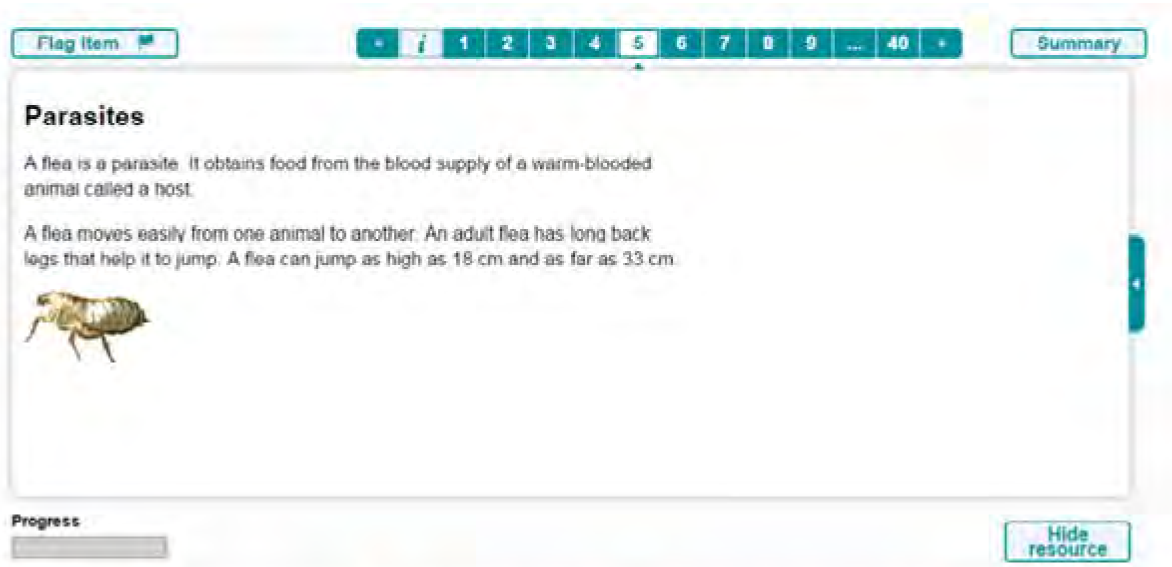
In terms of science literacy, NAP–SL is testing a student's ability to apply their science knowledge to real world science concepts. This requires analysis within particular contexts and an ability to connect the inherent science with the provided observations and data to the given context. That is, it requires students to use their thinking skills.

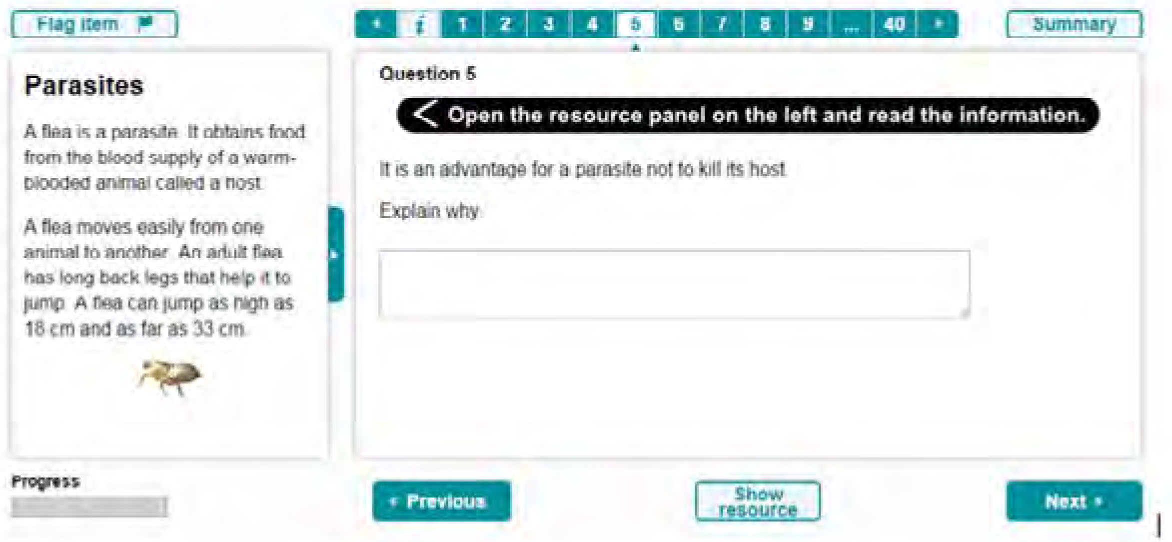
While the Australian Curriculum: Science does not specify a pedagogical approach, research supports the use of an inquiry approach to develop thinking skills. The Australian Curriculum: Science is a major contributor to the development of Critical and Creative Thinking capability and the Ethical Understanding capability. Science also contributes to the development of all the other capabilities, which is the intention for all learning areas.

Teachers might consider using the sample responses to questions to engage in discussions with their students about the appropriateness of responses to support the developmental process of learning to think critically, and then to apply this process to their own investigations in class.

Please note, the wording of the student responses on the following pages has been copied verbatim. Importantly, students are marked on the content of their responses and are not penalised for spelling or grammatical errors.

Example 1 and 2: Parasites





Example 1: This question asks students to examine information presented about parasites.

The proficiency level is 3.3 (see chapter 5 for a description of each level).

The question is linked to the content description in the Year 4 science understanding strand: Living things depend on each other and the environment to survive (ACSSU073).

Notably, the achievement standard for Year 4 requires students to describe relationships that assist the survival of living things. The elaborations provide some possible examples of contexts through which this could be taught:

- observing and describing predator-prey relationships
- predicting the effects when living things in feeding relationships are removed or die out in an area
- recognising that interactions between living things may be competitive or mutually beneficial.

Teaching to this content description should provide students with opportunities to explore the relationships that exist between living things and the environment, with the requirement to describe the nature of these relationships and how those factors assist in their survival.

The correct student responses below indicate an ability to connect the science with the observations presented in the question, or to put it another way, the effect with the cause. The notion of cause and effect is an important part of the key idea in science, patterns, order and organisation. The opportunity here, in teaching Year 4, is to incorporate this key idea into class discussion and to integrate the science as a human endeavour strand, science involves making predictions and describing patterns and relationships (ACSHE061). The content is actually secondary to the processes and simply provides the specific learning context.

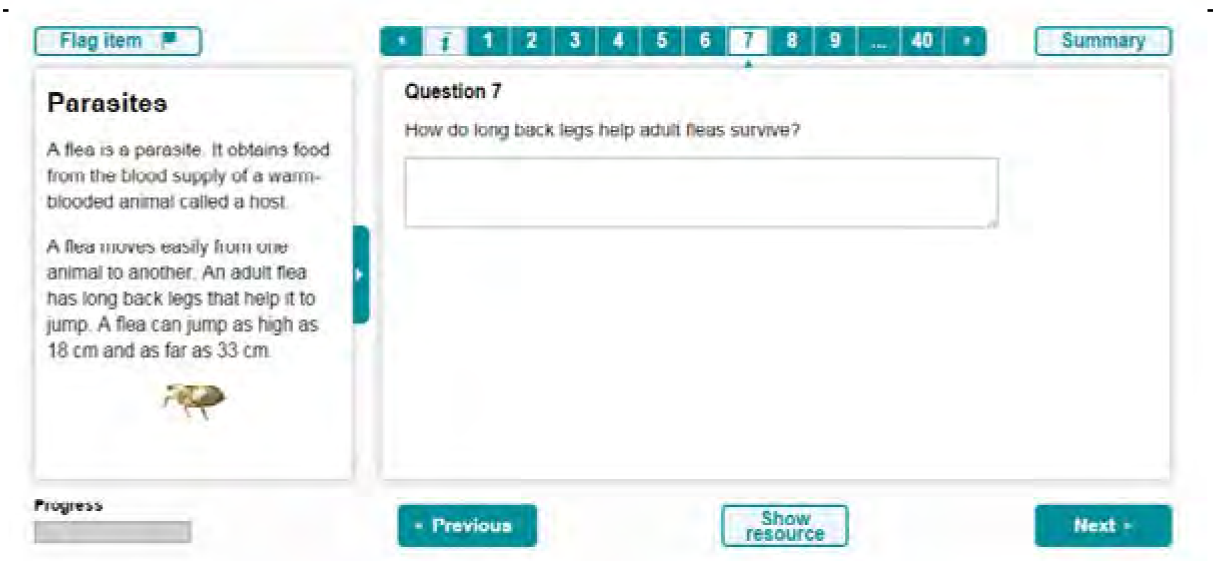
The following are examples of correct student responses gleaned from the NAP–SL 2015 assessment:

- It is an advantage because the longer the host lives, the longer the parasite gets a food supply.
- if it killed the host it would then have to go and find another host for it's food source. if the host stays alive it can feast all day long without finding another host.
- A flea needs lots of blood to stay alive so the don't kill the host so they have a source to drink blood from
- All a parasite need is to "obtain food" so it has no reason to kill the host

The following are examples of incorrect student responses from NAP–SL 2015: [comments]

- it is an advantage for a parasite not to kill it's host because then if it wants more blood they can just get more healthy blood [part way there, but does not answer the question]
- A flea cant really hurt an animal [irrelevant]
- A flea is a small creature and won't take as much blood from the animal compared with a bigger animal taking blood from it [irrelevant]
- An advantage for a parasite for it not to kill its host is because it can recieve more food by the host spreading it to other living creatures such as mammals [part way there, but does not answer the question]
- an advantage for a parasite not to kill its host is because fleas normally land on dogs and their owners spay them with flea killer which kills them [this child probably has responsibility for the family dog, but unfortunately it does not answer the question]
- becase then it will die of hunger so then there will be less fleas around [does not explain the cause and effect]

Example 2: Parasites



This question follows the previous. The level is 3.2 (see chapter 5 for details of proficiency levels) and it is linked to the Year 5 content description in the science understanding strand, living things have structural features and adaptations that help them to survive in their environment (ACSSU043).

The achievement standard requirement related to this content description is: students can analyse how the form of living things enables them to function in their environments. This links very clearly to the year level description for Year 5 that states: students are introduced to cause and effect relationships through an exploration of adaptations of living things and how this links to form and function. The key idea form and function is very clear here and builds on the Year 4 curriculum work as discussed above.

As for the previous question, students need to learn the thinking processes that are important by considering the relationship of flea and mammal (predator/host) and functionality of the legs (how long legs make a difference to the flea's mobility and how that will help it to survive). Specific investigations and examples of form and function, which allow students to develop these critical thinking processes, could be provided through the elaborations or other suitable contexts.

The following are examples of correct student responses gleaned from the NAP–SL 2015 assessment:

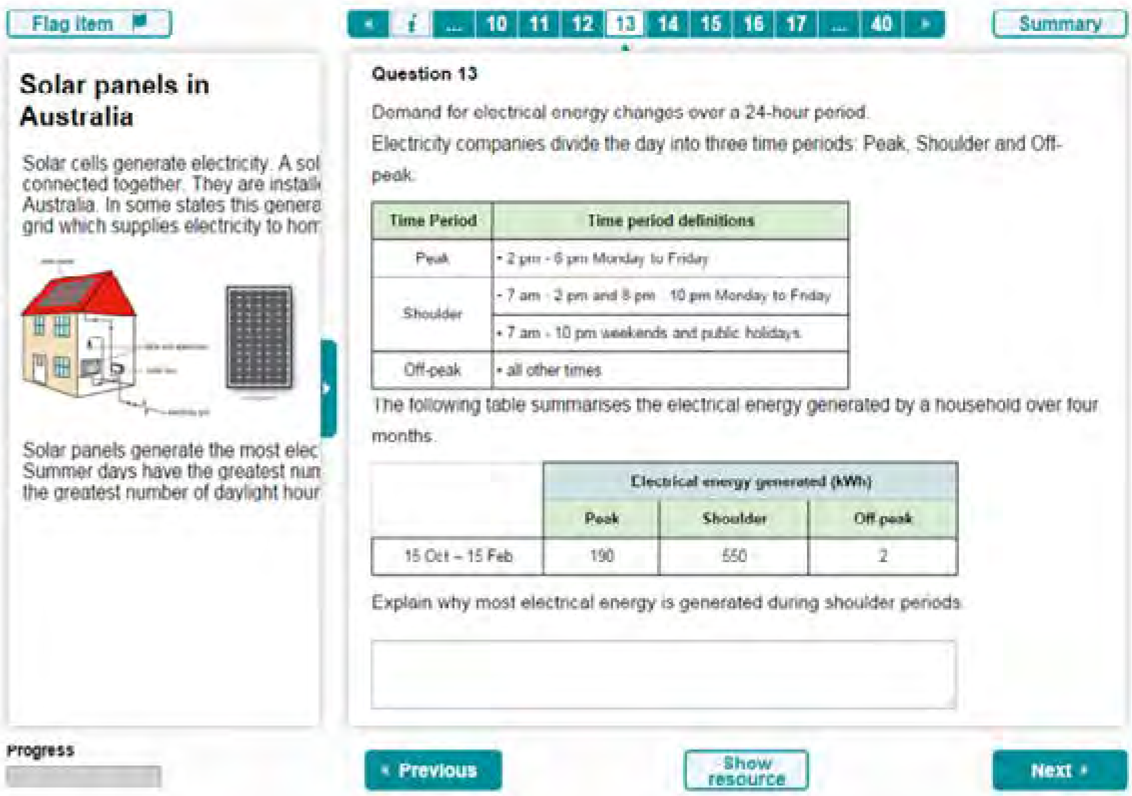
- A adult flea uses its back legs to survive by using its back legs to jump from animal to animal
- A fleas back legs can jump as high as 18 centimetres and as far as 33 centimetres which is a very far way for such a small creature, the legs of a flea / parasite would be considered as a way to jump from mammal to mammal which gives it a variaty or different mammals to choose from
- A flea's long black legs help it jump higher and greater distances. This way, it can quickly escape from danger

- because if the fleas are under attack they can quickly spring off
- because when you go to scratch where the flea bit you it can jump high from using its back legs.
- I think it is because the flea will have more leverage to jump from host to host
- it can jump away from potential threats

The following are examples of incorrect student responses from NAP–SL 2015:

- I think because if they grow it would grow bigger [has not answered the question]
- a long time until the owner or vet kills the parasite [irrelevant]
- because the adult fleas can jump really high with long legs [true, restates question, but does not answer the question]
- it blends in to black colour surfaces [irrelevant]
- it give the flea an advantage to jump higher and further [does not connect to its survival]

Example 3: Solar panels in Australia



This question has a proficiency level of 4 and above. Most students did not get this correct. It is linked to the content description in the Year 6 science inquiry skills strand, construct and use a range of representations, including tables and graphs, to represent and describe observations, patterns or relationships in data using digital technologies as appropriate (ACSIS107) and the achievement standard relating to this is: students construct tables and graphs to organise data and identify patterns in the data. The context for this learning could include the science as a human endeavour strand statement, science involves testing predictions by gathering data and using evidence to develop explanations of events and phenomena and reflects historical and cultural contributions (ACSHE098), which strongly connects the inquiry processes to the nature of science itself, that is, by learning about the way science works by doing it.

The specific content for this question relates to solar cells. This could be addressed through the Year 6 science understanding content description, electrical energy can be transferred and transformed in electrical circuits and can be generated from a range of sources (ACSSU097). However, students do not need this content to answer the question as the required processes are the thinking skills which are developed through the science inquiry skills strand and connected with the ability to use patterns to interpret data, compare data with predictions and use as evidence in developing explanations (ACSIS221).The achievement standard is: they collect, organise and interpret their data.

Finally, the year level description further identifies the importance of these processes for Year 6 with this statement: they learn how to look for patterns and to use these to identify and explain relationships by drawing on evidence. This accesses the key idea: patterns order and organisation and use of data such as these, especially when primary data are used (that is, data they have collected themselves) – it becomes their chance to use the evidence to support their conclusions, rather than reporting the results they 'should have obtained'.

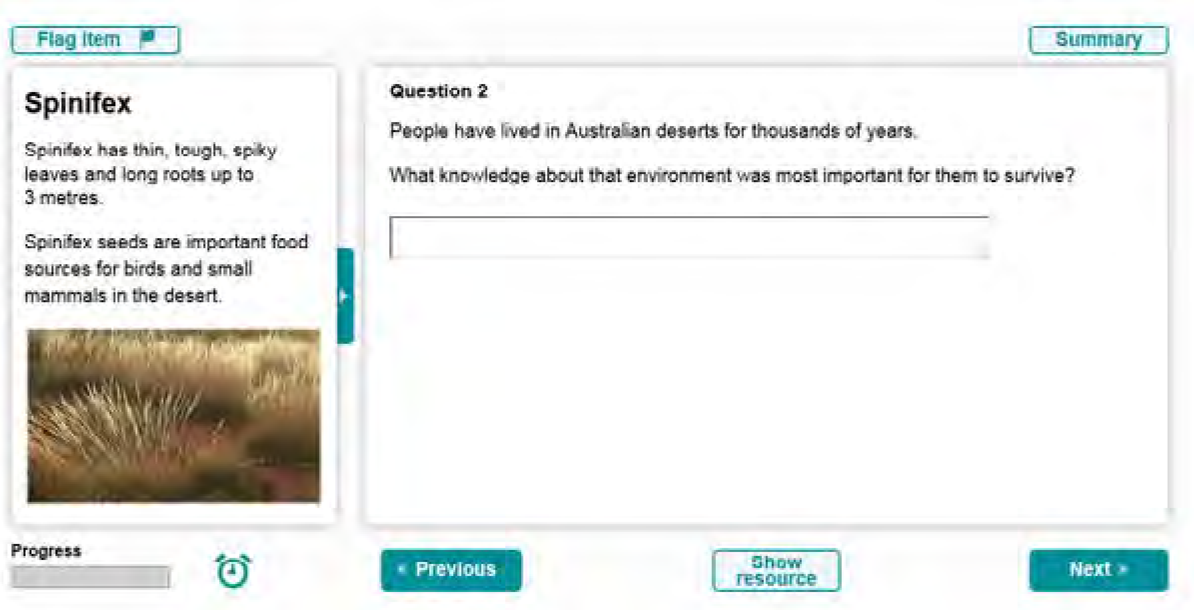
The following are examples of correct student responses gleaned from the NAP–SL 2015 assessment:

- Because during the shoulder periods the sun is out more especially during those moths of the year in Australia.
- because it is sunnier and it can generate more power from the sun.
- because it is the most time of the day when the sun is shining the most and it is hot and really bright
- Because that is the time that the sun is most likely the be high in the sky where all the sunlight reaches if there is nothing in the way. The shoulder time also includes midday which is when the sun is in the high peak in the sky which is the best time to collect energy from the sun.

