National Assessment Program

Science Literacy 2023

Technical Report

AUSTRALIAN CURRICULUM, ASSESSMENT AND REPORTING AUTHORITY

This technical report was written by Kate O'Malley, Dominique Hall, Dr Eunjung Lee, Frances Eveleigh, Clare Ozolins, Bethany Davies and Nina Martinus from the Australian Council for Educational Research (ACER).

NAP–Science Literacy team from ACARA

Wei Buttress – Senior Manager, Field Operations

Stephen Phillip – Senior Manager, Measurement and Evaluation

Leigh Patterson – Senior Psychometrician

Simon Collier – Curriculum Specialist (Science)

Danielle Cavanagh – Curriculum Specialist (General Capabilities)

Sudan Chandra Subedi – Senior Project Officer

Joel Teoh – Project Officer

NAP–Science Literacy team from ACER

Frances Eveleigh and Kate O'Malley – Project co-Directors

Dr Celia McNeilly and Gayl O'Connor – Lead Test Developers

Dr Tim Friedman – Lead Questionnaire Developer

Dr Eveline Gebhardt and Dr Eunjung Lee – Psychometricians

Dominique Hall – Field Operations Lead

Michael Beattie – Project Officer

Dr Martin Murphy, Nina Martinus and Emma Camus – Sampling

Acknowledgement of Country

ACARA acknowledges the Traditional Owners and Custodians of Country and Place throughout Australia and their continuing connection to land, waters, sky and community. We pay our respects to them and their cultures, and Elders past and present.

Copyright

Contact details

Australian Curriculum, Assessment and Reporting Authority
Level 13, Tower B, Centennial Plaza, 280 Elizabeth Street Sydney NSW 2000
T 1300 895 563 | F 1800 982 118 | www.acara.edu.au

Table of Contents

List of tables

List of figures

List of acronyms

Acronym	Full form
ACARA	Australian Curriculum, Assessment and Reporting Authority
ACER	Australian Council for Educational Research
AEST	Australian Eastern Standard Time
AMS	ACER Marking System
ASGS	Australian Statistical Geography Standard
CCT	Critical and Creative Thinking
DIF	differential item functioning
ICT	information and communication technology
ID	identification
IRT	Item Response Theory
ISM	Information Security Manual
JRR	jackknife repeated replication
KPM	Key Performance Measure
MCEETYA	Ministerial Council on Education, Employment, Training and Youth Affairs
NAP	National Assessment Program
NAPLAN	National Assessment Program Literacy and Numeracy
NAP–Civics and Citizenship	National Assessment Program Civics and Citizenship
NAP–Information and Communication Technology Literacy	National Assessment Program Information and Communication Technology Literacy
NAP—Science Literacy	National Assessment Program Science Literacy
OECD PISA	Organisation for Economic Co-operation and Development Programme for International Student Assessment
PCA	principle component analysis
PMRT	Performance Measurement and Reporting Taskforce
PSAP	Primary Science Assessment Program
PV	plausible value
SBD	student background data

Acronym	Full form
SCO	School Contact Officer
SE	standard error
SEIFA IEO	Socio-Economic Index of Education and Occupation
STEM	science, technology, engineering and mathematics
STSO	School Technical Support Officer
TA	Test Administrator
TIMSS	Trends in International Mathematics and Science Study
WLE	weighted likelihood estimates

Terms used in this report

Term	Definition
Assessment platform	The Online National Assessment Platform enables the online delivery of National Assessment Program events including NAP–Science Literacy, NAP–Civics and Citizenship and NAPLAN.
Confidence interval	An estimate derived from a sample is subject to uncertainty because the sample may not reflect the population precisely. The extent to which this variation exists is expressed as the confidence interval. The 95% confidence interval is the range within which the estimate of the statistic based on repeated sampling would be expected to fall for 95 of 100 samples that might have been drawn. Confidence intervals are provided in each of the data tables in this report.
Correlation coefficient	A statistical measure that indicates the degree to which 2 variables are related. The values range between -1.0 (a perfect negative correlation) and 1.0 (a perfect positive correlation). A coefficient of 0.0 shows no linear relationship between the 2 variables being studied.
Critical and Creative Thinking	In this report, when the initial letters of the term “Critical and Creative Thinking” are capitalised, it refers to the general capability of Critical and Creative Thinking in the F–10 Australian Curriculum. When the term is written without capitals, it refers to the broader thinking skills of reason, logic, innovation and creativity.
Effect size	The difference between group means divided by the standard deviation. Effect size provides a comparison of the difference in average scores between 2 groups with reference to the degree in which the scores vary within the groups. When the effect size is large, it means that the difference between average scores is large relative to the spread of the scores. The difference could therefore be considered “important”. Conversely, when the effect size is small, it means that the observed difference is relatively small compared with the spread of the scores and thus arguably less “important”.
Enemy item	Items that should not appear in the same test form because they are too similar or because one gives away the answer to the other.
Exempt	Students with very limited English language proficiency and students with significant intellectual or functional disabilities may be exempted from NAP sample testing.
Geolocation	The Australian Statistical Geography Standard (ASGS) Remoteness Structure is used to classify relative geographic remoteness across Australia. In this report, the 5 classes (major cities, inner regional, outer regional, remote and very remote) are collapsed into 3 classes (major cities, regional and remote) for the purposes of classifying the remoteness of individual schools.
Indigenous status	A student’s Indigenous status refers to whether a student identifies as being of First Nations Australian Aboriginal and/or Torres Strait Islander origin. The term “origin” is considered to relate to people’s First Nations Australian Aboriginal or Torres Strait Islander descent and for some, but not all, their cultural identity. A student who identifies as a First Nations Australian student is also considered to be of Aboriginal and/or Torres Strait Islander origin.

Term	Definition
Inquiry task	A set of contextualised independent items in the NAP–Science Literacy instrument that aims to engage students and assess methods of scientific inquiry. A student is led through a whole scenario content sequence and asked to apply scientific skills to answer predominantly open-ended questions across various response formats.
Jurisdiction	Each of the 3 educational sectors (government, Catholic and independent) that sit within an Australian state or territory. The state/territory level is the most granular level of analysis undertaken for the purposes of NAP sample reporting.
Language other than English spoken at home	A language other than English spoken in the home by a student. If a student speaks more than one language other than English at home, the language other than English the student speaks most often is reported.
Limited assessment language proficiency	The student is unable to read or speak the language of the assessment and would not be expected to overcome the language barrier in the assessment situation. Typically, a student who had received less than one year of instruction in the language of the assessment would be excluded.
NAP–Science Literacy Assessment Framework	The overarching assessment design that describes the content to be assessed, the cognitive engagement that is expected of students, the types of assessment tasks, contextual information and overall structure of the assessment.
NAP–Science Literacy scale	A continuous scale that provides a measure of student achievement in science literacy.
Parental education	The highest level of parental school or non-school education that a parent/guardian has completed. This includes the highest level of primary or secondary school completed or the highest post-school qualification attained. For the purposes of this report, where a student has parental education data for 2 parents/guardians, the higher of the 2 values is used.
Parental occupation	The occupation group that includes the main work undertaken by the parent/guardian. If a parent/guardian has more than one job, the occupation group that reflects their main job is reported. For the purposes of this report, where a student has parental occupation data for 2 parents/guardians, the higher of the 2 values is used.
Percentage	A number or ratio that can be expressed as a fraction of 100. In this report, the percentages of students represented in the tables have been rounded and may not always sum to 100.
Proficiency level	A defined range of the NAP–Science Literacy scale that describes the knowledge and competencies that students at that level are capable of successfully demonstrating.
Proficient standard	A point on the scale that represents a “challenging but reasonable” expectation of student achievement at that year level.
Rasch model	A psychometric model of Item Response Theory for analysing categorical data. It is the chosen model of analysis for cognitive and contextual data across all NAP sample assessments

Term	Definition
Response rate	Response rates are the percentages of sampled students that participated in the assessment. Response rates are calculated as the number of assessed students from whom data were recorded as a percentage of the total number of sampled students in the year level.
Sample	A subset of a population selected so that reliable and unbiased estimates of statistics for the full population can be inferred.
Science literacy	The 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.
Sector	The 3 educational sectors of government, Catholic and independent. All schools throughout Australia belong to one of these 3 school sectors. It is important to note that student responses for NAP sample assessments, in their most disaggregated form, are not analysed or reported by sector but are instead examined at the jurisdictional level.
Severe functional disability	A moderate to severe permanent physical disability that severely limits a student’s capacity to participate in the test.
Severe intellectual disability	A mental or emotional disability and/or cognitive delay that severely limits a student’s capacity to participate in the test.
Significant	In this report, the term significant refers only to differences that are statistically significant. Once a difference has been identified as statistically significant, the size of this difference (ranging from a small to very large effect size) can be considered.
Significant difference	Refers to the likelihood of a difference being a true reflection of the measured outcomes rather than the result of chance.
Standard deviation	A measure of variability or dispersion in student scores from the mean (or average).
Test form	A collection of selected items sequenced, balanced and grouped together to measure a student's knowledge, skills and understanding of a subject area.
Trend item	An item (test question) used in at least one of the previous NAP–Science Literacy assessment cycles.

Chapter 1: Introduction

Chapter 1: Introduction

The National Assessment Program

The National Assessment Program (NAP) was established to measure student achievement and to monitor progress towards the education goals first outlined in the 1999 Adelaide Declaration on National Goals for Schooling in the 21st Century. As part of the NAP, ministers for education in Australia agreed that nationally comparable data across jurisdictions would be collected in the domains of literacy, numeracy, science literacy, information and communication technology (ICT) literacy, and civics and citizenship.

The NAP–Science Literacy assessment is one of 3 national sample assessments developed and managed by the Australian Curriculum, Assessment and Reporting Authority (ACARA) under the auspices of the Education Ministers Meeting. Together with the NAP–Civics and Citizenship (NAP–CC) and the NAP–Information and Communication Technology Literacy (NAP–ICT Literacy), the NAP–Science Literacy assessment supports the measurement of progress towards the goals first set out in the Adelaide Declaration. These goals were upheld in the subsequent Melbourne Declaration (2008) and Alice Springs (Mparntwe) Education Declaration (2019), and they continue to provide the impetus for the NAP sample assessments.