The following are examples of incorrect student responses from NAP–SL 2015:

- largest amount of time [not true]
- because it is the time of year when we have more sunlight [also not true]
- because that's the most time that the electricity will be used [not necessarily true]
- At night people need light. Also 2 pm to 8 pm is when people are getting home and getting ready for dinner, maybe a movie or footy game, or maybe just to chill out to some relaxing music. [true but irrelevant]
- because there is more time in the shoulder period than the other time periods [incorrect]
- because during the "shoulder" hour the sun has fully reached the destination of where you live and in the "Peak" the sun has just peaked a bit because the world has not fully reached the country you live in and "Off-peak" means that the sun is leaving your country [response has nothing to do with the question]
- Because it is mostly when the sun is at its warmest and when it is warm people use air-conditioning [true but irrelevant]
- because it is on the weekend and public holidays and most people are home on the weekend or public holidays [irrelevant]

Example 4: Spinifex



This question has a proficiency level of 3.2 and it is linked to this content description in the Year 6 science as a human endeavour strand, science involves testing predictions by gathering data and using evidence to develop explanations of events and phenomena and reflects historical and cultural contributions (ACSHE098) but the question is also linked to the Year 6 science as a human endeavour content description, scientific knowledge is used to solve problems and inform personal and community decisions (ACSHE100). The achievement standard requirements are: students explain how scientific knowledge helps us to solve problems and inform decisions and identify historical and cultural contributions. This material could be explored when students are studying the content from the science understanding strand, the growth and survival of living things are affected by physical conditions of their environment (ACSSU094) or could have been studied in Year 5 when addressing, living things have structural features and adaptations that help them to survive in their environment (ACSSU043). The Year 5 science as a human endeavour strand content description (ACSHE081) is identical to the Year 6 human endeavor strand content description. However, the content is not essential for the students to be able to answer the question.

As for the previous questions, students need to engage with the thinking process that is important here by considering the process of knowledge acquisition for desert people in the context of the observations they are presented with.

The following are examples of correct student responses gleaned from the NAP–SL 2015 assessment:

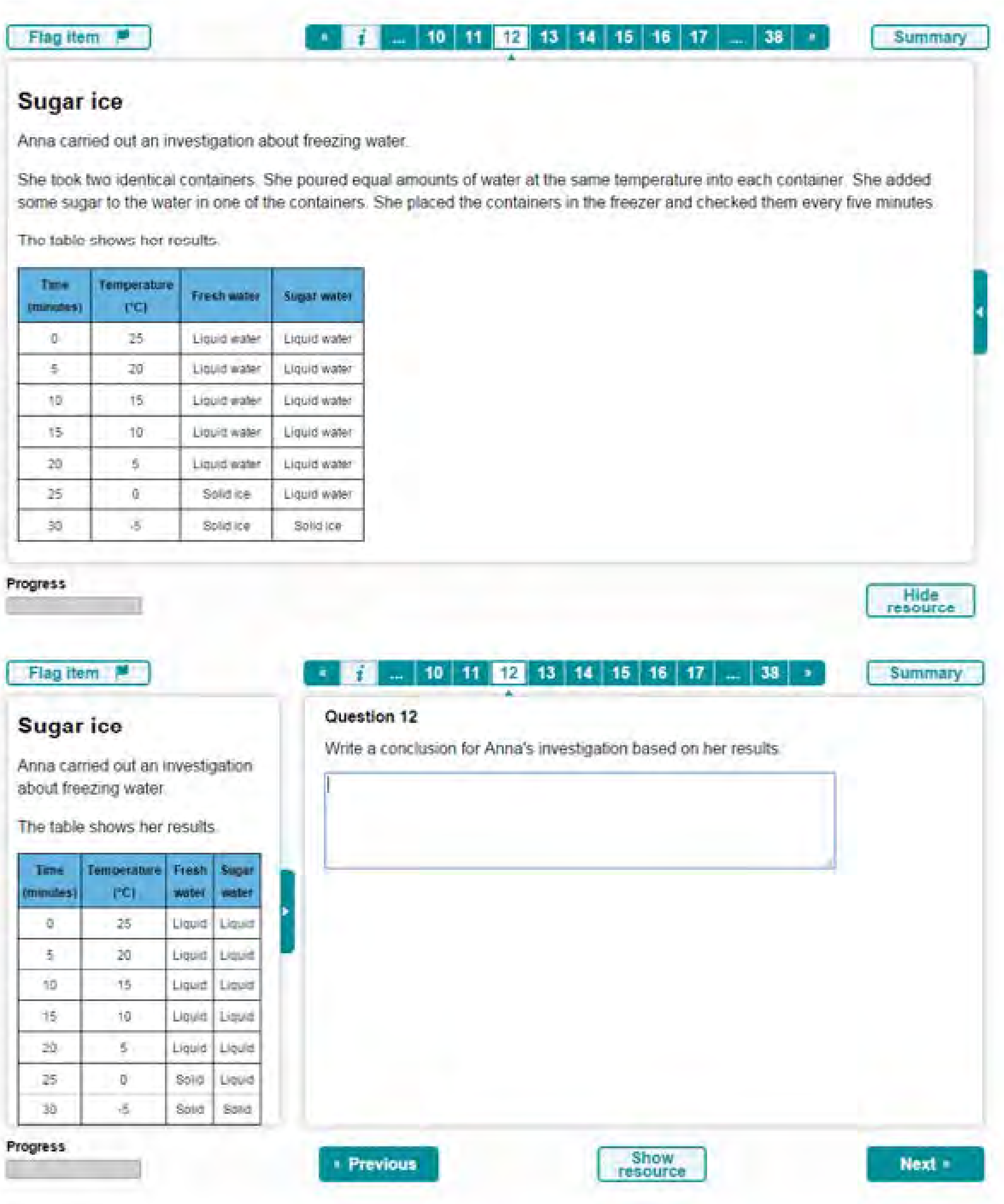
- i think the most important knowledge for them is what and what not to eat because eating the wrong thing can make you very sick or kill you
- It would be important to know where you can get food source from and water in plants. Also ways to make shelter out of plants.

- All the dangerous animals that could hurt such as snakes. The areas where they go to could also be dangerous it could be useful information for them
- Australia is a very dry country and because of this reason it would be most important for the people that live in Australia to know the wet and dry season and also no how to get water and where it is.
- how the seasons changed so they would know where to live in the colder times, and what types of food were okay to eat/not poison
- How to defend themselves and stay protected. They learnt what animals are ok to eat.

The following are examples of incorrect student responses from NAP–SL 2015:

- because they now how take from the land and some how give back [response is general; take what? Student needs to reference either food or water; or if plants edible/inedible; or if animals are dangerous]
- spinifex seeds is an important food source for small mammals in the desert [true but irrelevant]
- That small mammals hide in their and if you step in it a animal could attack you [restates given information, but does not answer question directly]

Example 5: Sugar ice



This question has a proficiency level of 3.3. It is the last question in a set of items presented to students. This item is linked to the content descriptions in the Year 6 science inquiry skills strand, compare data with predictions and use as evidence in developing explanations (ACSIS221) and communicate ideas, explanations and processes using scientific representations in a variety of ways, including multi-modal texts (ACSIS110). The achievement standard is: They describe and analyse relationships in data using appropriate representations and construct multimodal texts to communicate ideas, methods and findings. The context for this learning could include the acience as a human endeavour strand statement, science involves testing predictions by gathering data and using evidence to develop explanations of events and phenomena and reflects historical and cultural contributions (ACSHE098), which again strongly connects the inquiry processes to the nature of science itself; that is, learn about the way science works by doing it, and be aware that you are doing it.

Content for this question could be taught through a study of the chemical science sub-strand, changes to materials can be reversible or irreversible (ACSSU095). However, students do not need this content to answer the question as the required processes are the thinking skills developed through the science inquiry skills strand.

Finally, the year level description further identifies the importance of these processes for Year 6 within this statement: They learn how to look for patterns and to use these to identify and explain relationships by drawing on evidence accessing the key idea, patterns, order and organisation.

This example question provides many insightful examples of student responses, both correct and incorrect, that teachers could use to discuss with their students. Many students did not do well on this question. The question requires a combination of skills including reviewing and interpreting the data, linking these to the science that is occurring (change of state) and communicating that conclusion.

The following are examples of correct student responses from NAP–SL 2015:

- Anna was doing an investigation to see what freezes faster the water with sugar in it or no sugar in it. The water with the sugar in it took longer to freeze than the water with no sugar by 5 minutes
- Adding sugar into a container of liquid water menans it takes longer to freeze into solid ice.
- After 25 minutes the fresh water had turned solid but the sugar water was still liquid, and after 30 minutes at -5 degrees they were both solid. In conclusion fresh water freezers faster the water with sugar in it. Basically sugar slows down the process of water freezing.
- As shown on the table of Anna's results fresh water freezes five minutes quicker at a temperature 5 degrees warmer than natural water. This means that water that contains sugar freezes slower than natural water.
- Even though each container started with the same temperature of water, the water with sugar added took 5 minutes longer to freeze. The fresh water (without sugar) took approximately 25 minutes to completely freeze. On the other hand, the sugar water took about 30 minutes to freeze. We can also see from this investigation that sugar water's freezing point is lower then regular water, sugar water freezing at -5 degrees.

The following are examples of incorrect student responses from NAP–SL 2015:

- A conclusion is that the more time she left the water and sugar in th freezer the more chance it will go solid.

- Anna's investigation says that the the water and sugar is getting colder every five minutes by 5 so it is losing temperature every minute
- In conclusion if you freeze fresh water and sugar water to -5 from 30 minutes it will be solid however when the temperature gets warmer the water stays as a liquid. So as a result it is better if you want to freeze something you should put the temperature to -5.
- it drops 5 degrees every 5 minutes
- I think that because the sugar made the temperature colder cause of the amount of sugar [not a valid conclusion]

Conclusion

This section provides an historical perspective to the Australian Curriculum: Science and an overview of the connections between the Australian Curriculum: Science and NAP–SL together with suggestions to support student science learning in the classroom.

It is important to note that the NAP–SL school release materials from previous cycles are still available on ACARA's NAP website. These tests can be downloaded and printed and used by teachers for assessment purposes, where the context is appropriate to classroom teaching and learning programs. Alternatively, these materials can be used by teachers as a model to develop their own assessments. For the 2015 cycle, an online public demonstration test comprising a small selection of items will provide opportunities for students, teachers and parents to familiarise themselves with the online platform.

Chapter 10

Conclusion

Chapter 10. Conclusion

Managing innovation and change is both a challenge and an opportunity for all large-scale assessment programs. Over a number of cycles, the NAP sample assessment – science literacy (NAP–SL) has successfully accommodated a raft of incremental changes. This has been achieved without compromising the instrument's capacity to measure student performance and provide valid comparisons over time.

Since 2003, NAP–SL has increased in the scope and depth of topics assessed and ideas examined. In 2003 the first science literacy scale was adopted. In 2006 a more inclusive test design was implemented, resulting in a sample that was more inclusive of remote schools and items that provided better discrimination among students. The number of objective tests and the number of assessment items were also increased in the 2006 assessment.

In 2009 the student survey was introduced, providing additional insights into how students perceive science. The survey was expanded in 2012, which also saw refinements to the progress map.

The transition from paper-based to online assessment in 2015 marks a significant milestone for NAP–SL. The new NAP online test delivery system has provided opportunities to explore the use of innovative item types and as a result, students have experienced a more interesting and engaging assessment event. With the advent of the Australian Curriculum: Science, the first step towards alignment was explored in this cycle. Curriculum specialists worked together to compare the existing assessment framework with the new Australian Curriculum: Science. A mapping exercise was conducted and strong alignment was identified. Furthermore, aspects of the Australian Curriculum sub-strand, science as a human endeavour, were incorporated in both stimulus material and two new survey questions.

These changes and innovative practices have provided additional opportunities to explore how Australian students engage with science, solve problems and make sense of the world around them.

The process of continuous improvement has enabled NAP–SL to provide richer insights into how young Australians make sense of science during their primary school years.

Student achievement in science literacy from 2006 to 2015

Student results are reported against five proficiency levels (Level 2 and below, Level 3.1, Level 3.2, Level 3.3 and Level 4 and above) with Level 3.2 being described as the proficient standard. The distribution of students across the achievement levels at the national level has remained relatively stable since 2006.

In 2015, 55.1 per cent of students at the national level attained the proficient standard or better in science literacy. In 2012, the percentage was 51.4 per cent, in 2009 the percentage was 51.9 per cent and in 2006 it was 54.3 per cent. The differences in the percentage of students attaining the proficient standard across the five assessment cycles are not statistically significant. Chapter 4 of this report provided detailed information about the performance of students nationally and at state or territory levels in NAP–SL.

Factors associated with achieving science literacy

As outlined in chapter 7 of this report, student background characteristics are related to achievement of science literacy. Background data about gender, Indigenous status, language background and geographical location were collected.

At the national level, female students significantly outperformed male students. This is a notable change for NAP–SL. In all previous cycles, the difference between performance levels of female and male students was not statistically significant.

Nationally, non-Indigenous students achieved significantly higher levels of science literacy than Indigenous students, as was also the case in 2012, 2009 and 2006. This finding is similar to that of other national assessment programs and once again indicates that strategies need to be found to address the gap in achievement between the two groups.

Students living in metropolitan areas achieved the highest mean score in science literacy. Their results were significantly different from those of students living in provincial areas and those of students living in remote and very remote areas. Similar findings are evident in the national assessment programs in literacy, numeracy, civics and citizenship, and information and communication technology literacy.

Students from English-speaking backgrounds also achieved slightly higher means nationally than students from language backgrounds other than English. However, the difference was not statistically significant. A similar trend was observed in the 2012 and 2009 assessments.

Student survey

As discussed in chapter 8, students were surveyed on their perception of and engagement with science. Various relationships were found between survey responses and other aspects of this study. The survey provided interesting insights into students' perceptions of and attitudes to science and their experiences with science learning at and outside school.

Approximately 86 per cent of students responded that they would like to learn more science at school, indicating that a positive attitude towards this subject area exists. Students also showed a high degree of agreement with the view of science outlined in the science as a human endeavour strand of the Australian Curriculum: Science. In particular, 87 per cent of students agreed that people from all cultural backgrounds in Australia are involved in science, and 96 per cent of students agreed that both women and men are involved in science.

The responses to the student survey will guide further survey development for future cycles and can provide impetus for discussion at school and jurisdictional levels regarding students' perceptions of, attitudes towards, and experiences of, science in their lives.

References

Jones, R. (2004) Geolocation Questions and Coding Index: A technical report submitted to the Performance Measurement and Reporting Taskforce of the Ministerial Council on Education, Employment, Training and Youth Affairs. MCEETYA.

Martin, M.O., Mullis, I.V.S., Foy, P. & Stanco, G.M. (2012) TIMSS International Results in Science. Amsterdam: IEA and Chestnut Hill, MA, USA: TIMSS and PIRLS International Study Center.

OECD Programme for International Student Assessment (1999) Measuring Student Knowledge and Skills: A new framework for assessment. Paris: OECD.

Thompson, S., (2013) PISA 2012: How Australia Measures up. Melbourne: Australian Council for Educational Research Ltd.

Appendix

1

National Year 6 Primary Science Assessment Domain

Appendix 1. National Year 6 Primary Science Assessment Domain

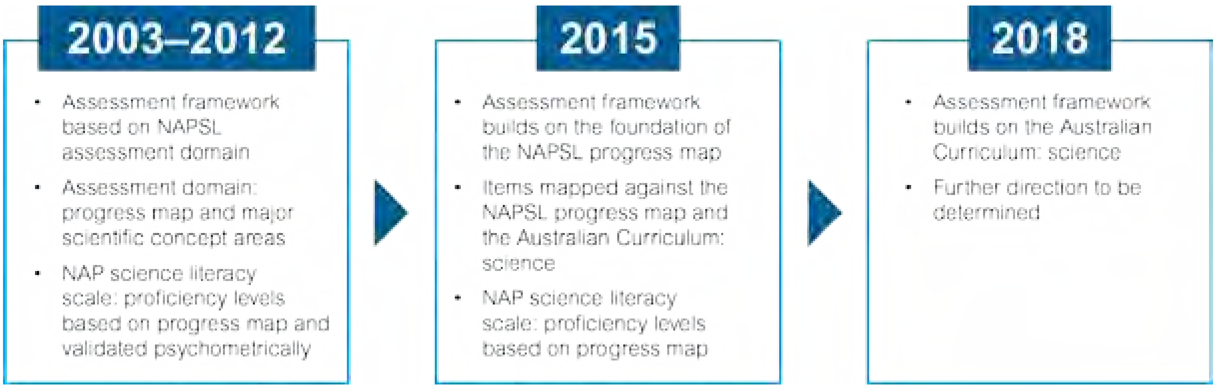
The complete NAP sample assessment – science literacy (NAP–SL) assessment framework bridging document can be found at the assessment frameworks page of the NAP website.

Introduction and background

This appendix provides a description and rationale for the framework that formed the basis of the 2015 NAP–SL. This framework is intended to serve as a bridge between past and future NAP–SL cycles (see figure A1.1).

Figure A1.1

Development of the NAP–SL assessment framework and science literacy scale



One of the main objectives of NAP–SL is to monitor trends in science literacy performance over time. To enable effective historical comparison, it was important that the underlying construct of the NAP–SL assessment was maintained. At the same time, the new Australian Curriculum: Science provided an opportunity to bring in aspects of science literacy that had not been considered in previous cycles.

As a consequence, this framework was intended to describe the existing NAP–SL construct in terms of the Australian Curriculum: Science and augment it in ways that reflect developments in the Australian Curriculum. It is expected that future NAP–SL assessments will be based mainly on the Australian Curriculum: Science.

Additionally, this cycle marked a major advance in the way the NAP–SL construct was assessed. For the first time, NAP–SL was assessed online, with items developed to be compatible with the IMS Question & Test Interoperability™ Specification.

Science literacy assessment domain

NAP–SL assesses science literacy in the context of a student's ability to apply broad conceptual understandings of science in order to make sense of the world; to understand natural phenomena; and to interpret media reports about scientific issues. It also includes the ability to ask investigable questions; conduct investigations; collect and interpret data; and make informed decisions. This construct was developed from an earlier definition of science 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, page 60)*

Science literacy progress map

For previous cycles of NAP–SL, a science literacy progress map was developed based on the agreed definition of science literacy and an analysis of the existing state and territory curriculum and assessment frameworks. The progress map describes the development of science literacy across three strands:

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

Strand B: 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: using science understandings for describing and explaining natural phenomena; and for interpreting reports about phenomena.

Major scientific concept areas

The existing science literacy domain comprises four major scientific concept areas: Earth and space; Energy and force; Living things; and Matter. Each concept area includes a set of concepts – broad statements of scientific understandings that Year 6 students would be expected to demonstrate – found most widely in the various state and territory documents and strongly aligned to the Australian Curriculum.

An illustrative list of examples for each of the concept areas provides elaboration of these broad conceptual statements and, in conjunction with the science literacy progress map, was used as a guide in the development of assessment items.

The Australian Curriculum: Science

In 2010, the federal, state and territory education ministers of Australia endorsed the release of the Australian Curriculum: Science. Since that time, the states and territories have been working on the implementation of the new curriculum in schools.

The Australian Curriculum: Science requires students to develop an understanding of important science concepts and processes; the practices used to develop scientific knowledge; and science's contribution to our culture and society and its applications in our lives.

Accordingly, the Australian Curriculum: Science has three interrelated strands – science as a human endeavour, science inquiry skills, and science understanding – which are designed to be taught in an integrated way. Together, these three strands provide students with understanding, knowledge and skills through which they can develop a scientific view of the world. Students are challenged to explore science, its concepts, nature and uses through clearly described inquiry processes. Table A1.1 lists the strands of the curriculum and the sub-strands within each strand.

Table A1.1

Strands and sub-strands in the Australian Curriculum: Science

Strands	Sub-strands
Science understanding	Biological sciences
	Chemical sciences
	Earth and space sciences
	Physical sciences
Science as a human endeavour	Nature and development of science
	Use and influence of science
Science inquiry skills	Questioning and predicting
	Planning and conducting
	Processing and analysing data and information
	Evaluating

Previous cycles of NAP–SL were developed with no common science curriculum across the states and territories. With the implementation of the Australian Curriculum: Science in all states and territories in 2014, it was important that the NAP–SL construct was mapped onto and described in terms of the new Australian Curriculum: Science as a first step in the alignment process.

Table A1.2 shows how the NAP–SL progress map strands map onto the strands/sub-strands of the Australian Curriculum: Science.

Table A1.2

The NAP–SL progress map links to the strands/sub-strands of the Australian Curriculum: Science

The NAP–SL progress map strands	Australian Curriculum: Science strands/sub-strands
Strand A: formulating or identifying investigable questions and hypotheses; planning investigations; and collecting evidence	Science inquiry skills – questioning and predicting Science inquiry skills – planning and conducting Science as a human endeavour
Strand B: interpreting evidence and drawing conclusions from students' own or others' data; critiquing the trustworthiness of evidence and claims made by others; and communicating findings	Science inquiry skills – processing and analysing data and information Science inquiry skills – evaluating Science inquiry skills – communicating Science as a human endeavour
Strand C: using science understandings for describing and explaining natural phenomena; and for interpreting reports about phenomena	Science understanding Science as a human endeavour

There is a high degree of alignment between NAP–SL progress map Strand A and Strand B and the science inquiry skills strand of the Australian Curriculum: Science.

The NAP–SL progress map Strand C provides an abstract representation of progression in students' use of science concepts for describing and explaining natural phenomena and interpreting reports about phenomena that makes no reference to particular science concepts.

In previous NAP–SL cycles, Strand C has provided guidance for the development of items that reflect levels of increasing complexity and abstraction in students' understanding of science concepts. In the absence of a common science curriculum across states and territories, the major scientific concept areas provided the contexts and specific concepts used to assess science understanding.

Table A1.3 shows how the NAP–SL major scientific concept areas map onto the sub-strand of the Australian Curriculum: Science – science understanding. The science understanding strand of the Australian Curriculum provides guidance about the specific concepts to be assessed in the NAP–SL tests. Table A1.7 shows the mapping between the major scientific concept areas and the science understanding strand at a finer level. There is no explicit equivalent of the abstracted progression articulated in Strand C in the Australian Curriculum: Science.

Table A1.3

The NAP–SL major scientific concept areas link to the sub-strand of the Australian Curriculum: Science – science understanding

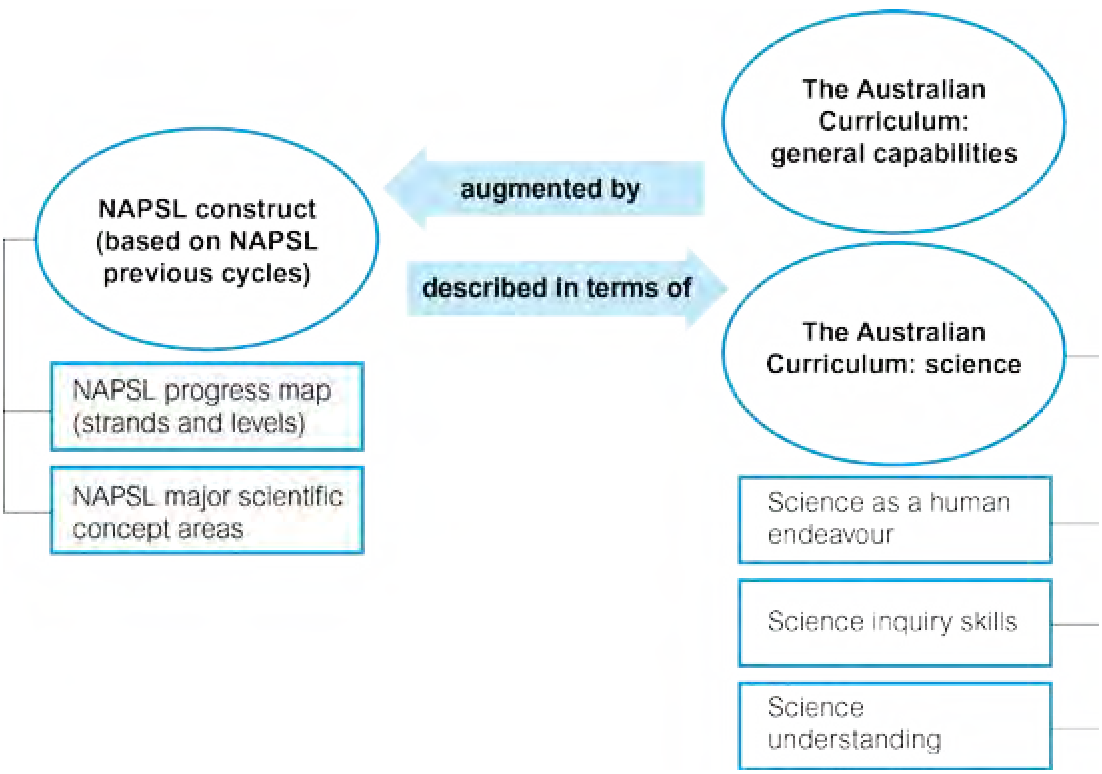
The NAP–SL major scientific concept areas	Australian Curriculum: Science – science understanding
Earth and space	Earth and space sciences
Energy and force	Physical sciences
Living things	Biological sciences
Matter	Chemical sciences

The 2015 NAP–SL Assessment Domain

One of the main objectives of NAP–SL is to monitor trends in science literacy performance over time. To enable effective historical comparison, it was important that the underlying construct of the NAP–SL assessment was maintained. At the same time, the Australian Curriculum: Science provided an opportunity to consider aspects of science literacy that had not been assessed in past cycles (see figure A1.2).

Figure A1.2

2015 NAP–SL construct



As in previous cycles, the conceptual framework for the assessment of science literacy comprised the NAP–SL progress map, which describes growth in science inquiry skills and conceptual understandings (see table A1.6). The progress map Strands A and B were mapped against and augmented by the content descriptions articulated in the *science inquiry skills strand of the Australian Curriculum: Science.*

Progress map Strand C describes a progression of how students use science concepts for describing and explaining natural phenomena and remained relevant for the 2015 cycle as it articulates performance levels that link directly to proficiency levels on the science literacy scale.

In previous cycles, the major scientific concept areas provided an indication of the pool of concepts from which item content related to science understanding was drawn. The science understanding strand of the Australian Curriculum: Science provided more specific content descriptions and were used in the 2015 NAP–SL cycle as the basis for developing items related to the understanding of science concepts. As noted in the previous section, there is a high degree of alignment between the major scientific concept areas and the science understanding strand content descriptions.

Items in previous NAP–SL cycles did not assess understandings related to the scientific endeavour (nature and development of science; use and influence of science). In the 2015 NAP–SL cycle, the science as a human endeavour strand of the curriculum informed both stimulus context and attitudinal aspects of the student survey. Where appropriate, items were also classified against content descriptions related to this strand. In future cycles, the assessment domain will be further expanded to explicitly include the assessment of student understanding of the nature and development of science.

Given that the NAP–SL test instruments were constructed within constraints of test length, the content covered in the test was intended to be a sampling of available concepts rather than an exhaustive assessment of all the concepts listed. The focus of the assessment was on concepts and skills from the Australian Curriculum: Science, Years 4–6. However, as the Australian Curriculum represents a continuum, it was appropriate that concepts and skills from Foundation through to Year 6 be considered. This is consistent with the approach taken in previous NAP–SL cycles, in which the progress map articulates a progression in development of understanding and skills.

The Australian Curriculum: general capabilities

The Australian Curriculum includes seven general capabilities. The capabilities identified as being most relevant and appropriate to the assessment of science, and hence reflected in NAP–SL, include:

Literacy: aspects of the literacy capability are found within the reading comprehension demands of both the stimuli and the items of NAP–SL.

Numeracy: aspects of the numeracy capability are found within NAP–SL, including the reading and construction of graphs and tables, calculations and measurement, as well as some elements of spatial reasoning.

Information and communication technology (ICT): aspects of the ICT capability arose from online delivery.

Critical and creative thinking: aspects of the critical and creative thinking capability arose from important cognitive skills inherent in scientific inquiry.

Items and stimulus could draw on aspects of the personal and social capability, the ethical understanding capability and the intercultural understanding capability, when appropriate. The following sections describe in more detail how the relevant capabilities were planned to be reflected in NAP–SL assessment. It should be noted though that the focus of the NAP–SL was the assessment of science literacy and not of general capabilities.

Literacy

In the NAP–SL tests, students were required to read and comprehend written and graphical stimuli. For some items, they were also expected to write coherent explanations of ideas.

While literacy plays an important role in science learning and assessment, it was important that the difficulty of items and stimuli did not derive mainly from the amount and the complexity of the stimulus material and instructions. The NAP–SL stimuli and items were written to a level appropriate for the students assessed. The literacy demand of items was monitored by expert review to ensure that it was at an appropriate level for the assessment.

Numeracy

Many elements of numeracy are evident in the Australian Curriculum: Science, particularly in *science inquiry skills*. These include practical measurement and the collection, representation and interpretation of data from investigations. Comparison between the Australian Curriculum: Science and the Australian Curriculum: Mathematics also shows a number of overlaps.

In the NAP–SL assessments, students were expected to show dispositions and capacities to use appropriate mathematical knowledge and skills as outlined in the Australian Curriculum: Science.

Information and communication technology

ICT plays a role when students take an online assessment. However, while students were interacting with technology to complete the assessment, it was not intended that the technology be a source of difficulty within NAP–SL. Consequently, it was important that access to, and navigation within, the test required relatively basic ICT skills.

Critical and creative thinking

The critical and creative thinking continuum is described in the Australian Curriculum as consisting of two interrelated aspects:

> *Critical thinking is at the core of most intellectual activity that involves students in learning to recognise or develop an argument, use evidence in support of that argument, draw reasoned conclusions, and use information to solve problems. Examples of thinking skills are interpreting, analysing, evaluating, explaining, sequencing, reasoning, comparing, questioning, inferring, hypothesising, appraising, testing and generalising.*
>
> *Creative thinking involves students in learning to generate and apply new ideas in specific contexts, seeing existing situations in a new way, identifying alternative explanations, and seeing or making new links that generate a positive outcome.*
>
> *This includes combining parts to form something original, sifting and refining ideas to discover possibilities, constructing theories and objects, and acting on intuition.*

The critical and creative thinking continuum organises these concepts into four elements.

Figure A1.3

Elements of critical and creative thinking

These elements are further subdivided to create 12 sub-elements (see table A1.4).

Table A1.4

Organising elements and sub-elements of critical and creative thinking

Organising element	Sub-element
Inquiring – identifying, exploring and organising information and ideas	Pose questions
	Identify and clarify information and ideas
	Organise and process information
Generating ideas, possibilities and actions	Imagine possibilities and connect ideas
	Consider alternatives
	Seek solutions and put ideas into action
Reflecting on thinking, actions and processes	Think about thinking (metacognition)
	Reflect on processes
	Transfer knowledge into new contexts
Analysing, synthesising and evaluating information	Apply logic and reasoning
	Draw conclusions and design a course of action
	Evaluate procedures and outcomes

All elements/sub-elements of critical and creative thinking have a place within science education. However, only some of them can be overtly assessed within the parameters of the NAP–SL construct.

Different aspects of critical and creative thinking play different roles within an assessment. For example, many items within previous NAP–SL tests have included strong elements of logical reasoning. Other items targeted the evaluation of procedures and outcomes. More generally, the use of complex and rich stimulus within the NAP–SL assessments results in nearly all item sets addressing, at varying degrees of sophistication, a student's ability to transfer knowledge into new contexts.

When appropriate, the NAP–SL items were tagged against relevant statements (capability descriptions) from this continuum. This was done when there was a reasonable inference that students who answer the item correctly will have engaged in the relevant cognitive skill. Language from the continuum also helped inform item descriptors.

Items were not written to specifically address critical and creative thinking. Instead, the tagging of items against this capability will enable further discussion and analysis of how the NAP–SL addresses this capability.

Online testing

Previous cycles of NAP–SL have been delivered using a printed document (otherwise known as paper-based testing). In 2015, NAP–SL was delivered as a computer-based assessment.

The item types available for 2015 NAP–SL included, but were not limited to, those in the following table (table A1.5).

Table A1.5

Item types available for 2015 NAP–SL

Main type	Description	Paper-based equivalent
Multiple choice/choices	Select one option from radio buttons	Multiple choice
	Select one or more options from check boxes	Multiple select
Extended text / text entry	Type text (manually scored) into box as separate paragraph	Short constructed response Long constructed response
Interactive gap match	Drag from a set of source options (text or images) into blank spaces within a passage of text or into a table	Not applicable
Match	Select which source objects match which destination categories by clicking a grid of radio buttons or checking checkboxes	Used for survey Likert style responses in the survey items
Hotspot	Select one or more predefined areas (circle, rectangle and polygon) on an image	Typically equivalent to multiple choice or multiple select with graphical options but with great freedom of layout and formatting
Interactive graphic gap match / position object	Drag objects (images/text) to hotspots on an image	Not applicable – but in some cases may be equivalent to multiple choice
Select point	Select a point on a background image	Used to simulate plotting points on a grid or graph
Composite	A combined item type that allows students to select an option and then write an explanation of their choice	This item type simulates a style of item used in past NAP–SL cycles, where students could choose from two alternatives and then explain their choice

Aside from the extended text item type, these item types can be more generally classified as forms of selected response. However, unlike with more traditional multiple-choice items, students can be presented with a much greater range of choices. Practice items were included in the assessment to make students familiar with the various item types used.

To help maintain a consistent construct with past NAP–SL cycles, a similar proportion of long constructed response items (using the extended text item type) were included. These items required students to give longer, more open-ended responses that were then marked by expert markers in a similar way to the marking of long constructed response items in previous cycles.