Background to the NAP–Science Literacy assessment

For the NAP–Science Literacy, the first collection of data was from a sample of Year 6 students in 2003[1]. Subsequent cycles of the assessment involving Year 6 students have been conducted on a rolling 3-yearly basis in 2006, 2009, 2012 and 2015.

In 2018, the assessment was extended to include Year 10 students so that both primary and secondary school student progress in science literacy could be measured by an assessment closely aligned with the Australian Curriculum. The inclusion of both Year 6 and Year 10 student data was maintained for the most recent assessment cycle in 2023[2].

This report describes the various technical, operational and administrative procedures of the NAP–Science Literacy 2023 assessment.

Sample

The NAP–Science Literacy 2023 assessment was based on a nationally representative sample of 589 participating schools with 9,502 participating students, of which 6,069 were from Year 6 and 3,433 were from Year 10. The weighted national school response rate when including substitute schools was 88% for Year 6 and 82% for Year 10.

Sampling followed a 2-stage cluster sampling design to ensure that each eligible student had an equal chance of being selected in the sample. In the first stage of sampling, schools were selected from a list of all schools in each jurisdiction with a probability proportional to the number of students in the relevant year level enrolled at that school. In the second stage, 20 students[3] were selected with equal probability from a list stratified by gender for each target year level.

[1] In 2003, the assessment was known as the Primary Science Assessment Program (PSAP).

[2] The 5-year gap between 2018 and 2023 was a result of disruptions caused by the COVID-19 pandemic.

[3] Or up to 20 students when schools had fewer than 20 students in the year level.

Stages of assessment development and implementation

For any large-scale assessment, a series of delineated stages must be planned to deploy the assessment in the field, and then to analyse and report on the data collected. For NAP–Science Literacy 2023, the development, implementation, analysis and reporting of the assessment can be separated into 6 distinct stages, namely:

Stage 1: review and revision of the assessment framework

Stage 2: development of items, units, clusters and test forms for field trial deployment

Stage 3: implementation of the field trial to trial both test items and operational procedures

Stage 4: psychometric analysis of test items and subsequent selection of content for main study deployment

Stage 5: implementation of the main study in a scientific sample of schools and students across Australia

Stage 6: psychometric analysis of main study data, production of school summary reports and development of public reports.

A description of, and approximate timeframe for, each of the 6 stages is provided in Table 1.1.

Table 1.1: Stages of NAP–Science Literacy 2023 development

No.	Stage	Description	Timeframe
1	Assessment framework review	• ACER, ACARA and the NAP–SL Working Group worked together to review, revise and enhance the NAP–SL assessment framework. • The revised framework is coherent with the Australian Curriculum: Science while aligning with developments in the assessment of scientific literacy competencies within and outside of Australia.	Oct 2021 – Mar 2022
2	Test development	• A total of 311 new items (101 Year 6 items, 113 Year 10 items and 97 Year 6/10 link items) were developed to complement the inclusion of secure items that had been used in previous cycles. • With input and guidance from ACARA and the working group, ACER reviewed the previous student questionnaires. Where possible, questions were retained without major changes. Some edits were required to update wording or retain relevance to a 2023 audience. Other questions were added so that themes such as student use of CCT and attitudes to the COVID-19 pandemic could be explored.	Jan – Sep 2022

No.	Stage	Description	Timeframe
		• Both the assessment and questionnaire content were authored in the assessment platform. Extensive quality assurance (QA) and user acceptance testing (UAT) were then performed across a variety of device types.	
3	Field trial	• A field trial was conducted in sampled schools to trial both the assessment instruments and related operational procedures. • In total, 1,488 Year 6 students and 1,378 Year 10 students from 128 schools participated in the trial (66 from Year 6 and 62 from Year 10). • Schools in NSW, Vic, Qld, WA and SA were selected to participate to avoid burdening the comparatively oversampled schools from the smaller jurisdictions. • Trained quality monitors attended over 5% of test sessions in schools to provide feedback on adherence to test protocol, occurrence of technical issues and levels of student engagement. • A centre-based marking operation was implemented for extended response items. A total of 62 items were marked by a team of trained markers with rigorous quality assurance processes implemented.	**Test administration** 17 Oct – 4 Nov 2022 **Marking operation** 30 Nov – 5 Dec 2022
4	Item analysis and instrument revision	• All field trial data were consolidated, cleaned and processed in line with agreed data processing protocol. • All cognitive (assessment) and contextual (questionnaire) data were psychometrically analysed to determine the success of each item. • The NAP–SL Working Group met to review the item analysis and discuss ACER's item inclusion/exclusion recommendations.	Dec 2022 – Mar 2023

No.	Stage	Description	Timeframe
		• Main study item selection was confirmed and final test forms were constructed. Test forms were designed to ensure inclusion of vertical and historical links, broad equity of content, difficulty, length and score points, and the avoidance of enemy items.	
5	Main study	• In total, 589 schools from across Australia participated in the main study. This included 368 schools at a Year 6 level and 221 at a Year 10 level. • A total of 271 new items (90 Year 6 items, 104 Year 10 items and 77 Year 6/10 link items), complemented by 90 trend items, were administered across 36 test forms. • Trained quality monitors attended 35 test sessions in schools across all states and territories in Australia. Again, they reported back on test protocol adherence, technical issue occurrence and the level of student engagement in the assessment. • Trained centre-based markers marked all main study extended response items. A total of 69 items (with a total of 81 scoring rubrics) were marked with rigorous quality assurance processes implemented.	**Test administration** 8 – 31 May 2023 **Marking operation** 31 May – 13 Jun 2023
6	Data analysis and public reporting	• All cognitive and contextual data were collated, cleaned, processed and analysed by psychometricians. • School summary reports were developed and distributed to participating schools at the beginning of Term 3. • Two reports were developed for publication. The NAP–Science Literacy 2023 Public Report contains findings from 2023 including comparisons, where appropriate, with findings from previous assessment cycles. This technical report provides more detailed information about the technical processes and analytical procedures applied in the study.	Jul – Dec 2023

Reporting the NAP–Science Literacy 2023 results

The Rasch model was used to establish the empirical components of the NAP–Science Literacy reporting scale. The scale was first developed in 2006 using data collected from Year 6 students during the main study assessment of that year. While the inaugural NAP–Science Literacy assessment took place in 2003, the later shift in methodology for school and student sampling, as well as a change to the structure of the assessment itself, meant that the 2006 assessment data provided a more suitable baseline for scale development.

In 2006, the NAP–Science Literacy scale was set with a mean of 400 and a standard deviation of 100. In all subsequent cycles, data from the common items across assessment cycles (i.e. historical link items) were used to equate the assessments and derive comparable student achievement scores on the established NAP–Science Literacy scale.

In 2018, the scale was extended to incorporate the newly added Year 10 assessment instrument. Common questions between the Year 6 and Year 10 assessments, known as vertical link items, were developed in cycles 2018 and 2023. This made it possible to equate the assessment items from Year 6 and Year 10 so that student achievement could be reported across both year levels on the same scale.

The NAP–Science Literacy 2023 assessment includes a proportion of items that were used in 2018 and previous cycles. Using common item-equating procedures enabled the recoding of the results for NAP–Science Literacy 2023 on the scale that had been established in 2006. Consequently, the results from NAP–Science Literacy 2023 are directly comparable with those from all previous cycles of the assessment.

The NAP–Science Literacy scale comprises 5 proficiency levels that describe the achievement of students in Year 6 and, from 2018 onwards, Year 10. Typically, students whose results are located within a proficiency level can demonstrate the understandings and skills associated with that level. They also possess the understandings and skills of lower proficiency levels. With the addition of Year 10 content to the scale in 2018, as well as the implementation of a standard-setting exercise in the same year, adjustments to the width of the proficiency levels were made so that it adequately covered the breadth of scale scores across the 2 year-level cohorts.

In addition to deriving the scale and the associated described levels of proficiency, proficient standards were established in 2006 for Year 6 and in 2018 for Year 10. The proficient standards are points on the achievement scale that represent a challenging but reasonable expectation for typical students in that year level to have reached. The proficient standard for Year 6 is 393 scale score points, which is the boundary between levels 2 and 3 on the NAP–Science Literacy scale. The proficient standard for Year 10 is 497 scale score points, which is the boundary between levels 3 and 4 on the scale. Year 6 students performing at level 3 or higher and Year 10 students performing at level 4 or higher have consequently met or exceeded their relevant proficient standard. In 2023, 57% of Year 6 students reached or exceeded the Year 6 proficient standard, whereas 54% of Year 10 students were at or above the proficient standard for Year 10. Further information about students' science literacy achievement in 2023, including comparisons with previous years, can be found in the NAP–Science Literacy 2023 Public Report.

Purpose and structure of the technical report

This technical report complements the NAP–Science Literacy 2023 Public Report. The purpose of the public report is to summarise the cognitive and contextual analysis of the data collected in the NAP–Science Literacy 2023 sample assessment, while the purpose of this report is to describe the technical aspects of the assessment.

This report outlines the main activities involved in the assessment design, sampling and data collection, and the analysis and reporting phases of the assessment. The structure of this report is as follows:

Chapter 1 introduces the NAP–Science Literacy assessment and provides an overview of content within this report.

Chapter 2 summarises the development of the assessment framework and describes the process of item development and construction of the instruments.

Chapter 3 outlines the sample design and describes the sampling process. It also describes the weighting procedures that were implemented to derive population estimates and the calculation of response rates.

Chapter 4 describes the data collection, processing and management procedures used. This includes the steps taken to ensure strict data security protocol was followed, as well as the various methods of data capture that were employed before, during and after the administration of the assessment. The procedures employed in the transfer, tracking, verification, cleaning and transformation of the data are also outlined.

Chapter 5 describes the scaling model and procedures, item calibration, the creation of plausible values and the standardisation of student scores. It discusses the procedures used for vertical (Year 10 to Year 6) and horizontal (2023 to 2018, 2015, 2012, 2009 and 2006) equating, and the procedures for estimating equating errors.

Chapter 6 outlines the NAP–Science Literacy proficiency levels and proficient standards.

Chapter 7 outlines the reporting of student results, including the procedures used to estimate sampling and measurement variance, and the multivariate analyses conducted with data from NAP–Science Literacy 2023.

Chapter 2: Assessment framework and instrument design

Chapter 2: Assessment framework and instrument design

The NAP–Science Literacy 2023 Assessment Framework was the central reference for development of the assessment and questionnaire instruments. While the described proficiency scale generated using the 2006 data (and supplemented with item data from 2009, 2012, 2015 and 2018) was used as an indicator of item and task difficulty to inform instrument development, the assessment framework was used as the substantive basis for instrument construction.