Science inquiry and modifications to practical assessment tasks

Previous cycles of NAP–SL have included a practical component. The purpose of this component was to provide students with an opportunity to experience practical aspects of science within a formal assessment and test the conventions of science literacy in more depth than was possible in the objective component.

The practical component was not intended to be an assessment of a distinctly separate construct of ‘practical science’ or to provide a sub-scale measure of practical skills that might provide complementary information to a more general science scale. Instead, the test items in the practical section of the NAP–SL were intended to be a part of the same science literacy scale. This meant that at the item development, the piloting and the post-trial levels, practical items were judged against their performance with the objective items.

The approach for previous NAP–SL practical tasks was to use a two-stage structure. In the first stage, students participated in a science practical group work task that was classroom-based. Students then individually answered a range of items on the practical task they had just completed. These items included assessments of skills related to completing an experiment, including data representation tasks, writing conclusions and evaluating aspects of an experiment.

The 2015 NAP–SL was delivered online. Consequently, the previous approach for the NAP–SL practical tasks (hands-on investigations carried out by a group of students followed by individual responses to items) was no longer viable. At the same time, it was desirable that within online delivery, the 2015 NAP–SL content be as comparable as possible with previous NAP–SL cycles to allow comparisons of science literacy performance over time. Therefore, the 2015 NAP–SL contained a component that achieved similar objectives as the previous NAP–SL practical, but by different means.

A new approach for the NAP–SL online ‘practical’ task:

- highlighted the value of practical work within the science curriculum and its role in science literacy
- delivered valid and reliable data in the same way as the other component of the NAP–SL test
- addressed the same skill areas as the practical items in past NAP–SL cycles
- assessed/aligned with the relevant skills in the historical NAP–SL progress map
- assessed/aligned with the relevant skills in the new Australian Curriculum: Science

- used items that could be written within the subset of item types available in the prescribed item authoring system
- was compatible with the test delivery system and the minimum technology requirements for participating schools (for example, screen size, appropriate file types, bandwidth limitations, etc.).

Linear staged task

An approach was developed after evaluation of different approaches to online tasks, taking into account the capabilities of the 2015 NAP–SL item-authoring and test-delivery systems and the available technology in schools.

The tasks presented were essentially closed, although the approach included some constructed response items. Students were placed in a role more like observers of a practical task rather than active participants. This means that students were not directly engaged in a practical activity but were tested on a range of relevant science inquiry skills. The approach was enhanced with video stimuli. To be sufficiently engaging, tasks with strong visual components were expected to be most effective.

To prevent dependencies, navigation needed to be more restricted than it was in the objective test. For example, if students were asked to make a prediction early in the task they needed to be prevented from changing that prediction once they had seen either the final results or the ‘prediction’ of an on-screen character.

It is hoped that this approach provided good examples for primary schools of online assessment tasks in the science inquiry skills strand.

Stimuli and items embedded in the online tasks (and more generally in the NAP–SL assessments) use the term ‘investigation’ to refer to the process of answering a question, exploring an idea or solving a problem that requires activities such as planning a course of action, collecting data, interpreting data, reaching a conclusion and communicating these activities. Students were not expected to know that an experiment is an investigation that tests a hypothesis, and the term ‘experiment’ was not used in item stimulus or items. This is consistent with the treatment of the concepts ‘investigation’ and ‘experiment’ in the Australian Curriculum, where the term ‘experiment’ first appears in Year 7.

The NAP–SL progress map

During the previous cycles of NAP–SL, a science literacy progress map was developed based on the construct of science literacy and an analysis of state and territory curriculum and assessment frameworks. The progress map describes the development of science literacy across three strands:

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 communication technologies.

Strand B: interpreting evidence and drawing conclusions from students' 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; questioning and examining the findings and conclusions of 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 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.

Science literacy has been described 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. The NAP–SL progress map (see table A1.6) describes progression in four levels from 1 to 4 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.

Strand C has been abstracted and makes no reference to particular science concepts or contexts.

NAP–SL focuses on Levels 2, 3 and 4 of the progress map, the levels of science literacy attained by students in Year 6. The agreed proficiency levels serve to further elaborate the progress map.

Table A1.6

NAP–SL progress map Levels 1 to 4

Level	Strands of science 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 students' 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 science understandings for describing and explaining natural phenomena, and for interpreting reports about phenomena.** **Conceptual strand: applies conceptual understanding.**
4	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. Summarises conclusions and explains patterns in science data. Makes 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.
3 [includes 3.1, 3.2, and 3.3]	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. Generalises and applies the rule by predicting future events.
2	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. Completes 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.
1	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.

The NAP–SL major scientific concept areas

Table A1.7

NAP–SL major scientific concept areas

Major scientific concept areas	Examples	Australian Curriculum: Science links
Earth and space		
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.	Features of weather, soil and sky, and effects on me. People use resources from the earth; need to use them wisely. Sustainability	Daily and seasonal changes in our environment, including the weather, affect everyday life (F, ACSSU004). The earth's resources, including water, are used in a variety of ways (Year 2, ACSSU032). The earth's surface changes over time as a result of natural processes and human activity (Year 4, ACSSU075).
The changing earth: The earth is composed of materials that are altered by forces within and on its surface.	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). Climate change.*	Observable changes occur in the sky and landscape (Year 1, ACSSU019). The earth's surface changes over time as a result of natural processes and human activity (Year 4, ACSSU075). Light from a source forms shadows and can be absorbed, reflected and refracted (Year 5, ACSSU080). Sudden geological changes or extreme weather conditions can affect the earth's surface (Year 6, ACSSU096). Global systems, including the carbon cycle, rely on interactions involving the biosphere, lithosphere, hydrosphere and atmosphere (Year 10, ACSSU189).*
Our place in space: The earth and life on earth are part of an immense system called the universe.	Rotation of Earth and night/day, spatial relationships between sun, Earth and moon. Planets of our solar system and their characteristics. Space exploration and new developments**	Observable changes occur in the sky and landscape (Year 1, ACSSU019). Earth's rotation on its axis causes regular changes, including night and day (Year 3, ACSSU048). Earth is part of a system of planets orbiting around a star (the sun) (Year 5, ACSSU078).

* Examples that can be found in the Australian Curriculum: Science, Years 7–10.
** Examples that cannot be explicitly found in the Australian Curriculum: Science, Foundation – Year 6.

Major scientific concept areas	Examples	Australian Curriculum: Science links
Energy and force		
Energy and us: Energy is vital to our existence and our quality of life as individuals and as a society.	Uses of energy, patterns of energy use and variations with time of day and season.	Energy from a variety of sources can be used to generate electricity (Year 6, ACSSU219).
	Energy sources, renewable and non-renewable.*	Some of earth's resources are renewable, but others are non-renewable (Year 7, ACSSU116).*
Transferring energy: Interaction and change involve energy transfers; control of energy transfer enables particular changes to be achieved.	Sources, transfers, carriers and receivers of energy, energy and change.	Light and sound are produced by a range of sources and can be sensed (Year 1, ACSSU020).
	Types of energy, energy of motion – toys and other simple machines – light, sound.	Heat can be produced in many ways and can move from one object to another (Year 3, ACSSU049).
		Light from a source forms shadows and can be absorbed, reflected and refracted (Year 5, ACSSU080).
		Electrical circuits provide a means of transferring and transforming electricity (Year 6, ACSSU097).
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.	Forces as pushes and pulls, magnetic attraction and repulsion.	A push or a pull affects how an object moves or changes shape (Year 2, ACSSU033).
		Heat can be produced in many ways and can move from one object to another (Year 3, ACSSU049).
		Forces can be exerted by one object on another through direct contact or from a distance (Year 4, ACSSU076).

* Examples that can be found in the Australian Curriculum: Science, Years 7–10.

Major scientific concept areas	Examples	Australian Curriculum: Science links
Living things		
Living together: Organisms in a particular environment are interdependent.	Living vs. non-living. Plant vs. animal and major groups. Dependence on the environment: Survival needs – food, space and shelter. Interactions between organisms and interdependence (e.g. simple food chains).	Living things have basic needs, including food and water (F, ACSSU002). Living things live in different places where their needs are met (Year 1, ACSSU211). Living things grow, change and have offspring similar to themselves (Year 2, ACSSU030). Living things can be grouped on the basis of observable features and can be distinguished from non-living things (Year 3, ACSSU044). Living things, including plants and animals, depend on each other and the environment to survive (Year 4, ACSSU073).
Structure and function: Living things can be understood in terms of functional units and systems.	Major structures and systems and their functions. Healthy lifestyle, diet and exercise.**	Living things have a variety of external features (Year 1, ACSSU017). Living things have structural features and adaptations that help them to survive in their environment (Year 5, ACSSU043).
Biodiversity, change and continuity: Life on Earth has a history of change and disruption, yet continues generation to generation.	Change over lifetime, reproduction and lifecycles. Adaptation to physical environment.	Living things have life cycles (Year 4, ACSSU072). Living things have structural features and adaptations that help them to survive in their environment (Year 5, ACSSU043). The growth and survival of living things are affected by the physical conditions of their environment (Year 6, ACSSU094).

** Examples that cannot be explicitly found in the Australian Curriculum: Science, Foundation – Year 6.

Major scientific concept areas	Examples	Australian Curriculum: Science links
Matter		
Materials and their uses: The properties of materials determine their uses; properties can be modified.	Materials have different properties and uses. Processing materials to make useful things produces waste, use of alternative materials to better care for the environment. Waste reduction – recycling. Nanotechnology.**	Objects are made of materials that have observable properties (Year F, ACSSU003). Everyday materials can be physically changed in a variety of ways (Year 1, ACSSU018). Different materials can be combined, including by mixing, for a particular purpose (Year 2, ACSSU031). Natural and processed materials have a range of physical properties; These properties can influence their use (Year 4, ACSSU074).
Structure and properties: The substructure of materials determines their behaviour and properties.	The properties of materials can be explained in terms of their visible substructure, such as fibres.*	The properties of the different states of matter can be explained in terms of the motion and arrangement of particles (Year 8, ACSSU151).* Differences between elements, compounds and mixtures can be described at a particle level (Year 8, ACSSU152).*
Reactions and change: Patterns of interaction of materials enable us to understand and control those interactions.	Materials can change their state and properties. Solids, liquids and gases.	A change of state between solid and liquid can be caused by adding or removing heat (Year 3, ACSSU046). Solids, liquids and gases have different observable properties and behave in different ways (Year 5, ACSSU077). Changes to materials can be reversible, such as melting, freezing, evaporating; or irreversible, such as burning and rusting (Year 6, ACSSU095).

* Examples that can be found in the Australian Curriculum: Science in Years 7–10.
** Examples that cannot be explicitly found in the Australian Curriculum: Science K–6.

Appendix 2

Sampling

Appendix 2. Sampling

Overview

The target population for the NAP sample assessment – science literacy (NAP–SL) consisted of all students enrolled in Year 6 in Australian schools in 2015.

The sample design for NAP–SL was a two-stage stratified cluster sample. Stratification involves ordering and grouping schools according to state, sector, size and school location. This helps ensure adequate coverage of all desired school types in the sample.

- Stage 1 consisted of selecting schools that had Year 6 students.
- Stage 2 involved the random selection of Year 6 students within each school selected in Stage 1.

The sample was designed with the aim of achieving a final sample size of approximately 12,000 students who enrolled at approximately 600 schools across Australia.

Ideally, the final sample would result in estimated mean scores for all jurisdictions that were of similar precision. However, it was recognised that reduced sample sizes would be needed for the smaller jurisdictions (that is, Australian Capital Territory, Northern Territory and Tasmania) because most schools in the smaller jurisdictions would need to participate to form a large enough sample. Additionally, as there are a number of national and international assessment projects implemented in Australia, many schools from the smaller jurisdictions would need to participate in multiple assessment projects. Consequently, there would be too much administrative burden on the schools if a larger sample size was used in the smaller jurisdictions.

Target population

The operational definition of the target population was a sampling frame, which consisted of a list of all Australian schools and their 2013 Year 6 enrolment sizes as supplied by Australian Curriculum, Assessment and Reporting Authority (ACARA).

Table A2.1 shows the 2015 estimate of the number of educational institutions and students in the sampling frame for each jurisdiction, as provided by ACARA.

Table A2.1

Estimated 2015 Year 6 enrolment figures as provided by ACARA

State/territory	Institutions	Students	Percentage of students
NSW	2,368	87,087	32.0
Vic.	1,816	65,396	24.0
Qld	1,364	57,378	21.1
WA	858	29,504	10.8
SA	600	18,888	6.9
Tas.	214	6,239	2.3
ACT	99	4,290	1.6
NT	158	3,289	1.2
AUST.	7,477	272,071	100.0

Note: Percentages may not add to 100 per cent due to rounding.

Table A2.2 shows the proportions of large, moderately small and very small schools within each jurisdiction. Schools with Year 6 enrolment sizes larger than or equal to the target cluster size (25) were classified as large schools. Those with enrolment sizes smaller than the target cluster size (TCS) but larger than 12 (TCS/2) were classified as moderately small schools. Schools with an enrolment of 12 (TCS/2) or less were classified as very small.

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 smaller than the desired target sample size.

Table A2.2

Proportion of schools by school size and jurisdiction

State or territory	School size	Number of schools	Percentage of schools	Number of students	Percentage of students
NSW	Large	1,388	58.6	76,475	87.8
	Moderately small	364	15.4	6,916	7.9
	Very small	616	26.0	3,696	4.2
	Total	**2,368**	**100.0**	**87,087**	**100.0**
Vic.	Large	1,073	59.1	56,778	86.8
	Moderately small	320	17.6	6,080	9.3
	Very small	423	23.3	2,538	3.9
	Total	**1,816**	**100.0**	**65,396**	**100.0**
Qld	Large	789	57.8	51,684	90.1
	Moderately small	187	13.7	3,366	5.9
	Very small	388	28.4	2,328	4.1
	Total	**1,364**	**100.0**	**57,378**	**100.0**

WA	Large	512	59.7	25,820	87.5
	Moderately small	134	15.6	2,412	8.2
	Very small	212	24.7	1,272	4.3
	Total	**858**	**100.0**	**29,504**	**100.0**
SA	Large	313	52.2	15,489	82.0
	Moderately small	129	21.5	2451	13.0
	Very small	158	26.3	948	5.0
	Total	**600**	**100.0**	**18,888**	**100.0**
Tas.	Large	115	53.7	5,051	81.0
	Moderately small	36	16.8	684	11.0
	Very small	63	29.4	504	8.1
	Total	**214**	**100.0**	**6,239**	**100.0**
ACT	Large	77	77.8	3,968	92.5
	Moderately small	14	14.1	266	6.2
	Very small	8	8.1	56	1.3
	Total	**99**	**100.0**	**4,290**	**100.0**
NT	Large	56	35.4	2,391	72.7
	Moderately small	22	13.9	418	12.7
	Very small	80	50.6	480	14.6
	Total	**158**	**100.0**	**3,289**	**100.0**

Student non-participation

In large-scale assessments of this kind, it is important to document reasons for non-participation so that interpretations of the main findings from the study can be appropriately made within the contexts of the assessment. Table A2.3 details the exemption and refusal categories for non-participating students.

Table A2.3

NAP–SL exemption and refusal codes

Category	Explanation
Absent	Student is absent from the test.
Functional disability	Student has a moderate to severe permanent physical disability such that he/she cannot perform in the testing situation. Functionally disabled students who can respond to the assessment should be included.
Intellectual disability	Student has a mental or emotional disability and is cognitively delayed so that he/she cannot perform in the testing situation. This includes students who are emotionally or mentally unable to follow even the general instructions of the assessment. Students should NOT be excluded from participating solely because of poor academic performance or disciplinary problems.

Limited test language proficiency	The student is unable to read or speak the language of the assessment (i.e. English) 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 language of the assessment may be excluded.
Student or parent refusal	Parent/caregiver requested that student not participate OR student refused to participate.

The numbers of non-participating students are provided in table A2.4, broken down by jurisdiction and reason for non-participation (where given).

Table A2.4

Student non-participation by jurisdiction

State/ territory	Non-inclusion code					Total
	Absent	Functional disability	Intellectual disability	Limited language proficiency	Student or parent refusal	
NSW	186	0	12	1	10	209
Vic.	154	1	8	0	10	173
Qld	176	0	12	0	7	195
WA	162	1	11	1	5	180
SA	161	0	12	2	35	210
Tas.	106	4	18	2	4	134
ACT	95	0	11	0	17	123
NT	117	2	2	5	1	127
AUST.	**1,157**	**8**	**86**	**11**	**89**	**1,351**

Stratification

The sampling frame was partitioned into 24 separate school lists with each list being a unique combination of state and territory and school type (government, Catholic and independent). This was done to ensure that an adequate number of students were sampled from each school type in each jurisdiction.

Within each of the separate strata, schools were ordered (implicitly stratified) firstly according to their NAPLAN data, then by geographic location and lastly according to the school measure of Year 6 enrolment size.

For most schools, the measure of size for a school was set to the 2013 Year 6 enrolment size of the school. A school's measure of size was adjusted if the school had a small or, alternatively, a very large number of Year 6 students. Sampling methods for both these school types are described in more detail in the 2015 NAP–SL Technical Report.

Replacement schools

Replacement schools were included in the sample to help overcome problems in relation to school non-participation. For example, if the non-participation rate is high, then the target sample sizes will not be achieved. Further, if non-participating schools tend to be lower performing schools, then a bias in the estimated achievement levels will likely occur. More detail can be found in the 2015 NAP–SL Technical Report.

Student selection

Each selected school provided Educational Assessment Australia (EAA) a list containing the details of all eligible Year 6 students. Twenty-six students from the student list were then randomly selected for inclusion in NAP–SL for each school. For schools with 26 students or fewer, all students were included in the study. The random selection of students within schools was a new approach for the 2015 cycle. In previous cycles, a Year 6 class was randomly selected where all Year 6 students in the selected class were included in the NAP–SL sample.

2015 NAP–SL sample results

Table A2.5 provides a breakdown of the sample at the school level according to jurisdiction.

Table A2.5

School participation rates by jurisdiction

State/territory	Number of schools sampled	Number of schools that participated	School participation (per cent)
NSW	93	91	97.8
Qld	91	89	97.8
Vic.	92	91	98.9
WA	92	88	95.7
SA	94	86	91.5
Tas.	62	60	96.8
ACT	55	55	100.0
NT	49	39	79.6
AUST.	**628**	**599**	**95.4**

Table A2.6 provides a breakdown of the sample according to jurisdiction. The target sample is the number of selected Year 6 students *enrolled at the time of testing* in the sampled schools. The achieved sample is the number of selected Year 6 students who participated (attempted the test).

Table A2.6

2015 NAP–SL target and achieved sample sizes by jurisdiction

State/territory	Number of selected students enrolled at the time of testing		Number of selected students who participated in the test	
	Students	Percentage of the sample	Students	Percentage of the sample
NSW	2185	15.1	1911	15.4
Vic.	2162	14.9	1930	15.6
Qld	2177	15.0	1833	14.8
WA	2126	14.7	1878	15.1
SA	2178	15.0	1790	14.4
Tas.	1366	9.4	1198	9.7
ACT	1366	9.4	1221	9.8
NT	920	6.4	649	5.2
AUST.	**14480**	**100.0**	**12410**	**100.0**

Note: Percentages may not add to 100 per cent due to rounding.

Table A2.7 provides a breakdown of the achieved sample in comparison with the number of Year 6 students in each jurisdiction.

Table A2.7

Achieved sample by student participation

State/ territory	Student population	Number of selected students	Number of students who participated	Within school exclusions	Within school exclusions (per cent)	Within-school student participation (per cent)
NSW	87,087	2,185	1,911	23	1.1	87.5
Vic.	65,396	2,162	1,930	19	0.9	89.3
Qld	57,378	2,177	1,833	19	0.9	84.2
WA	29,504	2,126	1,878	18	0.8	88.3
SA	18,888	2,178	1,790	49	2.2	82.2
Tas.	6,239	1,366	1,198	28	2.0	87.7
ACT	4,290	1,366	1,221	28	2.0	89.4
NT	3,289	920	649	10	1.1	70.5
AUST.	**27,2071**	**14,480**	**12,410**	**194**	**1.3**	**85.7**

Sample characteristics

Table A2.8 provides a breakdown of the achieved sample across states and territories according to gender, Indigenous status, students' language background and school geographic location.

Table A2.8

Percentage distribution of Year 6 sample characteristics by jurisdiction

	State/territory (per cent)								AUST. (per cent)
	NSW	Vic.	Qld	WA	SA	Tas.	ACT	NT	
Student gender									
Female	47.1	47.6	50.7	52.2	47.0	50.5	48.4	50.7	49.1
Male	52.9	52.4	49.3	47.8	52.3	49.5	51.6	49.3	50.8
Missing	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.1
Indigenous status									
Indigenous	5.2	1.7	8.4	3.7	3.4	7.3	2.6	29.1	5.9
Non-Indigenous	93.2	98.2	91.2	93.3	94.9	87.9	90.5	70.7	92.0
Missing	1.5	0.1	0.4	2.9	1.7	4.8	6.9	0.2	2.2
Language background									
English speaking background	71.0	74.6	86.6	61.8	80.4	87.8	64.3	43.0	73.3
Language background other than English	29.0	24.2	13.0	36.1	18.7	12.2	35.6	54.1	25.8
Missing	0.1	1.1	0.3	2.1	0.9	0.0	0.1	2.9	0.8
Geographic location									
Metropolitan areas	72.7	79.2	71.5	70.3	73.9	44.7	100.0	0.0	69.5
Provincial areas	27.3	20.8	24.8	23.8	23.5	54.8	0.0	59.8	26.5
Remote and very remote areas	0.1	0.0	3.8	6.0	2.7	0.5	0.0	40.2	4.0
Number of students	**1,911**	**1,930**	**1,833**	**1,878**	**1,790**	**1,198**	**1,221**	**649**	**12,410**

Notes: Percentages may not add to 100 per cent due to rounding. Percentages differ from values in the main report due to weightings.

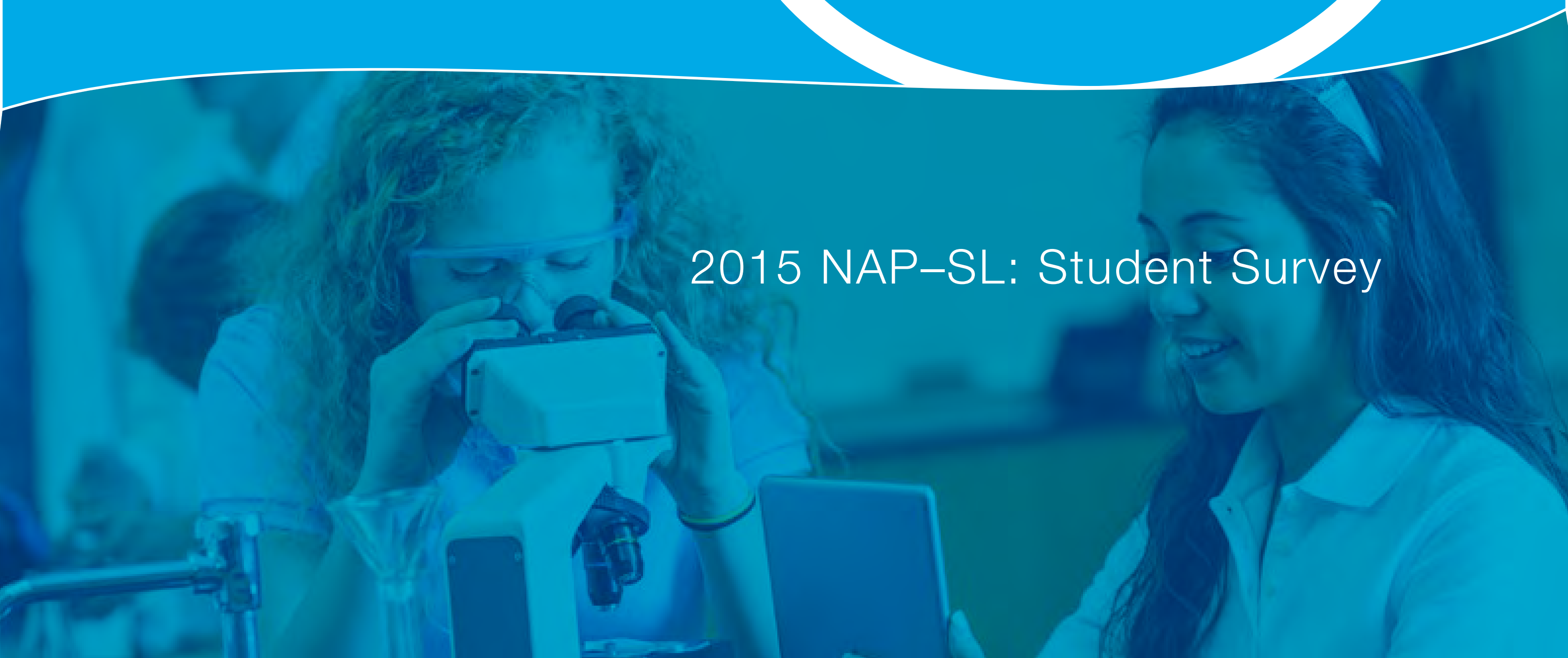
Table A2.9 provides a breakdown of the number of students in the achieved sample by Indigenous status across the three geographic location categories.

Table A2.9

Achieved sample size by Indigenous status and geographic location

Geographic location	Number of students by Indigenous status			Total
	Indigenous	Non-Indigenous	Missing	
Metropolitan areas	255	8,141	229	8,625
Provincial areas	294	2,959	35	3288
Remote and very remote areas	178	316	3	497
AUST.	**727**	**11,416**	**267**	**12,410**

Appendix 3



2015 NAP–SL: Student Survey

Appendix 3. 2015 NAP–SL: Student Survey

The survey was administered after the inquiry task on the same online platform as the rest of the assessment. Students were given the following instructions before starting the survey:

> *In the final section of the test, you will be asked to complete a survey.*
>
> *This survey is to find out your opinions and ideas about science.*
>
> *Please read each sentence carefully and answer as accurately as you can. You may ask for help to read the questions or if you are not sure how to show your answer.*
>
> *Remember: there are no right or wrong answers when answering the survey. Your answers should be the ones you think are best for you. Are there any questions?*

Each group of items appeared together on the screen one group at a time.

Table A3.1

Student survey questions and response options

Item	Text	Response options			
Group 1: How much do you agree with the statements below?					
1	I would like to learn more science at school.	Strongly agree	Agree	Disagree	Strongly disagree
2	I think it would be interesting to be a scientist.	Strongly agree	Agree	Disagree	Strongly disagree
3	I enjoy doing science.	Strongly agree	Agree	Disagree	Strongly disagree
4	I enjoy learning new things in science.	Strongly agree	Agree	Disagree	Strongly disagree
Group 2: How much do you agree with the statements below?					
5	I learn science topics quickly.	Strongly agree	Agree	Disagree	Strongly disagree
6	I can usually give good answers to science questions.	Strongly agree	Agree	Disagree	Strongly disagree
7	I can understand new ideas about science easily.	Strongly agree	Agree	Disagree	Strongly disagree

Item	Text	Response options			
Group 3: How much do you agree with the statements below?					
8	Science is an everyday part of my life.	Strongly dgree	Agree	Disagree	Strongly disagree
9	Science is important for lots of jobs.	Strongly agree	Agree	Disagree	Strongly disagree
10	Science is important: it changes how we live.	Strongly agree	Agree	Disagree	Strongly disagree
11	Scientific information helps people make good decisions.	Strongly agree	Agree	Disagree	Strongly disagree
Group 4: How much do you agree with the statements below?					
12	Science is about remembering facts.	Strongly agree	Agree	Disagree	Strongly disagree
13	Science is about doing experiments.	Strongly agree	Agree	Disagree	Strongly disagree
14	Science is finding out about how things work.	Strongly agree	Agree	Disagree	Strongly disagree
15	Science is about solving problems.	Strongly agree	Agree	Disagree	Strongly disagree
16	Science is quite easy for most people to understand.	Strongly agree	Agree	Disagree	Strongly disagree
Group 5: How often do you do these things outside of school?					
17	I view TV programs, DVDs or websites about science topics at home.	Frequently (more than 2 times a week)	Often (1 or 2 times a week)	Sometimes (less than once a week)	Never
18	I read books, newspapers or magazine articles about science topics.	Frequently (more than 2 times a week)	Often (1 or 2 times a week)	Sometimes (less than once a week)	Never
19	I talk about science ideas with my friends and family.	Frequently (more than 2 times a week)	Often (1 or 2 times a week)	Sometimes (less than once a week)	Never
Group 6: How often do you do these things at school?					
20	I view TV programs, DVDs or websites about science topics at school.	Frequently (More than 2 times a week)	Often (1 or 2 times a week)	Sometimes (less than once a week)	Never
21	We read books, newspapers or magazine articles about science topics at school.	Frequently (More than 2 times a week)	Often (1 or 2 times a week)	Sometimes (less than once a week)	Never
Group 7: How often do you do these things at school?					
22	During science lessons I get to plan and carry out my own investigations.	Always	Mostly	Sometimes	Never
23	When our class investigates things in science, we work in groups to carry out the investigation.	Always	Mostly	Sometimes	Never

Item	Text	Response options			
24	I use a computer for research or to present my science ideas and findings.	Always	Mostly	Sometimes	Never
25	Our class has in-depth discussions about science ideas.	Always	Mostly	Sometimes	Never
Group 8: Which of these science topics have you studied at school?					
26	Earth and space (Earth and space sciences) - for example, weather, soil, rocks, gravity, using Earth's resources, the planets, the sun and the moon.	Yes	No	I don't know	
27	Energy and force (physical sciences) – for example, how toys and other machines work, electricity, heat, light, sound, magnets.	Yes	No	I don't know	
28	Living things (biological sciences) – for example, living and non living things, how animals and plants survive in their environment, life cycles, interdependence.	Yes	No	I don't know	
29	Matter (chemical sciences) – for example, the different properties of materials such as plastics and metals, the different uses of materials, changes to materials (solids, liquids and gases).	Yes	No	I don't know	
Group 9: How often do you have science lessons at school?					
30	How often do you have science lessons at school?	More than once a week	Once a week	Less than once a week	Hardly ever
Group 10: Do you agree with the statements below?					
31	My classroom teacher teaches science to our class.	Yes	No		
32	I think my teacher enjoys teaching science.	Yes	No		
33	Our teacher invites visitors to school to talk to us about science topics.	Yes	No		
34	Our class goes on excursions related to the science topics we are learning about.	Yes	No		

Item	Text	Response options			
Group 11: How much do you agree with the statements below? Science is about…					
35	making observations about the world.	Strongly agree	Agree	Disagree	Strongly disagree
36	asking questions about objects and events.	Strongly agree	Agree	Disagree	Strongly disagree
37	making predictions and testing them.	Strongly agree	Agree	Disagree	Strongly disagree
38	describing patterns and relationships.	Strongly agree	Agree	Disagree	Strongly disagree
39	using evidence to develop explanations.	Strongly agree	Agree	Disagree	Strongly disagree
Group 12: How much do you agree with the statements below?					
40	People from many different countries have made important contributions to science.	Strongly agree	Agree	Disagree	Strongly disagree
41	Women and men are involved in science.	Strongly agree	Agree	Disagree	Strongly disagree
42	People from all cultural backgrounds in Australia are involved in science.	Strongly agree	Agree	Disagree	Strongly disagree
43	People of all ages are involved in science.	Strongly agree	Agree	Disagree	Strongly disagree

Appendix 4

Comparison of 2012 and 2015 Survey Results

Appendix 4. Comparison of 2012 and 2015 Survey Results

The following tables provide a comparison of survey questions that appeared in both the 2012 and the 2015 student survey. The percentages have been weighted to account for sampling affects across states and territories and so differ slightly from the percentages shown in chapter 8.

The significance of the difference has been calculated taking into account a Bonferroni adjustment.

Table A4.1

Comparison of student responses to common questions in the 2009, 2012 and 2015 student surveys

Group and number	Survey question		2012 Per cent	SE	2015 Per cent	SE	2012 to 2015 difference (percentage points)	Significant?	Change from 2012
G01	**How much do you agree with the statements below?**								
1	I would like to learn more science at school.	Strongly disagree	3.73	0.30	2.18	0.20	-1.55	YES	decrease
		Disagree	15.08	0.56	10.69	0.45	-4.39	YES	decrease
		Agree	57.10	0.71	58.09	0.74	0.99	NO	
		Strongly agree	24.08	0.67	28.20	0.77	4.12	YES	increase
2	I think it would be interesting to be a scientist.	Strongly disagree	12.33	0.51	7.27	0.37	-5.07	YES	decrease
		Disagree	28.93	0.60	23.46	0.56	-5.47	YES	decrease
		Agree	44.67	0.75	51.17	0.63	6.50	YES	increase
		Strongly agree	14.06	0.43	18.10	0.49	4.04	YES	increase
3	I enjoy doing science.	Strongly disagree	3.32	0.27	2.57	0.21	-0.75	NO	
		Disagree	11.34	0.45	10.53	0.44	-0.81	NO	
		Agree	56.68	0.66	56.38	0.72	-0.30	NO	
		Strongly agree	28.67	0.70	30.53	0.80	1.86	NO	
4	I enjoy learning new things in science.	Strongly disagree	2.41	0.25	1.90	0.18	-0.51	NO	
		Disagree	8.48	0.40	5.92	0.33	-2.55	YES	decrease
		Agree	53.49	0.67	49.51	0.78	-3.98	YES	decrease
		Strongly agree	35.63	0.71	42.67	0.83	7.04	YES	increase

Group and number	Survey question		2012 Per cent	SE	2015 Per cent	SE	2012 to 2015 difference (percentage points)	Significant?	Change from 2012
G02	How much do you agree with the statements below?								
5	I learn science topics quickly.	Strongly disagree	6.26	0.35	3.76	0.25	-2.50	YES	decrease
		Disagree	34.88	0.64	27.25	0.63	-7.63	YES	decrease
		Agree	48.39	0.70	54.95	0.71	6.56	YES	increase
		Strongly agree	10.48	0.37	14.04	0.45	3.56	YES	increase
6	I can usually give good answers to science questions.	Strongly disagree	4.81	0.29	3.02	0.23	-1.80	YES	decrease
		Disagree	29.82	0.64	23.13	0.56	-6.69	YES	decrease
		Agree	55.11	0.77	60.35	0.68	5.24	YES	increase
		Strongly agree	10.26	0.38	13.51	0.45	3.25	YES	increase
7	I can understand new ideas about science easily.	Strongly disagree	4.79	0.32	2.66	0.21	-2.13	YES	decrease
		Disagree	28.47	0.64	22.15	0.54	-6.32	YES	decrease
		Agree	51.71	0.72	55.43	0.65	3.72	YES	increase
		Strongly agree	15.04	0.46	19.77	0.56	4.73	YES	increase
G03	How much do you agree with the statements below?								
8	Science is an everyday part of my life.	Strongly disagree	19.22	0.65	12.25	0.41	-6.97	YES	decrease
		Disagree	43.62	0.71	41.43	0.61	-2.19	NO	
		Agree	24.66	0.58	32.23	0.61	7.57	YES	increase
		Strongly agree	12.50	0.60	14.09	0.52	1.60	NO	
9	Science is important for lots of jobs.	Strongly disagree	2.96	0.26	1.86	0.16	-1.11	YES	decrease
		Disagree	17.40	0.58	14.62	0.45	-2.78	YES	decrease
		Agree	52.72	0.57	52.18	0.61	-0.54	NO	
		Strongly agree	26.92	0.64	31.35	0.63	4.43	YES	increase
10	Science is important: it changes how we live.	Strongly disagree	2.87	0.25	1.50	0.16	-1.37	YES	decrease
		Disagree	10.89	0.43	7.47	0.34	-3.42	YES	decrease
		Agree	45.20	0.70	41.67	0.71	-3.53	YES	decrease
		Strongly agree	41.04	0.78	49.36	0.81	8.31	YES	increase
11	Scientific information helps people make good decisions.	Strongly disagree	3.55	0.30	2.29	0.19	-1.26	YES	decrease
		Disagree	18.01	0.52	13.58	0.44	-4.42	YES	decrease
		Agree	50.58	0.59	50.81	0.70	0.23	NO	
		Strongly agree	27.87	0.61	33.32	0.64	5.45	YES	increase