The NAP–Science Literacy 2023 Assessment Framework provides historical information about the origin and development of the NAP–Science Literacy assessment. It describes the content to be assessed, the cognitive engagement that is expected of students, the types of assessment tasks, the contextual information collected, and the overall structure and purposes of the assessment.

The NAP–Science Literacy assessment framework

In 2003, the inaugural NAP–Science Literacy assessment was administered to a sample of Year 6 students across Australia. It was informed by a framework predating the establishment of the Australian Curriculum. Following the formulation and adoption of the national curriculum, efforts were made to align some NAP–Science Literacy items with the Australian Curriculum for the 2015 cycle. In 2017, further initiatives were undertaken to devise a comprehensive framework outlining specifications for both the Year 6 assessment and the introduction of a Year 10 assessment commencing in 2018. These endeavours also encompassed the transition to an online assessment platform and the integration of innovative science assessment strategies. The redeveloped framework provided the structural guidance for the implementation of the 2018 assessments.

For 2023, the updated NAP–Science Literacy framework maintains the foundational structure established in 2018 while incorporating refined specifications tailored to both Year 6 and Year 10 science literacy assessments. Drawing upon insights from the 2019 national Alice Springs (Mparntwe) Education Declaration on education goals for all Australians, the 2023 framework mirrors recent enhancements to the Foundation – Year 10 Australian Curriculum. This framework provides the basis for an effective measure of students' science literacy over time.

Defining science literacy

NAP–Science Literacy measures science literacy as defined in the Australian Curriculum: Science as: "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" (ACARA n.d.).

NAP–Science Literacy content dimension

The NAP–Science Literacy Assessment Framework organises the content domains and sub-domains according to the strands and sub-strands of the Australian Curriculum: Science, respectively. The content strands and sub-strands are:

1. **Science understanding**, which refers to the selection and integration of appropriate science knowledge to explain and predict phenomena, and to the application of that knowledge to new situations. Science knowledge refers to facts, concepts, principles, laws, theories and models that have been established over time.

a. **Biological sciences,** which is concerned with understanding living things including animals, plants and microorganisms, and their interdependence and interactions within ecosystems.

b. **Earth and space sciences,** which is concerned with Earth's dynamic structure and its place in the cosmos.

c. **Physical sciences,** which is concerned with understanding the nature of forces and motion, and matter and energy.

d. **Chemical sciences,** which is concerned with understanding the composition and behaviour of substances.

2. **Science as a Human Endeavour**, which refers to the nature of science, including the role of science inquiry in developing science knowledge, and the factors that affect the use and advancement of science.

   a. **Nature and development of science**, which refers to the unique nature of science and scientific knowledge, including that scientific knowledge is based on empirical evidence and can be modified in light of new or reinterpreted evidence.

   b. **Use and influence of science**, which explores how science knowledge and applications affect individuals and communities, including informing their decisions and identifying responses to contemporary issues.

3. **Science Inquiry**, which is concerned with the diverse ways that scientists study the natural world and propose explanations based on evidence (National Research Council, 2000).

   a. **Questioning and predicting**, which refers to identifying and constructing investigable questions, proposing hypotheses and predicting possible outcomes.

   b. **Planning and conducting,** which refers to making decisions about how to investigate or solve a problem, and carrying out an investigation.

   c. **Processing, modelling and analysing**, which refers to analysing and representing data in meaningful ways and identifying trends, patterns and relationships in data.

   d. **Evaluating,** which refers to considering the quality of available evidence and the merit or significance of a claim, proposition, explanation or argument with reference to that evidence.

   e. **Communicating,** which refers to conveying information or ideas to others in ways appropriate to the purpose and audience.

NAP–Science Literacy cognitive dimension

The NAP–Science Literacy Assessment Framework cognitive dimension describes the science-focused thinking skills students are expected to use as they respond to assessment tasks. It represents the cognitive processes required in the application of science concepts. The cognitive areas are:

1. **Knowing and using procedures**, which refers to knowledge of facts and definitions, the ability to illustrate scientific concepts by providing or identifying examples, knowing and being able to perform simple science processes or procedures.

2. **Reasoning, analysing and evaluating**, which refers to the ability of students to engage in applying knowledge, skills and processes, as well as the analysis and evaluation of information, evidence and arguments with respect to quality, relevance and sufficiency of data.

3. **Synthesising and creating**, which refers to the consideration of a number of different factors, variables or concepts to compile elements in new or different ways to form a coherent hypothesis, argument or explanation.

Critical and Creative Thinking

The general capability of Critical and Creative Thinking (CCT) is integrated into NAP–Science Literacy through the cognitive dimension of the NAP–Science Literacy Assessment Framework. Aspects of CCT arise from important cognitive skills inherent in scientific inquiry and in broader scientific thinking. The elements and sub-elements of the CCT learning continuum from the Australian Curriculum have guided the development of assessment tasks and reflect the thinking skills and intellectual processes students are expected to use as they respond to the assessment tasks.

Within the context of NAP–Science Literacy, CCT represents important ways of thinking that help students inquire into the world around them. Within the cognitive dimension of the NAP–Science Literacy Assessment Framework, critical thinking involves students analysing and assessing possibilities, constructing and evaluating arguments, and using information, evidence and logic to draw reasoned conclusions and to solve problems. Thinking creatively involves students generating new ideas, considering alternative explanations and possibilities, and transferring knowledge and skills to new and unfamiliar contexts.

NAP–Science Literacy and the Australian Curriculum

The NAP–Science Literacy items included in the 2023 assessment cycle are aligned with the Australian Curriculum strands and sub-strands, as described in the previous section. Where applicable, items are also:

- aligned with the general capabilities of the Australian Curriculum (including the CCT capability described previously)
- aligned with the cross-curriculum priorities including Sustainability and Aboriginal and Torres Strait Islander Histories and Cultures
- reflective of the key ideas of the Australian Curriculum: Science, which represent key aspects of a scientific view of the world, and bridge knowledge and understanding across the disciplines of science.

Assessment instrument

The NAP–Science Literacy 2023 assessment instrument was based on the design principles established in 2006, which continued through the assessment cycles in 2009, 2012, 2015 and 2018. As in previous cycles, the assessment was computer-based and included a broad range of task formats including multiple-choice, interactive match, hotspot and constructed text/numerical response items.

As outlined in the NAP–Science Literacy 2023 Assessment Framework, the assessment instrument aligns with both the organisation and content of the Australian Curriculum: Science. The instrument addresses a range of proficiency levels required for the effective measurement of scientific literacy across the curriculum. Table 2.1 shows the content domains, sub-domains and target percentages for the NAP–Science Literacy assessment.

Table 2.1: Target percentages for content domains and sub-domains in the Year 6 and Year 10 assessment

Content domain	Target percentage	Content sub-domain
Science Understanding	45%	Biological sciences
		Earth and space sciences
		Physical sciences
		Chemical sciences
Science as a Human Endeavour	15%	Nature and development of science
		Use and influence of science
Science Inquiry	40%	Questioning and predicting
		Planning and conducting
		Processing, modelling and analysing
		Evaluating
		Communicating

The assessment content was designed to be congruent with the previous cycles of NAP–Science Literacy with the content domains and target percentages used in 2023 being broadly consistent with those from previous cycles. This includes the use of common items as vertical links between Year 6 and Year 10. The use of trend (common) items across cycles furthermore allowed the 2023 results to be reported against the existing NAP–Science Literacy scale.

The NAP–Science Literacy instrument used a cluster rotation design where each test form was linked through common clusters to other forms. To achieve the rotation, the items were written in contextual units. Clusters were then constructed by grouping units together, and clusters were then grouped together to create test forms. In total, there were 36 test forms developed for 2023, with 18 Year 6 forms and 18 Year 10 forms. The assessment platform enforced a time limit of 60 minutes for Year 6 and 75 minutes for Year 10 students.

All test forms included one Inquiry task that was structured as a scientific investigation. Students were provided a context and components of the scientific method for a simulated investigation linked to the context, and then required to apply the results to the original context.

Table 2.2 shows the implemented NAP–Science Literacy instrument rotation design with clusters and inquiry tasks positioned across the test forms.

Table 2.2: Test instrument cluster rotation design

Year 6	A	B	C
Testform 1	C01	C07	Task1
Testform 2	C02	C08	Task5
Testform 3	C03	C09	Task3
Testform 4	C04	C10	Task4
Testform 5	C05	C11	Task2
Testform 6	C06	C12	Task6
Testform 7	C07	Task 6	C04
Testform 8	C08	Task1	C05
Testform 9	C09	Task2	C06
Testform 10	C10	Task3	C01
Testform 11	C11	Task4	C02
Testform 12	C12	Task5	C03
Testform 13	Task1	C06	C11
Testform 14	Task2	C01	C12
Testform 15	Task3	C02	C07
Testform 16	Task4	C03	C08
Testform 17	Task5	C04	C09
Testform 18	Task 6	C05	C10

Year 10	A	B	C
Testform 19	C03	C18	Task8
Testform 20	C05	C13	Task9
Testform 21	C07	C14	Task4
Testform 22	C09	C15	Task5
Testform 23	C01	C16	Task6
Testform 24	C11	C17	Task7
Testform 25	C13	Task8	C09
Testform 26	C14	Task9	C01
Testform 27	C15	Task4	C11
Testform 28	C16	Task5	C03
Testform 29	C17	Task6	C05
Testform 30	C18	Task7	C07
Testform 31	Task4	C01	C17
Testform 32	Task5	C11	C18
Testform 33	Task6	C03	C13
Testform 34	Task7	C05	C14
Testform 35	Task8	C07	C15
Testform 36	Task9	C09	C16

Questionnaire instrument

First introduced as the “student survey” in 2009, the NAP–Science Literacy student questionnaire was administered to all Year 6 and Year 10 students immediately following the assessment. Unlike the assessment, the questionnaire was not timed, with most students completing it in approximately 20 minutes.

The questionnaire collected information about students’ experiences, attitudes, values and engagement with various aspects of science literacy, with content belonging to one of 3 broad themes:

1. Science as a Human Endeavour
2. Teaching and learning in science
3. Student engagement with science.

Further, the questionnaire gathered information about Year 10 students' perceptions of the relevance of science for future study and career opportunities in fields related to science, technology, engineering and mathematics (STEM).