Group and number	Survey question		2012 Per cent	SE	2015 Per cent	SE	2012 to 2015 difference (percentage points)	Significant?	Change from 2012
G04	**How much do you agree with the statements below?**								
12	Science is about remembering facts.	Strongly disagree	6.75	0.32	5.30	0.30	-1.45	YES	decrease
		Disagree	36.27	0.70	36.19	0.67	-0.08	NO	
		Agree	45.83	0.62	46.86	0.65	1.03	NO	
		Strongly agree	11.15	0.45	11.65	0.45	0.50	NO	
13	Science is about doing experiments.	Strongly disagree	2.19	0.21	2.06	0.19	-0.14	NO	
		Disagree	15.00	0.46	18.94	0.52	3.94	YES	increase
		Agree	49.88	0.60	49.98	0.63	0.10	NO	
		Strongly agree	32.92	0.62	29.02	0.60	-3.90	YES	decrease
14	Science is finding out about how things work.	Strongly disagree	1.18	0.20	0.59	0.09	-0.59	YES	decrease
		Disagree	4.78	0.26	3.95	0.24	-0.83	NO	
		Agree	52.54	0.65	50.24	0.67	-2.30	NO	
		Strongly agree	41.50	0.62	45.22	0.70	3.72	YES	increase
15	Science is about solving problems.	Strongly disagree	4.21	0.30	3.23	0.23	-0.97	YES	decrease
		Disagree	22.31	0.55	16.29	0.48	-6.02	YES	decrease
		Agree	50.59	0.62	53.02	0.65	2.43	YES	increase
		Strongly agree	22.90	0.55	27.46	0.67	4.56	YES	increase
16	Science is quite easy for most people to understand.	Strongly disagree	10.27	0.41	8.98	0.39	-1.29	NO	
		Disagree	43.08	0.67	46.29	0.71	3.21	YES	increase
		Agree	37.50	0.62	36.46	0.65	-1.04	NO	
		Strongly agree	9.14	0.37	8.26	0.40	-0.89	NO	
G05	**How often do you do these things outside of school?**								
17	I view TV programs, DVDs or websites about science topics at home.	Never	30.60	0.71	28.05	0.63	-2.54	YES	decrease
		Sometimes	41.16	0.63	40.80	0.61	-0.36	NO	
		Often	18.89	0.55	19.63	0.54	0.74	NO	
		Frequently	9.35	0.39	11.51	0.42	2.16	YES	increase
18	I read books, newspapers or magazine articles about science topics.	Never	40.95	0.82	36.02	0.69	-4.93	YES	decrease
		Sometimes	36.77	0.68	36.56	0.68	-0.21	NO	
		Often	15.82	0.47	18.66	0.51	2.84	YES	increase
		Frequently	6.46	0.32	8.76	0.37	2.30	YES	increase
19	I talk about science ideas with my friends and family.	Never	43.60	0.83	35.40	0.68	-8.20	YES	decrease
		Sometimes	33.84	0.66	33.01	0.56	-0.83	NO	
		Often	14.99	0.43	20.37	0.50	5.38	YES	increase
		Frequently	7.57	0.32	11.22	0.45	3.65	YES	increase

Group and number	Survey question		2012 Per cent	SE	2015 Per cent	SE	2012 to 2015 difference (percentage points)	Significant?	Change from 2012
G06	**How often do you do these things at school?**								
20	I view TV programs, DVDs or websites about science topics at school.	Never	19.94	0.79	15.79	0.51	-4.15	YES	decrease
		Sometimes	47.80	0.85	41.89	0.75	-5.91	YES	decrease
		Often	26.23	0.83	33.19	0.75	6.96	YES	increase
		Frequently	6.03	0.48	9.14	0.42	3.10	YES	increase
21	We read books, newspapers or magazine articles about science topics at school.	Never	23.50	0.83	24.03	0.61	0.54	NO	
		Sometimes	45.71	0.76	42.02	0.64	-3.69	YES	decrease
		Often	23.96	0.75	26.22	0.63	2.26	NO	
		Frequently	6.83	0.47	7.73	0.38	0.90	NO	
G07	**How often do you do these things at school?**								
22	During science lessons I get to plan and carry out my own investigations.	Never	23.33	0.90	13.71	0.55	-9.62	YES	decrease
		Sometimes	49.66	0.74	48.34	0.67	-1.33	NO	
		Mostly	21.93	0.71	30.09	0.61	8.16	YES	increase
		Always	5.07	0.42	7.86	0.34	2.78	YES	increase
23	When our class investigates things in science, we work in groups to carry out the investigation.	Never	7.45	0.60	3.95	0.31	-3.50	YES	decrease
		Sometimes	32.15	0.85	25.08	0.66	-7.07	YES	decrease
		Mostly	45.33	0.87	48.82	0.76	3.49	YES	increase
		Always	15.07	0.60	22.15	0.87	7.08	YES	increase
24	I use a computer for research or to present my science ideas and findings.	Never	16.99	0.78	11.86	0.70	-5.13	YES	decrease
		Sometimes	39.70	0.78	35.74	0.74	-3.96	YES	decrease
		Mostly	30.90	0.74	35.70	0.82	4.80	YES	increase
		Always	12.41	0.55	16.71	0.66	4.29	YES	increase
25	Our class has in-depth discussions about science ideas.	Never	15.49	0.82	10.23	0.53	-5.27	YES	decrease
		Sometimes	44.64	0.79	38.67	0.74	-5.97	YES	decrease
		Mostly	26.75	0.73	32.27	0.67	5.52	YES	increase
		Always	13.12	0.67	18.83	0.61	5.72	YES	increase
G08	**Which of these science topics have you studied at school?**								
26	Earth and Space (Earth and space sciences)	Yes	76.64	0.87	76.58	0.74	-0.06	NO	
		No	9.34	0.52	8.40	0.43	-0.94	NO	
		I don't know	14.02	0.55	15.02	0.52	1.00	NO	
27	Energy and Force (Physical sciences)	Yes	74.87	0.96	70.36	0.91	-4.51	YES	Decrease
		No	12.09	0.59	14.00	0.62	1.91	NO	
		I don't know	13.04	0.59	15.64	0.56	2.60	YES	Increase
28	Living Things (Biological sciences)	Yes	72.19	0.89	68.51	0.85	-3.68	YES	Decrease
		No	12.34	0.56	14.84	0.60	2.50	YES	Increase
		I don't know	15.47	0.56	16.65	0.58	1.18	NO	

Group and number	Survey question		2012 Per cent	SE	2015 Per cent	SE	2012 to 2015 difference (percentage points)	Significant?	Change from 2012
29	Matter (Chemical sciences)	Yes	63.44	0.98	67.43	0.97	3.99	YES	Increase
		No	15.43	0.57	14.67	0.67	-0.76	NO	
		I don't know	21.14	0.69	17.90	0.59	-3.23	YES	Decrease
G09	**How often do you have science lessons at school?**								
30	How often do you have science lessons at school?	hardly ever	19.81	1.23	14.99	0.91	-4.82	YES	Decrease
		less than once a week	18.03	0.93	17.13	0.87	-0.91	NO	
		once a week	39.71	1.50	45.76	1.51	6.05	YES	Increase
		more than once a week	22.44	1.23	22.12	1.22	-0.32	NO	
G10	**Do you agree with the statements below?**								
31	My classroom teacher teaches science to our class.	No	29.30	1.55	26.69	1.28	-2.62	NO	
		Yes	70.70	1.55	73.31	1.28	2.62	NO	
32	I think my teacher enjoys teaching science.	No	17.58	0.93	14.92	0.66	-2.66	YES	decrease
		Yes	82.42	0.93	85.08	0.66	2.66	YES	increase
33	Our teacher invites visitors to school to talk to us about science topics.	No	69.11	1.37	69.40	0.99	0.29	NO	
		Yes	30.89	1.37	30.60	0.99	-0.29	NO	
34	Our class goes on excursions related to the science topics we are learning about.	No	59.64	1.38	61.21	1.07	1.57	NO	
		Yes	40.36	1.38	38.79	1.07	-1.57	NO	

Notes: Percentages may not add to 100 per cent due to rounding. In Groups 5 and 6 (items 17–21) 'Frequently' means more than 2 times a week. 'Often' means 1 or 2 times a week. 'Sometimes' means less than once a week. Group 8 (items 26–29) were amended in 2015 to include the equivalent strand names from the Australian Curriculum: Science.

Appendix 5

Pilot Report Summary

Appendix 5. Pilot Report Summary

Educational Assessment Australia (EAA) conducted pilots of four 2015 NAP–SL inquiry tasks in November 2014.

Each inquiry task provided students with an opportunity to work through an investigation within a formal assessment and test the conventions of science literacy in more depth than is possible in the objective component. Each inquiry task employed written, pictorial and video stimulus.

The pilots had multiple purposes:

- to study student engagement with the tasks
- to study the practicalities of running the tasks online
- to study student interaction with the test platform.

The observations from the pilot were intended to inform the following:

- modifications and edits to items and stimulus in the inquiry tasks
- advice to schools involved in the NAP–SL trials and main study on how to conduct the tests
- advice to ACARA on issues relating to the test delivery platform.

At each pilot, test-developers from EAA observed students attempting an inquiry task. Once the students had completed the task, a group debrief and discussion was held with the students. During this debrief, students were asked questions and given a chance to voice their opinions about the task.

The schools

Pilots were conducted in five schools in the greater Sydney area. Schools included government primary schools, Catholic and independent schools.

While the small sample of schools could not be fully representative of the full diversity of schools in Australia, the schools and students that participated in the pilots were quite varied.

Classroom and technology set ups

Technology requirements for pilot participation were the minimum standards to access the system and a sufficient number of computers for a Year 6 class to participate.

Table A6.1

Technical requirements (pilot)

Hardware	Operating system	Web browser for student access
Laptop/desktop PC (iPads/tablets not supported) Mouse and sound card	Mac OS X MS Windows: XP, Vista, 7 or 8	Chrome 21+ (XP, Vista, 7, 8, OSX)
Settings		
1024 x 768 monitor resolution, or higher. Browser runs JavaScript Bandwidth to internet (for the school): recommend 512 KB bandwidth minimum for up to 20 users		

Each school presented different combinations of technology and accommodation.

Table A6.2

Technology and accommodation

School	Platform used for pilot	Platform used in lessons	Classroom arrangement for pilot
School 1	Chromebooks	Chromebooks	Portable devices in a student's normal classroom
School 2	Mac notebooks	BYODD iPads	Learning centre, flexible use room
School 3	Windows netbooks	Chromebooks	Two adjoining classrooms. Students sat 1 per desk, test style
School 4	Windows desktop PC	Windows desktop PC	IT lab; computers arranged in a horseshoe around three walls with an extra row of computers in the centre of the room
School 5	Windows desktop PC	Desktop PCs and iPads	IT lab as part of a library/resource area; computers arranged in rows facing the 'front' wall of the classroom

Note: BYODD - Bring Your Own Designated Device

All schools used the recommended Chrome browser. No school reported any issues installing Chrome. However, for three schools it was not the default browser for the computers and students had to be directed on how to open it.

School 1 had a class set of Chromebooks in the classroom. Students were already seated in their usual seating arrangement. The Chromebooks were handed out to students in a manner that suggested this was a usual practice. This set up was minimally disruptive to the students' normal lesson routine. Once the pilot sessions had finished, students moved seamlessly back into their normal lesson.

School 2 made use of a purpose-built learning centre. The room was designed to be used for multiple purposes, one of which was for specific IT lessons and IT related activities (for example, a lunchtime 'Minecraft' club). The room had a variety of seats, including some high bar-like tables with stools. The room also came equipped with a class set of Mac notebooks.

School 3 used two adjoining classrooms. This allowed them to spread most students out so that most students could not easily see another student's screen. A small number of students had to sit closer together. This school had recently transitioned to Chromebooks but as Chrome OS was not listed in our technical readiness advice, the school had assumed they could not use them. Instead, students used older 'netbook' style Windows machines. These computers were quite old and had small screens. In a few cases keys were missing from keyboards.

School 4 used a specialist computer room. The room was arranged with students not facing the 'front' wall of the classroom where the whiteboard was positioned. Students were sitting relatively close to each other and could see their neighbour's screen.

School 5 used a specialist computer room that was a part of a larger library/resource area. Computers were arranged in rows so that all students were sitting facing the main whiteboard. This arrangement was the easiest for showing students how to access the system via the interactive whiteboard. However, the seating arrangement meant that some students could see not only their neighbour's screen but also the screens of the students sitting in front of them.

Recommendations:

- The availability of the test for Chromebooks needs to be confirmed.
- While some classroom arrangements may be more suitable than others, schools will have little control of this in the short term.

Headphones

At none of the five schools participating in the pilot did students bring their own headphones for the test. School 2 reported that students commonly have headphones in their bags, but on this occasion they had left their bags in their regular classroom. School 1 and School 4 had a class set of headphones that accompanied the computers. School 5 had a small set of headphones available in some classrooms (approximately 10).

For convenience, headphones were provided by EAA in Schools 2, 3 and 5.

Students did not express any major objections to wearing headphones in any of the schools. Some students did say that, while they did not mind the headphones overall, their ears were a little uncomfortable by the end of the session.

All students appeared to be comfortable with adjusting the audio volume to a level with which they were comfortable.

Recommendations:

- The importance of having access to headphones needs to be made clear to schools conducting 2015 NAP–SL.

Starting a session

The start of each of the pilot sessions differed according to the classroom arrangement in the school.

At School 1, students were already at their usual seat. This meant that very little time was spent on seating students appropriately.

At School 2, the learning centre had a variety of seats, which made it more difficult to quickly seat students appropriately. The more informal setting also made it difficult to spread students out appropriately, particularly when some seating areas were clearly designed for collaborative work.

At School 3, the teacher had set out the computers in appropriate positions at the start of the day. The class teacher allocated students to particular positions, following a process similar to one a teacher might employ for a formal paper test.

At Schools 4 and 5, students quickly sat in appropriate positions next to computers. At School 5, class teachers repositioned some students to aid classroom management.

At all schools logging onto the school network, opening Chrome and logging onto the internet caused some issues. In all schools except School 1, students had to enter login details after they had opened a browser and attempted to access the internet. This login process was to access the jurisdiction's internet gateway. Some students were confused by this login step and attempted to use the login details provided for the NAP–SL pilot.

At School 2, logging in proved to be difficult for some computers. The class teachers present were unfamiliar with the system and found it difficult to help students with this step. Additionally, the process may have been different from that which students were used to when using iPads. It is possible that the problem was with some specific computers, as some students successfully logged on when they tried a different computer. Alternatively, it may have been an issue with the school's own network.

Once logged on and able to access the internet, students typed in the URL for the pilot test. Most students did this successfully but some mistyped the address. Students had few issues logging onto the test but some students were unsure about what to do when the browser prompted them to indicate whether the password should be remembered. School 4 tried to prevent students from having to type in the URL by putting a HTML page on their network with a link to the test. Unfortunately, the URL used was the technical readiness test (TRT) URL rather than the test URL. While this plan could not be used on the day of the pilot, it was a sensible idea.

The issues listed above are fairly normal for a teacher attempting to have a class of students all access the same website at the same time. The troubleshooting issues were largely minor and (aside from School 2) within the capability of the classroom teacher to resolve quickly. However, they did present a classroom management issue for the test as some students had to sit and wait while the teacher dealt with specific issues.

Recommendations:

- Schools should be advised to try a dress rehearsal so that the supervising teacher has some experience with the process of getting students onto the system.
- It is unclear if bandwidth issues played a role in some of the issues School 2 had. Notably, at School 1 other classes had been advised not to use the network during the pilot. We would suggest that schools during the trial and main study follow a similar policy to minimise network traffic during the test.
- The schools saving a local web page with a link to the test URL is a sensible idea that will make it easier for students to get onto the system. Alternatively, the school could bookmark the test in the browser. Unfortunately, it would be hard to give schools general guidance on this.
- The simplest troubleshooting technique at the start of the process is for students encountering difficulties to change to a different computer. Ensuring that there are functioning spares might not always be possible, but if schools can do this, the process will be easier.
- Having two members of staff in the room is advisable but may not be feasible for schools.
- As the test has a pre-set duration, it is not strictly necessary for all students to start at the same time. A staggered start might be more manageable for some schools.

Accessing the test

The test ran with few technical problems in most of the schools. Only in School 2 did students have issues accessing the materials. In all cases, this was with computers that had problems accessing the internet initially. The TRT was run on one of the machines but no issues were found.

At School 5, three students encountered a notification advising ‘Bad Gateway: Please refresh’. The students refreshed the page they were on and it returned them to the test at the item they were on. The students managed this issue themselves; unfortunately, observers did not see this issue directly.

At School 3, one student did not have all images load for one item in Sunscreens.

In a few cases at all schools, some students found themselves temporarily unable to access the next item. In these cases the active item was appearing as normal (the item number appearing on the navigation bar as blue text on a white background) but the next item appeared as white text on a feint background rather than on a dark background. Skipping back a couple of items seemed to ‘un-pause’ the test.

Overall, the test ran smoothly and without issue for most students. Even at School 2, most students completed the test, although a few students had to wait until another student had finished.

Timing

In every school, students took from 10 to 30 minutes to finish the test. This range occurred in all schools and classes, regardless of the overall ability of the class. Three students ran out of time – one student at School 4 and two students at School 5. At School 5, a student who finished at 30 minutes (that is, 0 minutes left) and the student who ran out of time had been identified by the teacher as being 'Science boffins'.

From the test data, it was possible to make estimates of the time students took to complete the tasks. The test data included the time spent by each student on each item. By adding these times together, estimates of the time spent on the whole task could be made.

Students who spent longer on the test were typically students who wrote more on the extended-response items. Students were observed deleting initial responses to extended-text items and then writing a new response. This flexibility with online testing should lead to improved responses overall but may add to a greater disparity between students on the time taken to complete the test.

This table gives the average time spent on each item type.

Table A6.3

Average time by item type

Item type	Average time (seconds)
Interactive order*	45
Multiple choices	48
Multiple choice	66
Interactive gap match	70
Interactive graphic gap match	74
Cloze*	83
Composite	102
Select point	111
Extended text	111
Average for all item types	**90**

* The 'Interactive order' and 'Cloze' item types were earlier forms of item types that were superseded before the NAP–SL trial.

In debrief sessions, many students commented that it was easier to type than write by hand. It is reasonable to assume that the online format allows some students to write more than they would on paper.

Given that students cannot change their responses in the inquiry tasks, it will not be possible for students who finish early to go back and check their responses. Potentially, some students would have to wait up to 25 minutes for their classmates to finish and this may cause classroom management issues.

Recommendations:

- The timer for the inquiry task should be set at 35 minutes.
- Supervising teachers should be advised that some students may finish much earlier.
- Students who have finished the inquiry task should move onto the survey.
- Students who have finished the inquiry task and the survey need to have a silent activity to keep them occupied and/or return to normal lessons. Schools should be advised of these options but individual schools may prefer one approach over another.

Test navigation

Students generally had no issue moving from one item to the next.

The stimulus pane proved to be confusing to some students. For item 1, students needed to be told that they could collapse the pane to see the first question. After that some students did open the stimulus pane but many students did not. A few students in all schools were unaware that the pane could be opened.

Many students maximised the videos to full-screen when watching them. This was particularly interesting as they were never told this was possible. The standard video controls that appear in the system were familiar to students and consequently students had no problems pausing videos or maximising them to full screen. The only downside to this was that the videos had been compressed to a level that had reasonable picture quality for the standard size but not for full screen size.

When asked in the debrief session for one thing that could be changed for the videos, some students commented that they would like them to be ‘full HD’ (high definition) so that they would still look good at full screen.

The warning dialogue box presented to students when they moved onto the next item was effective for informing students that they couldn’t then change their answers. Many students commented on this issue during the debrief session. Many students were not happy that they could not change their answers later but several correctly identified why this test had the restriction.

Recommendations:

- General navigation features seem to be effective and do not need changing.
- The dialogue box warning students that they cannot change their answers was effective. While some students found it a little annoying, it should be retained.
- Higher definition videos would be good but without extensive testing in schools with whole classes it was better to stick with low file size to ensure videos load reliably.
- Text on the first item in a set, reminding students that they can open the stimulus pane, needed to be introduced.

Videos

The use of video in the test was very popular among students. During debrief sessions students often mentioned unprompted that they liked having video. The level of popularity of the videos surprised observers.

Students liked having the stimulus information read and shown to them in a video. They commented that it was easier to understand and follow. Complex stimulus material such as UV light, photosensitive paper, etc., seemed to be understood by students better because of the video.

Students actively engaged with the videos. Students of all abilities were seen watching videos all the way through and sometimes students watched the same video more than once. As mentioned above, many students chose to watch the videos full screen. Some students noted that topics such as earthquakes or black holes would have been interesting as stimulus material.

When asked what they would like to change about the videos, many student commented on the sound. Sound levels were not always consistent and in some cases the audio cut off a little sharply at the end. Students also felt the videos ended too abruptly and felt that they would have liked a clearer indication that a given video had finished.

For tasks that featured video of a student conducting an investigation (all tasks except Pendulums), students were asked about the student portrayed. Most students did not feel the student should be younger. This was the case even for Sunscreen and Mustard Seeds, which featured a 14-year-old student.

One class at School 3 indicated that they would prefer a younger student, but the reasons given did not relate to engagement (for example, that a younger student would use fewer big words - even though in the video the student did not speak).

The use of animated characters was discussed with students. A modified version of some of the Pendulum videos was shown to students in School 2. Generally, this was regarded positively by students although some students felt that no character (real or animated) was needed for Pendulums. One student in School 5 said that they preferred not to have an animated character as it was less real.

In several schools students indicated that they would have liked the videos to be longer.

Recommendations:

- The use of videos in online tests seemed to be well supported based on student reaction.
- One or two of the videos should use an animated character.
- Audio on all videos needed further improvement, including consistent sound levels.
- All videos should start and end with title screens to clearly indicate the beginning and end of each video.

Item types

Students appeared to interact with all item types with confidence. When asked during the debrief session to identify the type of items they preferred, students in all schools identified multiple choice as their preferred item type. However, many students also identified the various kinds of more interactive items as being good. A minority of students in all schools spoke favourably about the extended-text items on the grounds that they could express their own ideas.

For two graph items, the select-point interaction was used to test students' ability to plot points. Students' reaction to these items was interesting. Many students identified these items as being difficult but students also identified them as a style of item that they liked.

Students appeared to engage with the text-entry items. Students appeared to understand the purpose of the composite items and some students identified them as an item type they liked as it allowed them to explain the choice they had made.

Recommendations:

- Student response data will be examined carefully to see if there are other issues.
- There were no suggested changes to item types.

Overall engagement

Students appeared fully engaged with the test as they worked on it. The level of concentration was remarked on by class teachers in several schools. The lowest level of engagement was seen in School 2 among students who had technical issues.

During the debrief session, students were asked about how interesting the test was. Generally, the response was neutral. At School 5 some students indicated that they would have preferred it if the test had been more game-like but another student disagreed saying that they preferred a test to be a test.

The videos and the colourful headphones seemed to contribute to student engagement.

Science content

Students were asked in the debrief session if they thought the test felt like a science test. Responses were quite mixed to this and showed more variation between schools than other questions. It is possible that many students did not have any set ideas of what a science test should be like and may not have sat many tests specifically on science. In primary schools, science may be included as one aspect of a broader unit of work and students may not see ‘science’ as a distinct curriculum subject.

Students were also asked if the investigation shown was like something they might do in class. The response to this was also quite varied.

At School 4, the class did the Mustard Seeds task and coincidentally had recently completed an investigation involving growing wheat seeds. Consequently, the correspondence between the two tasks was immediate and obvious.

At School 2, students were less sure but also found it difficult to describe what kinds of topics they might investigate in class.

Recommendations:

- Content appears appropriate based on student responses.

Appendix 6

Trial Report Summary

Appendix 6. Trial Report Summary

Introduction

This is an extract of the 2015 NAP sample assessment – science literacy (NAP–SL) trial report compiled by Educational Assessment Australia (EAA). It provides a brief summary of the development and trialling of a suite of assessment instruments in preparation for the national sample assessment of Year 6 students conducted in October 2015.

Project background

The equating design

To place the 2015 results on the science literacy scale established following the 2006 NAP–SL assessment, the common-item method with multiple-linking linkage plan was used. Since 2015, NAP–SL was the fifth cycle of this assessment, there was an opportunity to draw link items from four previous test cycles (2003, 2006, 2009 and 2012).

To this effect, 39 items were identified from previous cycles that had strong psychometric properties and which would be amenable to being delivered online.

The items were grouped in eight clusters, C1–C8. Each objective test consisted of two clusters. This allowed eight tests to be constructed. The inquiry task items made up four inquiry task units, each of which was contained in an independent form.

Table A7.1

Structure of test forms by cluster

Trial forms	Cluster 1	Cluster 2	Source of items
Objective form 1	C1	C2	2006, 2009 & 2012
Objective form 2	C2	C3	2012 & 2015
Objective form 3	C3	C4	2015
Objective form 4	C4	C5	2015
Objective form 5	C5	C6	2015

Objective form 6	C6	C7	2015
Objective form 7	C7	C8	2015
Objective form 8	C8	C1	2006, 2009 & 2015
Inquiry task 1			2015
Inquiry task 2			2015
Inquiry task 3			2015
Inquiry task 4			2015

Each cluster contained approximately 20 items. Based on the assumption that each item takes 1.5 minutes, each cluster contained approximately 30 minutes of material. The trial tests contained 204 different items, including inquiry tasks and associated items. In each class, the eight objective test forms were randomly assigned.

A total of 961 students participated in the trial. Each item appeared in two out of eight objective tests and an average of 232 student responses were collected for each objective item. Students sat one of the four inquiry tasks, providing around 218 responses for each inquiry task item.

Development and trialling of assessment instruments

The following sub-sections outline the process followed for the development of items for trialling; the types of items developed; the distribution of items across items types, levels and strands; the development of the student survey; and the trial process.

Item development workflow

In advance of commencement of item development, the test development manager and project manager prepared a work plan, which was approved by the project director (EAA).

The work plan outlined the tasks for the development and review of the stimulus, items and tasks in-house prior to releasing materials for Australian Curriculum, Assessment and Reporting Authority (ACARA) and Science Literacy Working Group (SLWG) reviews.

Specific criteria were developed by EAA and ACARA to guide the ACARA and SLWG reviews. Processes were also established for recording feedback on tasks and items as the reviews proceeded, and associated documentation was prepared.

All inquiry tasks were also piloted as a part of the review process and the feedback from these pilot studies was taken into account in the review process. Based on the feedback received, items were amended accordingly.

Definition of item types

Items developed were either contained within contextualised units or within thematic inquiry tasks. Items were classified by type in multiple ways to ensure cross-compatibility with previous cycles, methods of psychometric analysis and with the native item types available in the item authoring system.

A full treatment of item classifications used can be found in the assessment framework bridging document.

Item pool

The composition of items selected for trial, including the historical items, is presented here.

Table A7.2

Characteristics of items selected for trial

Australian Curriculum: Science strand	Objective items	Inquiry task items	Total
Science as a human endeavour	14	1	15
Science inquiry skills	65	40	105
Science understanding	79	5	84
2012 NAP–SL concept area			
Earth and space	37		37
Energy and force	53	25	78
Living things	36	21	57
Matter	32		32
2012 NAP–SL strand			
A	20	17	37
B	41	21	62
C	97	8	105
Type			
Composite	20	10	30
Extended text	46	17	63
Hotspot	6		6
Interactive gap match	8	1	9
Interactive graphic gap match	7	3	10
Multiple choice	60	11	71
Multiple choices	7	2	9
Position object	2		2
Select point	1	2	3
Text entry	1		1
Total	**158**	**46**	**204**

Student survey

An innovation for 2009 NAP–SL was the inclusion of a student survey. Following discussions with ACARA, the survey items included in the 2012 study were re-authored into the item authoring system for online delivery.

The survey items were also reviewed against the science as a human endeavour strand of the Australian Curriculum: Science. Aspects of that strand, which were not covered by existing survey items and which were suitable for the survey, were identified. As a consequence, two new clusters of survey items were developed and reviewed by the SLWG.

Trial process

The NAP–SL trial was administered online in March 2015. The trial had two purposes:

- to obtain item level data in order to inform the final item pool for main study
- to trial the administration procedures and technology.

Students from approximately 50 schools selected from New South Wales, Australian Capital Territory, Queensland and Victoria participated in the trial. The trial schools were selected to reflect the range of educational contexts around the country. This included government, Catholic and independent; metropolitan and regional; large and small; low and high socio-economic areas; and students with language backgrounds other than English.

Each student completed one of the eight trial objective tests and one of the four inquiry tasks. Tests were allocated randomly before the trial period so that all tests and all inquiry tasks were undertaken at each participating school.

Classroom teachers were required to administer the national sample assessment in October 2015, so it was important that the trial be conducted in the same way. Classroom teachers were designated as test administrators and provided with an administration manual before the trial to allow them to familiarise themselves with the test procedures. At the completion of each session, the test administrator completed a session report form to provide feedback about various aspects of the trial. This feedback, in conjunction with a range of other sources of feedback, informed refinements to the administration manual.

Before the assessment, all participating schools conducted the online technical readiness test (TRT) on assessment-designated computers. As backup technology was not available for the trial, it was necessary to replace one Queensland government school due to a known issue relating to insufficient internet connectivity.

EAA first level helpdesk

During the trial, a helpdesk was available to respond to any technical issues that arose. Forty-seven schools (out of the 51 schools that participated) called the helpdesk. Of the total calls received, half came from New South Wales, a quarter from Victoria and a quarter from Queensland. The table below shows the breakdown of calls:

Table A7.3

Calls to helpdesk by type

Query/reason for call	No. of calls
Test date query (including schools confirming their date, wanting to move the date with test observers attending)	17
Document request (including requesting email or document to be resent)	7
Mode-effect study query	2
Code request (including student list resent with login information, school code)	10
Absent student (including students who have left the school)	12
Computer/network issue (including freezing computers, computers just not working)	9
Students locked out	17
Special needs students	5
Sound issues	2
Other (including contact details updated, queries on selected students, requests for re-selection of students as not in the same class, time each session takes, parental permission required, individual reports for students)	27

Helpdesk calls, test administrator and test observer feedback summary

Issues from both helpdesk calls and test administrator / test observer feedback were summarised, with a total of 23 technical issues and 15 item related issues recorded. As multiple schools reported on the same issues, the total number of unique issues from schools was 127. The breakdown of these issues between local, procedural, system and unknown issues is represented below:

Table A7.4

Summary of technical issues

Category	Number of issues
Local	77
Procedural	32
System	10
Unknown	8
Total	**127**

Feedback was received from external test observers who attended half of the schools. Of the remaining half (25 schools), all but six schools provided responses.

Video

EAA conducted pilots of the 2015 NAP–SL inquiry tasks in November 2014. Four tasks were piloted: Bouncing balls, Mustard seeds, Pendulums and Sunscreen. All four tasks were included in the trial. During the trial, several schools reported that video files were often slow to start because of video buffering.

Thirteen schools reported a range of technical issues. These included: constant refreshing of cache needed; freezing of video; students playing videos in unexpanded form to avoid the lagging; and one school reported the median time for most students to successfully download was four minutes. A further seven schools reported incidents of videos not working altogether during the inquiry task.

It appears that when all students opened the inquiry task at the same time, loading issues that were not identified during the TRT became apparent.

Headphones

Schools were expected to provide headphones. The helpdesk team reported six schools experiencing difficulty with headphones. The issues with headphones and sockets resulted in students having to change computers or headphones, being delayed and having to work through lunch breaks with distractions, or abandoning the test altogether.

Logging in processes

Schools logging onto the school network, opening browsers and logging onto the internet continued to challenge some schools. EAA's helpdesk team reported seven schools experiencing difficulty with the logging on process; however, it was possible that other schools experienced similar issues but managed to resolve them without needing to contact helpdesk support.

Students at one school had issues logging onto the school's own computers because students usually logged onto their own personal laptops. One school reported it took 40 minutes to have all students logged on correctly. Another school reported a security certificate issue with the site being blocked, and students needing to accept the URL, which caused confusion. Once on the required website, a further three schools reported students having difficulties with logging in to the website, mostly due to loading issues.

Observation visits

Observations were conducted across half of the participating schools by staff employed by EAA. Observers were given training prior to the trial and provided with a trial test observers manual and feedback sheets.

In addition to these visits, project managers from EAA and Educational Services Australia (ESA) visited three schools.

Feedback from the observations were used to inform the final assessment materials.

Technology used

As in the NAP–SL pilot, technology requirements for trial participation were the minimum standards to access the system with a sufficient number of computers for a Year 6 class to participate.

Table A7.5

Technical requirements for schools (trial)

Hardware	Operating system	Web browser for student access in order of preference
• Device • Laptop/desktop PC • Chromebook • iPads* • Android tablets* • Mouse • Sound card plus speakers/ headphones	• Mac OS X • MS Windows XP, Vista, 7 and 8 • Chrome OS 38+	1. Chrome 28+ (XP, Vista, 7, 8, OSX, Chrome OS) 2. Firefox 20+ (XP, Vista, 7, 8, OSX) 3. Internet Explorer 10+ (7, 8) 4. Internet Explorer 9 (Vista, 7) 5. Safari (OSX) 6. Internet Explorer 8 (XP, Vista, 7) Note: IE8 is the least preferred browser due to an older technical design. In some configurations the student experience can be diminished.

* Supported tablet devices online

- Compatibility exists with most tablet devices, but not all configurations have been able to be verified as fully supported for the customised NAP–SL test player.
- It is essential to conduct a satisfactory technical readiness test with any intended tablet device before the assessment day. If a tablet does not render correctly then an alternative device will need to be used.
- For the best student experience it is highly recommended to use an external keyboard with tablet devices.