The incorporation of this contextual dimension allows us to examine the attitudinal and behavioural data of students while also measuring the extent to which certain factors are associated with variations in student achievement. The questionnaire responses were scaled to provide construct indicators of students' perception and engagement, with outcomes reported in the NAP–Science Literacy 2023 Public Report. This included analyses of the correlation with students' overall achievement in science literacy at the national, as well as state and territory levels.

A copy of the student questionnaire can be found in the appendices to this report (Appendix A).

Chapter 3: Sampling and weighting

Chapter 3: Sampling and weighting

This chapter describes the NAP–Science Literacy 2023 main study sample design, the achieved sample and the procedures used to calculate the sampling weights. The sampling and weighting methods were used to ensure that the data provided accurate and efficient estimates of the achievement outcomes for the Australian Year 6 and Year 10 student populations.

Information on the field trial sampling can be found in the field trial section of Chapter 4.

Sampling

The target populations for the study were Year 6 and Year 10 students enrolled in schools across Australia.

A 2-stage stratified cluster sample design was used in NAP–Science Literacy, similar to that used in other Australian national sample assessments and in international assessments such as the Trends in International Mathematics and Science Study (TIMSS). The first stage involved drawing a sample of schools. The sampling frame was explicitly stratified by state or territory and school sector, and separate, independent samples were drawn from each.

Schools were implicitly stratified within each explicit stratum, by the following variables:

- school type (primary, secondary, combined)
- school NAPLAN performance quintile
- a measure of school socio-economic status known as the Socio-Economic Index of Education and Occupation (SEIFA IEO)[4]
- school Australian Statistical Geography Standard (ASGS) remoteness class (Major Cities, Inner Regional, Outer Regional, Remote and Very Remote)[5]
- enrolment size at the target year level (either Year 6 or Year 10).

The second stage involved drawing a random sample of 20 students across the entire year level in sampled schools.

Up to 2 substitute schools were assigned to each sampled school at the time of sampling. Substitute schools were chosen to be as similar as possible to the sampled school with respect to the implicit stratification variables listed above. This enabled the sample size and representativeness to be maintained if a sampled school was unable to participate. To maintain the integrity of the original sample as much as possible, the use of substitute schools was kept to a minimum.

The sampling frame

Schools were selected from ACARA's Australian Schools List, a comprehensive list of all schools and campuses in Australia, comprising schools from all Australian states and territories, updated annually.

[4] This is a measure of the socio-economic status based on the socio-economic conditions, such as education and employment, of the geographic location of the school.
[5] This is a measure of geographic location of the school.

School exclusions

At the school level, exclusions from the target population included:

- schools that had participated in the NAP–Science Literacy field trial
- very remote schools[6] in all jurisdictions (except in the Northern Territory)
- schools listed on the ACARA Australian Schools List with fewer than 5 students in the target year level
- non-mainstream schools.[7]

At the time of sampling, these students accounted for 3.9% of the Year 6 student population and 5.6% of the Year 10 student population.

The decision to include very remote schools in the Northern Territory sample for 2023 was made because very remote schools comprised 25% of the Year 6 population and 19% of the Year 10 population in the Northern Territory, while this population was less than 1% of the total student population of Australia.

The designed sample

Sample sizes for both Year 6 and Year 10 were chosen to provide accurate estimates of achievement outcomes for all states and territories. The expected 95% confidence intervals were estimated in advance to be within approximately ±0.15 to ±0.2 of the population standard deviation for estimated means of the larger states. This level of precision was considered an appropriate balance between the analytical demands of the study, the burden on individual schools and the overall costs of the study. An effective sample size of around 100–150 students[8] is required to meet confidence intervals of this magnitude in the larger states. Smaller sample sizes were deemed as sufficient for the smaller states and territories because of their relatively small student populations. Table 3.1 shows the target populations and designed samples for each state and territory.

[6] Very remote schools are considered Australian Statistical Geography Standard (ASGS) Code = 4.

[7] Non-mainstream includes schools such as correctional schools, schools with a non-English curriculum (e.g. French immersion schools), language schools, special schools, schools for distance education (including Schools of the Air), hospital schools, short-term provision (e.g. environmental education support centre), mature age and preschools (all enrolments below year 0).

[8] The effective sample size is the sample size of a simple random sample that would produce the same precision as that achieved under a complex sample design.

Table 3.1: Year 6 and Year 10 target population and designed samples by state and territory

	Year 6			Year 10		
State/territory	**Enrolment**	**Schools in Population**	**Designed Sample**	**Enrolment**	**Schools in Population**	**Designed Sample**
NSW	97,534	2,102	55	89,425	797	60
VIC	77,154	1,687	55	70,990	554	55
QLD	67,514	1,170	55	61,004	469	39
SA	20,611	537	50	19,934	196	14
WA	33,056	739	50	29,742	245	40
TAS	6,594	197	43	6,327	83	6
NT	3,258	122	33	2,604	52	4
ACT	5,932	97	22	5,489	41	4
Aust.	311,653	6,651	363	285,515	2,437	222

First sampling stage

The sample design developed for the project was a stratified cluster sample. Prior to sampling, schools were explicitly stratified by state and sector. That is, separate samples were drawn for each sector[9] within states and territories for a total number of 24 explicit strata for Year 6. In Year 10 however, for ACT and NT, due to the low number of schools, both Catholic and independent schools were collapsed into one stratum, resulting in a total of 22 explicit strata. Schools within each stratum were ordered by school type, school NAPLAN performance quintile, SEIFA IEO, school Australian Statistical Geography Standard (ASGS) remoteness class, and enrolment size at the target year level. With systematic selection of the schools, these variables became implicit stratifiers.

The selection of schools was conducted using a systematic probability-proportional-to-size (PPS) method. For large schools, the measure of size (MOS) was equal to the enrolment at the target year level. To minimise variation in weights, the MOS for very small schools (between 5 and 9 students) was set to 10, and the MOS for small schools (between 10 and 19 students) was set to 20.

After sorting the sampling frame according to the stratification variables listed above, the standard process for the selection of schools with PPS was as follows:

- The MOS was accumulated from school to school and the running total was listed next to each school. The total cumulative MOS was a measure of the size of the population of sampling elements. Dividing this figure by the number of schools to be sampled provided the sampling interval.
- The first school was sampled by choosing a random number between one and the sampling interval. The school whose cumulative MOS contained the random number was the first sampled school. By adding the sampling interval to the random number, a second school was identified. This process of consistently adding the sampling interval to the previous selection number resulted in a PPS sample of the required size.

An analysis of small schools (schools with fewer enrolments than the assumed cluster sample size of 20 students) was undertaken prior to sampling. On the basis of this analysis, the school sample size in some strata was increased in order to ensure that the number of students sampled was close to expectations. As a result, after the small school analysis, the actual numbers of schools sampled for

[9] The 3 Australian school sectors are: government, Catholic and independent

Year 6 and Year 10 were 378 and 226, respectively. Both were slightly larger than the designed sample. The actual sample drawn is referred to as the “target sample”.

Table 3.2: Year 6 and Year 10 designed and target samples by state and territory

	Year 6		Year 10	
State/territory	**Designed Sample**	**Target Sample**	**Designed Sample**	**Target Sample**
NSW	55	57	60	60
VIC	55	57	55	56
QLD	55	56	39	39
SA	50	52	14	14
WA	50	51	40	40
TAS	43	45	6	8
NT	33	38	4	5
ACT	22	22	4	4
Aust.	363	378	222	226

As each school was selected, the next school in the sampling frame was designated as a substitute school to be included in cases where the sampled school did not participate. The adjacent school immediately before the sampled school was designated as the second substitute. It was used if neither the sampled nor the first substitute school participated. In some cases (such as primary schools in the Northern Territory), there were not enough schools available for 2 substitutes to be drawn. Due to the stratified sampling frame, the 2 substitute schools were similar (with respect to geographic location, socio-economic status, NAPLAN performance and size) to the originally sampled school for which they were assigned as a substitute.

After the school sample had been drawn, a number of sampled schools were identified as meeting the criteria for exclusion. When this occurred, the sampled school and its substitutes were removed from the sample and removed from the calculation of response rates. Three schools were removed each from the Year 6 and the Year 10 sample respectively. These exclusions are included in the exclusion rates reported earlier.

Second sampling stage

The second stage of sampling involved the systematic selection of 20 students within each participating school from a list of all eligible students at each target year level sorted by gender. This approach ensured that the distribution of students sampled by gender matched to the distribution at the school. If fewer than 20 eligible students were enrolled in the target year level (in smaller schools, for instance), *all* students in the year level were selected to participate.

Student exclusions

In each of the sampled schools, individual students were exempted from the assessment if they met any one of the following criteria:

- Severe functional disability: the student had a moderate to severe permanent physical disability such that they could not be expected to perform in the assessment situation.
- Severe intellectual disability: the student had a mental or emotional disability and/or cognitive delay such that they could not be expected to perform in the assessment situation.

- Very limited assessment language proficiency: the student was unable to read or speak the language of the assessment (English) and would not be expected to overcome the language barrier in the assessment situation. Typically, a student who had received less than one year of instruction in English would be exempted.

Table 3.3 and Table 3.4 detail the numbers and percentages of students excluded from the NAP–Science Literacy assessment, according to the reason given for their exclusion. The number of student-level exclusions was 170 at Year 6 and 125 at Year 10. This gives weighted exclusion rates of 2.2% of the sampled Year 6 students and 2.5% of sampled Year 10 students.

Table 3.3: Year 6 breakdown of student exclusions according to reason by state and territory

	Student Exclusion				
State/territory	**Functional Disability**	**Intellectual Disability**	**Limited Language Proficiency**	**Total**	**Weighted Proportion of Sampled Students in Year 6**
NSW	5	4	7	16	1.4%
VIC	8	11	2	21	1.6%
QLD	14	13	9	36	3.2%
SA	9	10	3	22	1.8%
WA	11	8	5	24	2.1%
TAS	11	13	2	26	3.1%
NT	2	5	4	11	1.5%
ACT	5	6	3	14	3.1%
Aust.	65	70	35	170	2.2%

Table 3.4: Year 10 breakdown of student exclusions according to reason by state and territory

	Student Exclusion				
State/territory	**Functional Disability**	**Intellectual Disability**	**Limited Language Proficiency**	**Total**	**Weighted Proportion of Sampled Students in Year 10**
NSW	14	11	5	30	2.2%
VIC	20	10	1	31	2.4%
QLD	20	3	8	31	3.2%
SA	2	2	1	5	2.0%
WA	6	6	3	15	2.3%
TAS	6	1	0	7	3.1%
NT	1	2	0	3	2.5%
ACT	1	1	1	3	4.0%
Aust.	70	36	19	125	2.5%

Weighting

While the multi-stage stratified cluster design provides a very economical and effective data collection process in a school environment, oversampling of sub-populations and non-response cause differential probabilities of selection for the ultimate sampling elements, the students. Consequently, one student in the assessment does not necessarily represent the same number of students in the population as another, as would be the case with a simple random sampling approach. To account for differential probabilities of selection due to the design and to ensure unbiased population estimates, a sampling weight was computed for each participating student. It was an essential characteristic of the sample design to allow the provision of proper sampling weights, since these were necessary for the computation of accurate population estimates.

The overall sampling weight is the product of weights calculated at the 2 stages of sampling:

1. the selection of the school at the first stage
2. the selection of students within the sampled schools at the second stage.

First-stage weight

The first-stage weight is the inverse of the probability of selection of the school, adjusted to account for school non-response within each explicit stratum.

The probability of selection of the school is equal to its measure of size (MOS) divided by the sampling interval (SINT), or one, whichever is lower. A school with a MOS greater than the SINT has a certain probability of selection and therefore has a probability of one.

The sampling interval is calculated at the time of sampling, and for each explicit stratum it is equal to the cumulative MOS of all schools in the stratum, divided by the number of schools to be sampled from that stratum.

The first factor of the first-stage weight, or the school base weight (BW_{sc}), was the inverse of this probability:

$$BW_{sc} = \frac{SINT}{MOS}$$

Following data collection, counts of the following categories of schools were made for each explicit stratum:

- the number of schools that participated (n_p^{sc})
- the number of schools that were sampled but should have been excluded (n_x^{sc})
- the number of non-responding schools (n_n^{sc}).

Note that $n_p^{sc} + n_x^{sc} + n_n^{sc}$ equals the total number of sampled schools from the stratum.

Examples of the second category (n_x^{sc}) were:

- a sampled school that no longer exists
- a school that, following sampling, was discovered to fit one of the criteria for school-level exclusion (e.g. very remote, very small), but which had not been removed from the frame prior to sampling.

In the case of the non-responding schools (n_n^{sc}), neither the originally sampled school nor its substitutes participated. Schools with a student response rate of less than 25% were also considered to be non-responding schools.

Within each explicit stratum, an adjustment was made to account for school non-response. This non-response adjustment (ASC) for a stratum was equal to:

$$ASC_{strt} = \frac{\left(n_p^{sc} + n_n^{sc}\right)}{n_p^{sc}}$$

The first-stage weight, or the final school weight, was the product of the base weight of the school and the school non-response adjustment:

$$FW_{sc} = BW_{sc} \times ASC_{strt}$$

Second-stage weight

Following data collection, counts of the following categories of students were made for each sampled school:

- the number of students at the relevant year level (n_{tot}^{st})
- the number of students who participated (n_p^{st})
- the number of sampled students who were exclusions (n_x^{st})
- the number of non-responding sampled students (n_n^{st}).

Note that $n_{samp}^{st} = n_p^{st} + n_x^{st} + n_n^{st}$ equals the total number of sampled students from the sampled school.

The first factor in the second-stage weight was the inverse of the probability of selection of the student from the sampled school.

$$BW_{st} = \frac{n_{tot}^{st}}{n_{samp}^{st}}$$

The student-level non-response adjustment was calculated for each school as:

$$AST_{sc} = \frac{\left(n_p^{st} + n_n^{st}\right)}{n_p^{st}}$$

The final student weight was:

$$FW_{st} = BW_{st} \times AST_{sc}$$

Overall sampling weight

The overall sampling weight (FWTOT) was simply the product of the weights calculated at each of the 2 sampling stages:

$$FWTOT = FW_{sc} \times FW_{st}$$

After computation of the overall sampling weights, the weights were checked for outliers that would have a large effect on the computation of the standard errors. A weight was regarded as an outlier if the value was more than 4 times the median weight within an explicit stratum. Weights exceeding this threshold were trimmed to 4 times the median weight. The final, trimmed weight was:

$$WT2023 = FWTOT_{trimmed}$$

Response rates

Separate response rates were computed:

1. with substitute schools included as participants[10]
2. with substitute schools regarded as non-respondents.

In addition, each of these rates was computed using unweighted and weighted counts. Regardless of the method used, school and student response rates were computed, and the overall response rate was the product of these 2 response rates. The differences in computing the 4 response rates are described below. These methods are consistent with the methodology used in TIMSS (Olson, Martin and Mullis 2013).

Unweighted response rates including substitute schools

The unweighted school response rate, where substitute schools were counted as participating schools, was computed as follows:

$$RR_1^{sc} = \frac{n_s^{sc} + n_{r1}^{sc} + n_{r2}^{sc}}{n_s^{sc} + n_{r1}^{sc} + n_{r2}^{sc} + n_{nr}^{sc}}$$

where n_s^{sc} is the number of responding schools from the original sample, $n_{r1}^{sc} + n_{r2}^{sc}$ is the total number of responding substitute schools and n_{nr}^{sc} is the number of non-responding schools that could not be replaced.

The student response rate was computed over all responding schools. Of these schools, the number of responding students was divided by the total number of eligible, sampled students:

$$RR_1^{st} = \frac{n_s^{st} + n_{r1}^{st} + n_{r2}^{st}}{n_s^{st} + n_{r1}^{st} + n_{r2}^{st} + n_{nr}^{st}}$$

where n_s^{st} is the total number of responding students in sampled schools, $n_{r1}^{st} + n_{r2}^{st}$ is the total number of responding students in substitute schools and n_{nr}^{st} is the total number of eligible, non-responding, sampled students in all participating schools.

The overall response rate is the product of the school and the student response rates.

$$RR_1 = RR_1^{sc} \times RR_1^{st}$$

Unweighted response rates excluding substitute schools

The difference of the second method from the first is that the substitute schools were counted as non-responding schools.

$$RR_2^{sc} = \frac{n_s^{sc}}{n_s^{sc} + n_{r1}^{sc} + n_{r2}^{sc} + n_{nr}^{sc}}$$

This difference had an indirect effect on the student response rate because fewer schools were included as responding schools, and student response rates were only computed for the responding schools.

$$RR_2^{st} = \frac{n_s^{st}}{n_s^{st} + n_{r1}^{st} + n_{r2}^{st} + n_{nr}^{st}}$$

[10] A school is considered to be participating if it has a student response rate of at least 50%. Schools with less than 50% response rate and students within schools with less than 50% response rate are given a weight of zero for response rate calculations.

The overall response rate was again the product of the 2 response rates.

$RR_2 = RR_2^{sc} \times RR_2^{st}$

Weighted response rates including substitute schools

For the weighted response rates, sums of weights were used instead of counts of schools and students. School and student base weights (BW) are the weight values before correcting for non-participation, so they generate estimates of the population being represented by the responding schools and students. The full weights (FW) at the school and student levels are the base weights corrected for non-participation.

School response rates are computed as follows:

$$RR_3^{sc} = \frac{\sum_i^{s+r1+r2}\left(BW_i \times \sum_j^{r_i}(FW_{ij})\right)}{\sum_i^{s+r1+r2}\left(FW_i \times \sum_j^{r_i}(FW_{ij})\right)}$$

where i indicates a school, $s + r1 + r2$ all responding schools, j a student and ri the responding students in school *i*. First, the sum of the student final weights FW_{ij} for the responding students from each school was computed. Second, this sum was multiplied by the school's base weight (numerator) or the school's final weight (denominator). Third, these products were summed over the responding schools (including substitute schools). Finally, the ratio of these values was the response rate.

The numerator of the school response rate (RR_3^{sc}) is the denominator of the student response rate (RR_3^{st}):

$$RR_3^{st} = \frac{\sum_i^{s+r1+r2}\left(BW_i \times \sum_j^{r_i}(BW_{ij})\right)}{\sum_i^{s+r1+r2}\left(BW_i \times \sum_j^{r_i}(FW_{ij})\right)}$$

The overall response rate is the product of the school and student response rates:

$RR_3 = RR_3^{sc} \times RR_3^{st}$

Weighted response rates excluding substitute schools

Practically, substitute schools were excluded by setting their school base weight to zero for computation of the school response rates and applying the same computation as above. More formally, the parts of the response rates are computed as follows:

$$RR_4^{sc} = \frac{\sum_i^{s}\left(BW_i \times \sum_j^{r_i}(FW_{ij})\right)}{\sum_i^{s+r1+r2}\left(FW_i \times \sum_j^{r_i}(FW_{ij})\right)}$$

$$RR_4^{st} = \frac{\sum_i^{s+r1+r2}\left(BW_i \times \sum_j^{r_i}(BW_{ij})\right)}{\sum_i^{s+r1+r2}\left(BW_i \times \sum_j^{r_i}(FW_{ij})\right)}$$

$RR_4 = RR_4^{sc} \times RR_4^{st}$

Reported response rates

In terms of the coverage of the sampled population, weighted response rates are a more accurate indicator of the representativeness of the sample. For the 2023 cycle, the weighted national school response rate in Year 6 was 88% when including substitute schools and 88% when excluding substitute schools. In Year 10, the respective percentages were 82% and 81%.

Overall unweighted response rates for Year 6 were 86% when including substitute schools and 86% when excluding substitute schools. Overall unweighted response rates for Year 10 were 78% when including substitute schools and 78% when excluding substitute schools.