Settings
1024 x 768 monitor resolution, or higher (this is the recommended minimum resolution to maintain optimal screen display).
JavaScript must be enabled in the browser.
Bandwidth to internet (for the school): recommend 512 Kbps (0.512 Mbps) minimum download speed for up to 20 users.

Note 1: Cookies do not need to be enabled.
Note 2: There is no use of pop-ups.

The following table shows as a percentage of participating schools (trial and mode-effect online study) the operating system and browser used based on a sample of students extracted from the test delivery system data.

Table A7.6 Platforms used as a percentage of schools

Operating system	Browser – percentage of schools				Total
	Chrome	Firefox	Internet Explorer	Safari	
Chrome OS	4	0	0	0	4
iPad	0	0	0	4	4
Mac OS	8	2	0	0	10
Windows 7	13	2	50	0	65
Windows 8.1	8	2	6	0	17
Total	**33**	**6**	**56**	**4**	**100**

Note: Percentages may not add to 100 per cent due to rounding.

As can be seen from the table, the most common platform (operating system + browser) was Windows 7 with Internet Explorer (IE).

Details of the platforms used were also collected by the test observers who visited a sample of the schools. Of the schools they visited, a majority held the test in a dedicated IT lab (or equivalent). The following table gives a breakdown by room and browser used.

Table A7.7 Room type and browser used

Room type	Chrome	Firefox	Internet Explorer	Mixed	Total
Classroom	2	0	3	2	7
IT lab	3	1	9	3	16
Library-mixed use	0	0	2	0	2
Total	**5**	**1**	**14**	**5**	**25**

The most common arrangement (36 per cent of observed schools) was for a school to use Internet Explorer in a dedicated IT lab.

Data from test observers on the number of computers available in the testing room are less clear. Observers were asked to count the number of machines available, but in many schools the machines allocated to a room were supplemented by extra notebooks or iPads that are normally shared across multiple classrooms.

Twenty-five was the modal number of computers available in the rooms used for testing (that is, not including any additional machines brought in to make up the numbers) but there was a wide range between schools. This was mitigated by most schools being able to supplement the number of available machines with portable devices from other classrooms or as a shared resource.

Technical recommendations

Overall, the volume of helpdesk calls was manageable. Troubleshooting issues were largely minor (aside from one school) and were within the capability of the classroom teacher to resolve quickly. However, classroom management remains an issue for the test as some students had to sit and wait while the teacher dealt with specific issues.

After a preliminary review of all technical issues reported, technology providers ESA and Janison initially concluded that:

- The vast majority of issues appear to be either local manifestations (due to bandwidth and network limitations) or procedural in nature.
- As far as possible, they would investigate to see if any technical design changes may reduce the incidence of the issues (for example, the hosting of the site and videos were to be upgraded to Microsoft Azure).
- There appeared to be a number of technical issues with devices and browsers that would have been expected to be resolved during the TRT. This suggested the technical testing at the school level could have been more comprehensive.
- It should be noted that the technical issues reported did not occur during testing of the application by ESA, and a number of schools did not experience significant technical disruptions. That is to say, the reported technical issues are not systemic bugs. The pattern that has emerged is more indicative of how online functionality may struggle when local connectivity or network throughput capacity becomes a constraint.

Results of trialling

The trial of items for 2015 generated several sources of evidence about the performance of the items and test as a whole. Feedback from test administrators and observers provided further information about aspects of the tests including overall length. Markers involved in scoring the constructed response items gave feedback about students' responses and the ability of the marking guide to capture their responses appropriately. Most importantly, data analysis of the students' scores provided empirical evidence about the statistical characteristics of the items.

Marking process

A team of five experienced markers was engaged for a seven-day period following the trial. The marking centre was located onsite at EAA. The test development manager and project director trained the markers and were onsite to oversee the marking process. Markers were briefed on the process, with a focus on the task of using the criteria provided in the draft marking guides to mark items with extended-text responses.

All items were marked online using a marking system devised to work with the test delivery system. EAA staff, in collaboration with the technology providers (ESA and Janison), set up the marking system prior to the marking centre. Markers were then trained in the use of the marking system.

Markers were given specific training on the marking schemes of all relevant items. They were briefed by the test development manager and project director, who led them through an overview of each unit or inquiry task, discussed the marking criteria for the relevant items and then provided feedback on an initial sample of items marked by the markers. As the test development manager and project director were onsite during marking, advice on any issues that arose was provided immediately. Useful extra text was gathered in this session, including illustrative answers for correct and incorrect student responses. This was used to update the marking guides accordingly.

To maintain quality control, double marking was used initially to ensure consistency of marking. Any inconsistency between marks was escalated to senior staff in the system. In addition to this, marks were sampled by senior staff to ensure that marking was correct and consistent throughout.

Data analyses

Data were exported from the test-delivery/marking system by ESA and then processed and analysed. The results of these analyses were compiled onto a psychometric input data sheet supplied by ACARA.

Key criteria for judging the performance of items were measures of item fit statistics (weighted MNSQ) and performance illustrated by Item Characteristic Curves (ICC). Percentage correct and point-biserial correlation were noted, but only informed a decision to eliminate an item if other indices were poor. Based on the initial analysis (first run), four items were eliminated because of poor indices and flagged as ‘reject’. An extra item was flagged as ‘reject’ due to technical issues with how the student response data were captured by the system.

The following diagram illustrates the distribution of all trialled items (indicated by item identifiers used during analysis). This diagram provides ‘at a glance’ the range of difficulty of the items, and how they align with the ability of students in the trial pool (each ‘x’ represents five students). The range of item difficulty was nearly 8 logits; the easiest item was I0113 at 4.06 logits. The most difficult item was I0123 at 3.83 logits.

Figure A7.1

Item-person map of 2015 NAP–SL trial items

```
==========================================================================================================
LOCATION              PERSONS   ITEMS [locations]
----------------------------------------------------------------------------------------------------------
   4.0                        |
                              | I0123
                              |
                              |
                              | I0194
   3.0                        |
                              | I0193 I0114 I0128 I0111
                              | I0133 I0035
                              | I0086 I0089 I0177 I0174 I0154 I0181
                              | I0192 I0083
   2.0                        | I0121 I0139
                            x | I0085 I0196 I0203 I0132 I0078 I0096 I0057
                            x | I0171 I0101 I0110
                            x | I0202 I0153 I0145 I0080 I0072 I0137 I0144 I0107
                           xx | I0106 I0074 I0156 I0094 I0098 I0122
   1.0                     xx | I0138 I0041 I0119 I0158 I0131 I0016 I0180 I0204 I0136  I0095
                      xxxxxxx | I0004 I0043 I0082 I0006 I0103 I0151 I0188
                     xxxxxxxx | I0161 I0147 I0109 I0176 I0069 I0125 I0190
                     xxxxxxxx | I0052 I0034 I0152 I0172 I0051 I0020 I0200 I0100 I0076  I0036 I0159 I0201
                  xxxxxxxxxxx | I0173 I0178 I0197 I0141 I0187 I0102 I0150 I0157 I0097  I0048 I0127 I0011
   0.0         xxxxxxxxxxxxxx | I0081 I0185 I0033 I0146 I0075
              xxxxxxxxxxxxxxx | I0084 I0149 I0093 I0092 I0019 I0073 I0116 I0120 I0066  I0195 I0191 I0170 I0155 I0104 I0091
             xxxxxxxxxxxxxxxx | I0198 I0015 I0169 I0143 I0162 I0077 I0012 I0058 I0087  I0070 I0126 I0056 I0118 I0088
               xxxxxxxxxxxxxx | I0071 I0032 I0026 I0049 I0160 I0046 I0014 I0067 I0002  I0063 I0060
           xxxxxxxxxxxxxxxxxx | I0028 I0186 I0029 I0105 I0068 I0148 I0134
  -1.0          xxxxxxxxxxxxx | I0042 I0013 I0001 I0009 I0040 I0023 I0047 I0065 I0030  I0005 I0175 I0062
                 xxxxxxxxxxxx | I0021 I0117 I0055 I0112 I0184 I0017 I0165 I0022 I0025  I0064 I0007 I0003
                 xxxxxxxxxxxx | I0124 I0168 I0140 I0179
                   xxxxxxxxxx | I0199 I0050 I0045 I0090 I0024 I0130 I0135
                     xxxxxxxx | I0039 I0164 I0183 I0059
  -2.0                xxxxxxx | I0008 I0099 I0129 I0166 I0010 I0031
                        xxxxx | I0018 I0182 I0061 I0079 I0108 I0038
                           xx | I0115 I0054
                           xx | I0027 I0053 I0044 I0142
                            x | I0037
  -3.0                      x | I0167
                              |
                              |
                              | I0189 I0163
                              |
  -4.0                        | I0113
----------------------------------------------------------------------------------------------------------
                  x = 5 Persons
----------------------------------------------------------------------------------------------------------
```

As can be seen from the diagram, the 2015 trial assessment achieved an excellent spread of item difficulties but did contain many difficult items for the Year 6 cohort. There were a number of items that all students found to be very easy, a number of items that were challenging (even for the most able students), and many items in the middle range.

Differential Item Functioning (DIF) analyses for gender were carried out for all remaining items. However, DIF analysis for language (LBOTE) could not be considered due to small sample size and the lack of information about specific language background provided by students who participated in the trial.

The DIF analyses were carried out using Conquest by fitting a facets model, where the interaction between an item and the gender group is estimated. In cases where items exhibited large DIF, content experts inspected the reasons for the observed bias. The items were flagged but not automatically removed simply based on statistical evidence of bias.

Based on this psychometric analysis as well as more general feedback from the trial, items were classified into four categories based on their overall quality. These categories were used to help inform the SLWG post-trial review.

Table A7.8

Categories used to classify trial items by quality

Category	Explanation
Keep	No obvious issues with the item
Review	Some issues that should be considered when including the item in a test
Low priority	The item could be viable but is not ideal for use
Reject	Not a suitable item for inclusion in a test

Student survey results

Students' responses to the survey items were exported from the test delivery system and processed by EAA.

Feedback from the Science Literacy Working Group (SLWG)

All trialled items were provided to the SLWG to view in the authoring system. Members were invited to view the stimuli and items, as well as the associated metadata, before the upcoming panel discussion in Sydney. They were also provided with a spreadsheet with selected metadata and a summary recommendation.

This pool was discussed at a meeting with the SLWG and approved for use in the 2015 assessment.

During the meeting, SLWG members were presented relevant psychometric data on the items that included:

- facility (per cent correct)
- weighted MNSQ
- discrimination (overall and for each distractor where applicable)

- gender DIF
- ICC graphs.

SLWG survey review

Also at this meeting, the results from the trial NAP–SL student survey were presented. SLWG members were invited to comment on the survey items and provide a priority for inclusion in the final selection. SLWG members discussed the results, recommended changes to a number of survey items and agreed on a final list of survey items to be included in the main assessment.

SLWG inquiry tasks review

The four inquiry tasks were discussed in detail. For the main study, two inquiry tasks of approximately 10 items were required and the SLWG discussed the performance of all four tasks to help evaluate which two tasks should be selected.

SLWG objective test review

The working group looked at all the objective test items that had been flagged as 'Review' and were also invited to raise issues with any other items.

Wording of items was discussed but to ensure that the psychometric properties of items did not change only minor changes in wording were agreed to. Additionally, some items flagged as 'Review' were moved to 'Low priority' due to issues highlighted during discussion.

It was also agreed to make some general editorial changes to ensure consistency of style in multiple-choices style items by putting references to number of options to be chosen in bold.

Table A7.9

Item classification of objective test items following the SLWG review

Unit title	Keep	Review	Low priority	Reject	Total
Historical units	21	21	1		43
New objective items	58	25	8	2	93
Inquiry tasks	22	11	12	1	46
Total	**101**	**57**	**21**	**3**	**182**

Table A7.10

Average facility for each inquiry task

Unit title	Average percentage correct
Task 1 – not used	29.7
Task 2 – backup task for 2015 main study	50.2
Task 3 – used in 2015 main study	47.5
Task 4 – used in 2015 main study	44.6

All feedback from the SLWG meeting was collated and formed the basis for deciding which items were chosen for the final item pool and any modifications to items.

Appendix

7

Glossary

Appendix 7. Glossary

See also the National Assessment Program glossary.

Term	Definition
Assessment domain	A subject or learning area that is the focus of a test. There are three domains for sample assessments: civics and citizenship information and communications technology literacy science literacy.
Assessment framework	A clear definition of a scope and method of testing in each of the three areas included in the NAP sample assessments.
Construct	The underlying cognitive abilities measured by an assessment.
Critical and creative thinking	One of the general capabilities of the Australian Curriculum.
Confidence interval	An indication of the statistical uncertainty to expect when estimating population parameters using statistics from a sample (e.g. a sample mean score).
Domain	A subject or learning area that is the focus of a test. There are three domains for sample assessments: civics and citizenship, information and communications technology literacy and science literacy.
Exempt	A students who is not assessed. Students with a language background other than English, who arrived from overseas less than a year before the tests, and students with significant intellectual disabilities or co-existing conditions may be exempted from NAP testing.
Experiment	A type of investigation that involves manipulation of one or more independent variables, and observes the effect on some outcome (dependent variable).
Factor analysis	A statistical technique that identifies underlying factors (called latent factors or latent commonalities) by the intercorrelations between data such as survey items.
General capabilities	A dimension of the Australian Curriculum that describes seven capabilities important for life and work in the 21st century.

Term	Definition
Geographic location	A locality of an individual school that is used to disaggregate data according to 'metropolitan', 'provincial', 'remote' and 'very remote'.
Indigenous status	A student is classified as being of Aboriginal and/or Torres Strait Islander origin, based on information from the school provided by the student, or their parent/guardian. The Melbourne Declaration and national data collections use the term 'Indigenous' to refer to Australia's Aboriginal and Torres Strait Islander Peoples.
Information and communication technology	One of the general capabilities of the Australian Curriculum.
Investigation	A scientific process of answering a question, exploring an idea or solving a problem that requires activities such as planning a course of action, collecting data, interpreting data, reaching a conclusion and communicating these activities.
Item	A general term referring to an individual problem, question, or task used within a test, i.e. a test item is a test question (or part of a question).
Jurisdiction	State or territory in Australia. Test administration authorities are responsible for the implementation and administration of the NAPLAN tests in their jurisdiction (e.g. New South Wales, Northern Territory, Tasmania etc.).
Language background other than English (LBOTE)	A student is classified as LBOTE based on information from the school provided by the student or their parent/guardian as to whether mainly a language other than English is spoken at home.
Learning continuum	A description of the relevant knowledge, skills, behaviours and dispositions within a learning area of the Australian Curriculum at particular points of schooling.
Literacy	One of the general capabilities of the Australian Curriculum.
Likert style items	Items to which respondents are required to select one of several categories (e.g. 'strongly disagree', 'disagree', 'agree', 'strongly agree') to indicate their intensity of response to the item.
Logits	A unit of measurement, which forms an equal interval linear scale, used to express item difficulty and person ability.
Metadata	Classification of data that provides a higher level of information about test items.
Multilevel modelling	A statistical technique used for analysing data, which varies at more than one level (for example, student performance may vary by both gender and socio-economic status).
Numeracy	One of the general capabilities of the Australian Curriculum.
OECD	The Organisation of Economic Cooperation and Development.
Online testing	A delivery of test items via computers delivered over the internet.
Organising element	A component of the Australian Curriculum consisting of related concepts, understandings and skills.
Personal and social capability	One of the general capabilities of the Australian Curriculum.
PISA	The Programme for International Student Assessment, a triennial international survey evaluating education systems by testing students at 15 years old. PISA is managed by the OECD.
Practical assessment	An assessment of practical and inquiry skills taught as a part of the Australian Curriculum: Science.
Proficiency levels	Ranges on a scale accompanied by descriptions of capabilities associated with each level.

Term	Definition
Proficient standard	A standard in each of the NAP sample assessments that represents a reasonably challenging level of performance where students need to show more than the minimal skills expected at that year level. Proficient standard is not the same as national minimum standard because the latter refers to the basic level needed to function at that year level, whereas proficient standard refers to what is expected of a student at that year level.
Progress map	A description of skills and understandings that are progressively more demanding across levels.
Psychometrics	An area of statistics concerned with educational and psychological measurement. This include construction of tests, construction of measurement scales and development and refinement of theoretical approaches to measurement.
Rasch model	A statistical model developed to analyse data and responses from assessments such as the NAP tests. It enables student abilities and item (question) difficulties to be placed on the same scale.
Regression analysis	A statistical technique used for quantifying a relationship between two or more sets of data.
Science literacy	An ability to use scientific knowledge, understanding and inquiry skills to identify questions, acquire new knowledge, explain science phenomena, solve problems and draw evidence-based conclusions in making sense of the world, and to recognise how understandings of the nature, development, use and influence of science help us make responsible decisions and shape our interpretations of information (Australian Curriculum).
Science literacy scale	A proficiency scale for NAP sample assessment – science literacy. Proficiency scales are developed for NAP sample assessments, and enable comparisons of results in each domain to be made from year to year. Proficiency scales are similar to the assessment scales for NAPLAN but are divided into proficiency levels instead of bands.
Statistically significant	A probability that a result may have occurred by chance. It is generally used with data collected from samples. If a difference between two groups is statistically significant, then the observed difference is likely to also occur in the entire populations from which the samples were drawn, and unlikely to have occurred because of some artefact of sampling.
Stimulus	Material used in assessments to provide context for assessing the knowledge and skills of students.