Table 3.5: Overall school and student response rates in Year 6

	Unweighted, including substitute schools			Unweighted, sampled schools only			Weighted, including substitute schools			Weighted, sampled schools only		
State/territory	**Overall**	**School**	**Student**	**Overall**	**School**	**Student**	**Overall**	**School**	**Student**	**Overall**	**School**	**Student**
NSW	0.89	1.00	0.89	0.89	1.00	0.89	0.87	1.00	0.87	0.87	1.00	0.87
VIC	0.88	0.96	0.91	0.88	0.96	0.91	0.88	0.96	0.91	0.88	0.96	0.91
QLD	0.89	1.00	0.89	0.89	1.00	0.89	0.89	1.00	0.89	0.89	1.00	0.89
SA	0.91	1.00	0.91	0.89	0.98	0.91	0.91	1.00	0.91	0.90	0.99	0.91
WA	0.86	0.98	0.88	0.83	0.94	0.88	0.88	1.00	0.88	0.86	0.97	0.88
TAS	0.88	1.00	0.88	0.88	1.00	0.88	0.88	1.00	0.88	0.88	1.00	0.88
NT	0.70	0.86	0.81	0.70	0.86	0.81	0.77	0.89	0.86	0.77	0.89	0.86
ACT	0.85	0.95	0.89	0.85	0.95	0.89	0.85	0.95	0.89	0.85	0.95	0.89
Aust.	0.86	0.98	0.89	0.86	0.97	0.89	0.88	0.99	0.89	0.88	0.99	0.89

Table 3.6: Overall school and student response rates in Year 10

	Unweighted, including substitute schools			Unweighted, sampled schools only			Weighted, including substitute schools			Weighted, sampled schools only		
State/territory	**Overall**	**School**	**Student**	**Overall**	**School**	**Student**	**Overall**	**School**	**Student**	**Overall**	**School**	**Student**
NSW	0.79	0.97	0.82	0.79	0.97	0.82	0.82	0.98	0.84	0.82	0.98	0.84
VIC	0.74	0.93	0.80	0.73	0.91	0.80	0.83	1.00	0.83	0.80	0.97	0.83
QLD	0.79	0.97	0.81	0.79	0.97	0.81	0.79	0.97	0.81	0.79	0.97	0.81
SA	0.78	0.95	0.83	0.78	0.95	0.83	0.84	1.00	0.84	0.84	1.00	0.84
WA	0.85	1.00	0.85	0.85	1.00	0.85	0.86	1.00	0.86	0.86	1.00	0.86
TAS	0.82	1.00	0.82	0.82	1.00	0.82	0.80	1.00	0.80	0.80	1.00	0.80
NT	0.86	1.00	0.86	0.86	1.00	0.86	0.87	1.00	0.87	0.87	1.00	0.87
ACT	0.83	1.00	0.83	0.83	1.00	0.83	0.83	1.00	0.83	0.83	1.00	0.83
Aust.	0.78	0.96	0.82	0.78	0.96	0.82	0.82	0.99	0.83	0.81	0.98	0.83

Chapter 4: Data collection, management and processing

Chapter 4: Data collection, management and processing

The collection and processing of cognitive, contextual and administrative data for NAP–Science Literacy is supported by a framework of high-quality and well-organised data management procedures. These procedures have been developed and refined by the Australian Council for Educational Research (ACER) over the course of many NAP sample cycles to ensure the integrity and quality of the data, while also minimising the administrative burden on participating schools.

This chapter outlines the data management procedures implemented for NAP–Science Literacy 2023. This includes the various methods of data collection that were employed before, during and after the administration of the assessment, as well as the procedures applied in the transfer, tracking, verification and transformation of the data collected.

Data management plan

ACER creates a detailed data management plan for the collection, transfer, processing and storage of data for NAP sample projects. Established plans and associated processes formed the basis for the NAP–Science Literacy 2023 cycle data management plan. The plan firstly identified the data elements, or information assets, that were relevant to NAP–Science Literacy. It then detailed where each of the information assets were stored, and described how they were to be secured over the life of the project. This plan was referred to and, where necessary, updated over the course of the project so that it would accurately describe the most current data management practices implemented by the project team.

Data security

ACER is extremely aware of the importance schools, educational authorities and wider society rightly place on the security of personal data. In the context of collecting, transferring and storing school- and student-level data, it is important to ensure that all systems, staff and processes are handling those information assets securely for the life of the project. Given that many of the NAP–Science Literacy information assets contain a level of personally identifiable data of Australian school children, all assets were marked as protected in accordance with both ACER's Data Classification Policy and its Cryptographic Policy.

ACER therefore implemented an Information Security Management System that is compliant with:

- ISO 27001:2013 Information technology – Security techniques – Information security management systems – Requirements
- ISO 27002:2015 Information technology – Security techniques – Code of practice for information security controls.

ACER's Information Security Management System also complies with:

- The Australian Government Information Security Manual (ISM) produced by the Australian Signals Directorate
- The Australian Government Protective Security Policy Framework.

ACER ensured that all the processes, systems and solutions used to support and implement the NAP–Science Literacy 2023 study complied with our Information Security Management System. This ensures that ACER systems, staff and processes are securely handling information assets.

Data identification

To track and monitor data throughout the life of the NAP–Science Literacy project, a system of identification (ID) codes was implemented. At the school level, a unique ID was created for each school at the time of sampling. This school ID was 7 digits in length and comprised a concatenation of codes relating to year level, state, sector, substitution status as well as a unique sequential number. The specific codes used for each variable are outlined in Figure 4.1.

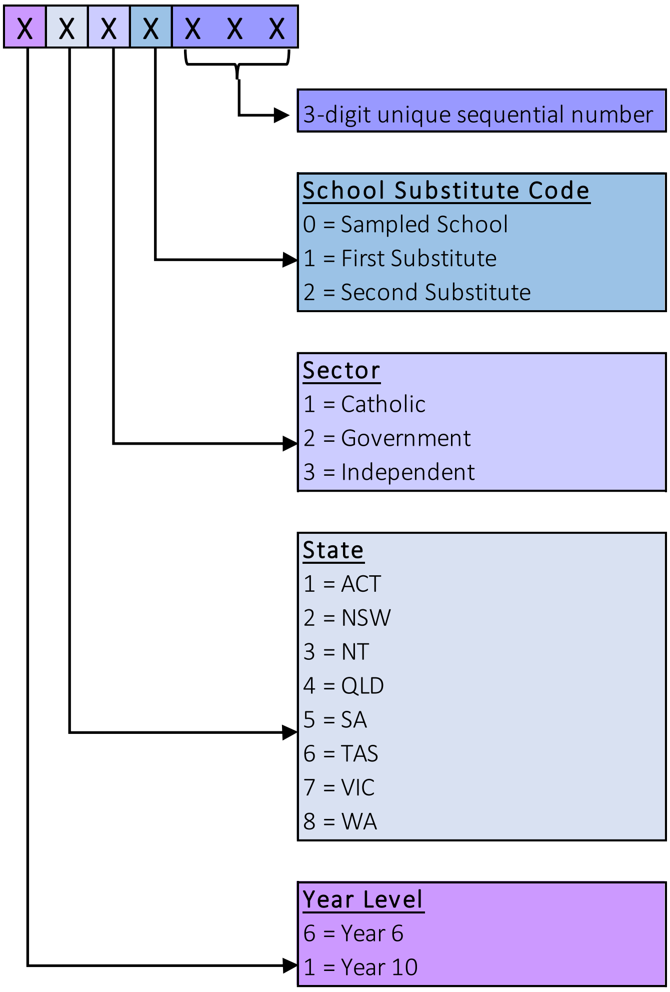

Figure 4.1: Breakdown of codes used in unique school ID

At a student level, an ID was created that comprised the 7-digit school ID followed by a 2-digit student number (01–20) that was unique to each sampled student within the school. This student ID was included in the student cognitive, contextual and student background data files so that data could be accurately matched and tracked throughout the data capture, cleaning and analysis stages. Five spare IDs were created for each school and were distributed if additional test login credentials were required. The spare ID comprised the 7-digit school ID followed by a 2-digit student number (21–25). The use of unique student IDs allowed for NAP–Science Literacy data to be viewed and analysed without the use of personally identifiable data (i.e. student names).

Data collected from schools and jurisdictions

The administration of the NAP–Science Literacy 2023 assessment required several stages of contact with the sampled schools to request or provide information. The contribution of both educational authorities and school staff to the data collection process is an essential part of the field administration.

To ensure the participation of sampled schools, education authority liaison officers were appointed for each jurisdiction. The liaison officers were expected to facilitate communication between ACER and the selected schools from their respective jurisdictions. The liaison officers helped to achieve a high participation rate for the assessment, which in turn helped to ensure unbiased, valid and reliable data.

Key personnel at each of the schools were nominated by the principal so that administrative and technical information could be collected in a timely manner. The roles of these nominated school personnel are outlined below:

- **The School Contact Officer (SCO):** The SCO was the main point of contact for ACER at the school and was responsible for coordinating and overseeing the assessment. SCOs provided ACER with information about the school's preferred assessment dates, student cohort lists and, if this could not be provided by the jurisdiction, student background data (SBD) for the selected students.
- **The School Technical Support Officer (STSO):** The STSO was responsible for ensuring that the school's computer system was test-ready by the scheduled assessment date. Primarily, the role involved conducting a series of technical checks on a sample of computers that were to be used for the assessment and helping to troubleshoot any issues ahead of assessment day.
- **The Test Administrator (TA):** The TA was responsible for administering the assessment to the sampled students, according to the standardised administration procedures provided in the TA Handbook. The SCO at the school would often also perform the duties of TA, though they could alternatively choose to nominate another staff member for this role.

An overview of the school liaison and data collection processes is provided in Table 4.1.

Table 4.1: School liaison and data collection processes

Stage	Jurisdictional activity	ACER project team activity	School activity
1	Educational authorities inform sampled schools of their selection in the assessment. If the jurisdiction confirms that a sampled school is unable to participate, the relevant substitute school is contacted.	ACER contacts principals of sampled schools to request the nomination of a SCO and STSO.	Principals of contacted schools supply requested contact information via a secure online form.
2		ACER contacts nominated SCOs and requests preferred assessment dates and student lists for target year level (Year 6 and/or Year 10 cohort).	SCOs submit preferred assessment dates and student list via a secure school administration website.

Stage	Jurisdictional activity	ACER project team activity	School activity
3		ACER contacts nominated STSOs and provides technical check instructions. ACER provides technical support and troubleshooting advice to STSOs via the Helpdesk.	STSOs undertake technical checks to ensure the school's computer resources are test-ready.
4		ACER notifies SCOs of finalised assessment date and selected students via the school administration website.	SCOs make relevant school-level test day arrangements (including room bookings and informing sampled students of their selection).
5	Educational authorities provide SBD for students in schools for which this information is held centrally.	Where SBD cannot be provided by the jurisdiction, ACER requests this information from SCOs for all sampled students.	SCOs provide SBD for all sampled students via the school administration website.
6		ACER provides detailed test administration manual and test login credentials to all nominated Test Administrators. ACER continues to provide support to schools via the Helpdesk.	Test Administrators familiarise themselves with the processes and procedures outlined in the test administration manual and consult with ACER Helpdesk staff to confirm understanding of protocol and circumvent any perceived issues prior to the scheduled assessment date.

The NAP–Science Literacy online school administration website

All information provided by SCOs to ACER was submitted via a secure website. The benefits of the NAP–Science Literacy school administration website were twofold: it eased the administrative burden on the selected schools, and provided a convenient, intuitive and secure repository for all school data relating to the study.

Schools were able to download all relevant administrative materials from this site, as well as use it to provide information to ACER regarding SCO details, assessment date preferences and student-related information as required. To access the website, SCOs needed to create a secure password and activate their school-specific account. Figure 4.2 shows a screenshot from the homepage of the website.

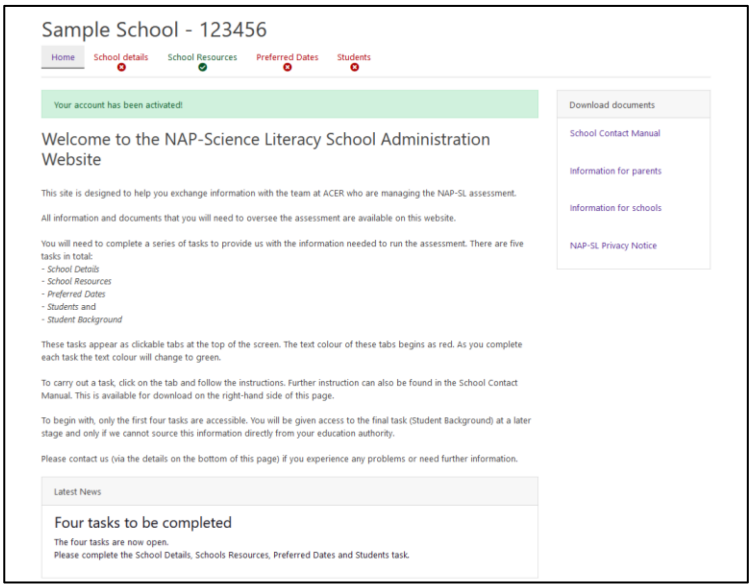

Figure 4.2: NAP–Science Literacy 2023 school administration website

The STSO technical checks

To ensure the smooth running of the assessment, STSOs needed to perform a series of technical checks on the devices selected for use. These checks consisted of a device check run through the test delivery system that checked the compatibility of the schools' devices, and a feedback questionnaire to report the results. An excerpt from the STSO manual, containing the device check instructions and steps, is provided in Appendix B.

After the technical checks were performed, the ACER project team would liaise with the STSOs who had reported issues. Technical issues were resolved through a process of troubleshooting with the ACER project team. This sometimes involved referring the matter to the test delivery system engineers or, in the case of access/security protocols, to the relevant central education authority of the applicable school.

Helpdesk provision and online support

A 1800 helpdesk support number and a dedicated email address were made available to schools for the entire main study administration phase (February – June 2023). Using these means, the ACER project team supported schools through all administrative, technical and operational tasks related to the administration of the NAP–Science Literacy assessment. Project staff were also on hand to provide any urgent assistance required during, or immediately preceding, the assessment session itself.

The helpdesk hours of operation during the assessment window were 8am–6pm AEST so that school hours across Australia's various time zones could be accommodated.

Collection of student background information

As per NAP protocol, student background data were collected for all participating students and matched to students' cognitive assessment and questionnaire responses for analysis and reporting purposes.

The structure of these student background variables follows NAP protocols as set out in the Data Standards Manual (ACARA 2022). The information collected included:

- age
- gender
- Indigenous status
- parental school education
- parental non-school education
- parental occupation
- main language spoken at home.

Schools are required to collect this information from the time of student enrolment. For NAP–Science Literacy 2023, student background data were collected in one of 2 ways: from the education authorities in each jurisdiction or from the schools themselves. Where possible, education authorities from each jurisdiction supplied these data directly to ACER so that schools were not unnecessarily burdened with this administrative task. Provision of student background data from education authorities occurred in 14 out of 24 of the jurisdictions across the country. The source of student background data for each of the jurisdictions is outlined in Table 4.2.

Table 4.2: Student background data provision

State/Territory	Sector	Source
NSW	Government	NSW DET
	Catholic	School
	Independent	School
VIC	Government	VIC DET
	Catholic	School
	Independent	School
QLD	Government	QLD DETE
	Catholic	School
	Independent	School
SA	Government	SA DECD
	Catholic	SA CEO
	Independent	School
WA	Government	WA DET
	Catholic	WA DET
	Independent	WA DET

State/Territory	Sector	Source
TAS	Government	Tas DoE
	Catholic	Tas CEO
	Independent	School
NT	Government	NT DET
	Catholic	School
	Independent	School
ACT	Government	ACT DET
	Catholic	ACT DET
	Independent	ACT DET

Where data collection from educational authorities was not possible, ACER collected this information from the schools themselves. To do this, the ACER project team created a template into which schools could enter the coded background details for each sampled student. This template was then uploaded by each school onto the secure NAP–Science Literacy school administration website. The code list for the student background data collected is presented in Table 4.3.

Table 4.3: Variable definitions for student background data

Category	Description	Codes
Gender	Gender of student	F = Female M = Male O = Other 9 = Not stated/unknown
Age	Date of birth of student	Free response DD-MM-YYYY
Indigenous status	A student is considered to be Indigenous if they identify as being of Aboriginal and/or Torres Strait Islander origin.	1 = Aboriginal but not Torres Strait Islander origin 2 = Torres Strait Islander but not Aboriginal origin 3 = Both Aboriginal and Torres Strait Islander origin 4 = Neither Aboriginal nor Torres Strait Islander origin 9 = Not stated/unknown.
Parental school education	The highest year of primary or secondary education a parent/guardian has completed.	1 = Year 9 or equivalent or below 2 = Year 10 or equivalent 3 = Year 11 or equivalent 4 = Year 12 or equivalent 0 = Not stated/unknown/Does not have Parent 2.
Parental non-school education	The highest qualification attained by a parent/guardian in any area of study other than school education.	5 = Certificate I to IV (including Trade Certificate) 6 = Advanced diploma/Diploma 7 = Bachelor's degree or above 8 = No non-school qualification 0 = Not stated/unknown/Does not have Parent 2.
Parental occupation	The occupation group that includes the main work undertaken by the parent/guardian.	1 = Senior management and professionals 2 = Other manager and associate professionals 3 = Tradespeople & skilled office, sales and service staff 4 = Machine operators, labourers, hospitality, and related staff 8 = Not in paid work in last 12 months 9 = Not stated/unknown/Does not have Parent 2.
Student/Parent language spoken at home	The main language spoken in the home by the respondent.	1201 = English Codes for all other languages as per the Australian Standard Classification of Languages (ASCL) Coding Index 2nd Edition

The ability of the ACER project team to collect student background data to the level required for data analysis purposes depends on how complete the records are kept at participating schools and central authorities. Where data variables were labelled as unknown or left blank by the school or jurisdiction, and the absence of data was confirmed upon follow up from the project team, these values were coded as missing. The percentage of missing values for the derived background data variables, along with the percentages for all valid codes, are presented in the NAP–Science Literacy 2023 Public Report.

Assessment administration

Field trial

The NAP–Science Literacy field trial was conducted from 17 October to 4 November 2022. In total, 2,866 Year 6 and Year 10 students from 128 schools across New South Wales, Victoria, Queensland, Western Australia and South Australia participated. The sample included students from major cities, and regional and remote areas. The students also came from a range of socio-economic backgrounds, and included a mix of government, Catholic and independent schools.

The purpose of the field trial was to test the assessment instruments and associated operational procedures. Thirty test forms were rotated across the participating students so that item-level coverage was evenly distributed.

Overall, the analysis of the collected data suggested that the test instrument, scoring guides and scoring procedures had been successful and would form a solid foundation for the 2023 main study. As a result of the findings, decisions were made as to which items would be used in the main study assessment instrument. The coverage and content of the assessment instrument is described in the following section.

Main study

The NAP–Science Literacy 2023 main study assessment was conducted across Australia during Term 2. Schools were permitted to schedule the assessment on a day that suited them within the official assessment period. The scheduled assessment window for all states and territories was Monday 8 May to Friday 26 May 2023.

During the testing period, the assessment window was extended from Friday 26 May to Wednesday 31 May to accommodate the testing of additional students.

The NAP–Science Literacy assessment consisted of a set of practice questions, the test and a student questionnaire. All components were to be administered on the same day. Schools were asked to allow approximately 2 hours for the entire assessment process to cater for settling the students, providing instruction and logging students into the online assessment platform.

Assessment platform

The 2023 cycle of the NAP–Science Literacy assessment was delivered exclusively via the Online National Assessment Platform developed to deliver NAPLAN Online and other NAP assessment events. The platform is managed by Educational Services Australia. As all the assessment and questionnaire data were collected electronically, scanning and manual data entry of student responses were not required.

Test session timing

The test administration times were designed to minimise the disruption of teaching and classroom patterns. Table 4.4 shows the timing of the test session.

Table 4.4: The timing of the assessment session

Component	Year 6	Year 10
Practice questions	10 mins (approx.)	10 mins (approx.)
Assessment	60 mins	75 mins
Student questionnaire	20 mins (approx.)	20 mins (approx.)

Follow-up test sessions

To maximise student participation, schools were asked to administer follow-up sessions in cases where a significant proportion (i.e. more than 20%) of students were absent on the scheduled assessment day. This helped ensure a minimum student participation rate of 80% at most participating schools.

Quality monitor visits

In line with quality assurance processes, ACER sent trained quality monitors to 5% of participating schools nationally. In total, 20 quality monitors were hired, who together visited 35 schools across all states and territories in Australia. The responsibility of the quality monitor was to ensure the uniformity and consistency of test administration procedures implemented across all participating schools. This was done by observing the Test Administrator before and during the administration of the assessment. The quality monitor then reported back to ACER via the online submission of a detailed, structured report. The quality monitor report template is provided in Appendix C.

Scoring student responses

Students completed the NAP–Science Literacy assessment using software that included a combination of different item types or formats. Student responses were either scored automatically by the testing system or scored during the later marking operation by a team of trained markers using a detailed scoring guide. The different formats and item types are described in Table 4.5 below.

Table 4.5: NAP–Science Literacy response formats and item types

Format	Item type	Item use in NAP–Science Literacy
Selected response format	Multiple-choice	• Students must select one of 4 options. • Options can be in word, graphical or pictorial format.
	Multiple-choices	• Students must select multiple options from a total of 5 or more options (e.g. “select all that apply”).
	Two-tier multiple-choice	• Students must select an option for a prediction, explanation, etc. and then select from a different set of options to justify reasoning.
	Interactive match (drag and drop)	• Students must select, drag and drop words, graphical or pictorial elements for classification purposes or to place items in order.
	Interactive match (draw lines)	• Students must connect 2 columns of options by drawing lines from an option in one column to an option in the second column. • Options can be images, numbers, words or descriptions.
	Interactive match (checkbox)	• Students must select a checkbox from columns within a table. • Multiple responses may be required to what is often a dichotomous “yes/no” type question.

Format	Item type	Item use in NAP–Science Literacy
	Interactive gap match	• Students must select words from multiple options to insert at various points in a sentence or passage.
	Hotspot	• Students must select one or more predefined areas on a diagram, graph or other image.
	Composite (inline choice)	• Students must select an answer from a drop-down menu. • Drop-down options are usually numbers, single words or short sentence fragments of 2 to 3 words. • An item may contain several inline choices where multiple responses are required.
	Composite (multiple interactions)	• Students must make 2 or more interactions from the item types listed above, where there are related concepts that constitute parts of a whole.
Constructed response formats	Short constructed	• Students must use one or 2 words, a phrase or numerical response.
	Single numerical	• Students must enter a single numerical answer in a text box, including setting values for input variables in simulations.
	Extended constructed	• Students must write between one sentence and several paragraphs of text. • This is particularly useful for probing students' deeper understanding and assessing higher proficiency levels.

Constructed response items

Some items required students to provide a typed response. These responses were captured by the test delivery system and later delivered to markers using a purpose-built online scoring system. Some of these items had scoring guides that allowed for dichotomous scoring (correct/incorrect), whereas others had scoring guides with partial credit scoring in which different categories of student responses could be scored according to the degree of knowledge, skill or understanding they demonstrated.

Centre-based marking operation

For the items that could not be autoscored by the test delivery system, responses were marked by a team of trained markers in a centre-based marking operation. The main study marking operation was conducted in the ACER Sydney Marking Centre from Monday 29 May to Tuesday 13 June 2023. Marking was conducted online using the ACER Marking System (AMS).

ACER employed a total of 21 markers, and 2 marking supervisors, to mark the 69 items needing to be human scored. These individuals were chosen from ACER's pool of highly experienced markers, many of whom had marked previous cycles of the NAP–Science Literacy field trial and main study. The marking design used for the operation of the marking centre involved a single team of day markers and 2 evening marking teams. Each group was closely supervised by a group leader who was responsible for check-marking and the maintenance of marking consistency across the group. The day team were experienced NAP sample markers, while the evening team comprised 13 practising Science teachers. The evening markers scored the more scientific and demanding items.

As per previous NAP sample marking operations, ACER utilised an ongoing training model (train-mark, train-mark) over the entire duration of the operation. This means that training in each item was conducted directly before marking of that particular item began, so that the rubric and construct were fresh in the minds of the markers as they began to mark an item.

The training approach included the completion of carefully selected sample responses that exemplified the nuances of the rubric, with time assigned for marker discussion and clarification of any questions prior to the commencement of marking.

Quality assurance during the marking process

Part of the role of the group leaders was to backread (spot check) a random sample of at least 10% of all responses scored by markers. Very few instances of off-task marking were observed, although in each group there were instances in which some retraining and remarking of individual items occurred as a result of interactions with student responses that evidenced answers not anticipated by the marking guide.

Control scripts for each item were pre-selected and added into the system for the markers to score as part of their allocated packet of responses. Very high compliance rates were observed on all control scripts, which again denotes a high level of marker accuracy.

Group leaders also referred to score distribution reports to ensure consistency in scoring patterns across the team during the scoring of each item.

Data cleaning and verification

Data cleaning and verification relate to processes of ensuring the integrity of the data collected. For NAP–Science Literacy, a series of data cleaning steps were undertaken on all data collected from jurisdictions, schools and students. With respect to student background data, the following steps were performed:

- Student names (for the purposes of school reporting) were corrected where there was obvious first name/surname reversal, or where foreign characters (e.g. ?, !, %) were included. Some instances of correction had to be confirmed with the school directly.
- Missing gender of the student was attributed where it could be inferred from the school type (e.g. where single-sex). Some instances of correction had to be confirmed with the school directly.
- All dates of birth were converted to the standard dd/mm/yyyy format, and any auto-formatting executed by the spreadsheet template that rendered dates of birth illegible was reversed and corrected.
- Any free text or abbreviated text was coded as per the variable coding schema presented in Table 4.3.
- Any out of range, implausible or missing values were double-checked with the school or jurisdiction that provided the data. Where possible, the correct values were inputted. Where no further information was provided or available, the data were recoded to missing.

Student background variables were also derived for the purposes of reporting achievement outcomes. Table 4.6 shows the derived variables and the transformation rules used to recode them.

Table 4.6: Transformation rules to derive student background variables for reporting

Variable	Label	Transformation rule
School location	ASGSRemote	The geographical classification of the school location according to the ABS remoteness classification (0 = major cities, 1 = inner regional, 2 = outer regional, 3 = remote, 4 = very remote). For the purposes of reporting, categories were then further collapsed into "regional" (inner and outer) and "remote" (remote and very remote).
Gender	GENDER	Classified by response; missing data treated as missing unless the student was present at a single-sex school.
Age	AGE	Derived from the difference between the date of assessment and the date of birth, transformed to whole years.
Indigenous status	INDIG	Coded as Indigenous (1) if response was "yes" to Aboriginal OR Torres Strait Islander OR Both. Coded as non-Indigenous (0) otherwise.
Language spoken at home	LBOTE	Each of the 3 Language spoken at home questions (student, Parent 1 or Parent 2) were recoded to "LBOTE" (1) or "Not LBOTE" (0) according to ASCL codes. The reporting variable (LBOTE) was coded as "LBOTE" (1) if response was "LBOTE" for any student, Parent 1 or Parent 2. If all 3 responses were "not LBOTE" then the LBOTE variable was designated as "not LBOTE" (0). If any of the data were missing, then the data from the other questions were used. If all of the data were missing, then LBOTE was coded as "missing".
Parental education	PARED	Parental education equalled the highest education level (of either parent) according to Australian Standard Classification of Education (ASCED). Where one parent had missing data, the highest education level of the other parent was used. Only if parental education data for both parents were missing would parental education be coded as "missing".
Parental occupation	POCC	Parental occupation equalled the highest occupation group (of either parent). Where one parent had missing data or was classified as "not in paid work", the occupation group of the other parent was used. Where one parent had missing data and the other was classified as "not in paid work", parental occupation equalled "not in paid work". Only if parental occupation data for both parents were missing would parental occupation be coded as "missing".

With respect to the student cognitive and questionnaire data, the following preliminary data cleaning steps were performed:

- Instances of invalid IDs were investigated and, after liaison with the test administration team, corrected where possible or else removed from the dataset.
- Instances of spare IDs were matched with valid student IDs and recoded accordingly. This often necessitated confirmation and cross-checking with the attendance roll data and notes from the test administration team.
- Patterns of missing values were explored and, where appropriate, recoded to “9” for embedded missing, “r” for not reached (cognitive data only) or “n” for not administered.

Further information regarding the scaling procedures implemented for the cognitive achievement data and student questionnaire data can be found in Chapter 5 of this report.

Student eligibility for respondent flag

Psychometric analysis of student cognitive and contextual data requires a minimum threshold of valid responses to be met. To include a student record in the database for scaling, each student must meet a combination of 3 criteria (as shown in Table 4.7) including:

- valid attempts of at least 5 cognitive items, or at least one valid attempt in the student questionnaire
- an appropriate attendance status
- not being listed as exempt.

Students who did not meet the minimum valid attempt criterion were flagged as “Ineligible” and subsequently “Non-respondent”.

Students who met the minimum valid attempt criterion were flagged as “Eligible” for consideration to be identified as “Respondent”. They were marked as “Respondent” only when their attendance status was “Participated”, “Other” or “Not in attendance file” and their exemption status was “Not stated”. The remaining “Eligible” students were flagged as “Non-respondent”.

Students flagged as “Respondent” were kept for the purposes of scaling and analysis only if the school response rate met the minimum requirement as outlined in Chapter 3.

Table 4.7: Rules of flagging students as respondents

Eligibility	Student attendance	Student exemption			
		Not stated	1	2	3
Ineligible	Participated	Non-respondent			
	Absent				
	Exempt				
	Left school				
	Parent refusal				
	Other				
	Not in attendance data file				

Eligibility	Student attendance	Student exemption			
		Not stated	1	2	3
Eligible	Participated	Respondent	Non-respondent		
	Absent	Non-respondent			
	Exempt				
	Left school				
	Parent refusal				
	Other	Respondent			
	Not in attendance data file				

Students identified to be eligible when:
a) at least 5 or more valid responses in cognitive items, or b) at least 1 valid response* in StQ*
Exemption code 1 = severe functional disability
Exemption code 2 = severe intellectual disability
Exemption code 3 = very limited English language proficiency

**Valid responses exclude missing, not reached and not administered*

Data processing for school reporting

Once all student responses were marked, the following data processing steps were implemented to produce the school summary reports that were distributed to the participating schools:

- collation of all marked student data and creation of a single data file for each year level
- removal of introductory practice items for each student and separation of student questionnaire data (which was not included in the analysis for school summary reports)
- checking of the student response data file against the codebook to ensure no major data anomalies
- computation of item per cent correct (unweighted, and excluding not reached responses)
- for partial credit items, computation of item per cent correct for each item in standard NAP sample format (e.g. 75,23 where 0,1,2 item becomes 75 [facility of 1 and 2], 23 [facility of 2 only])
- formatting of data file to required specifications for export into school-specific MS Excel reports.

Providing the school summary reports to schools

After all test data were collected, cleaned, marked and analysed, ACER provided access to an interactive Excel report for all participating schools via the NAP–Science Literacy school administration website.

The NAP–Science Literacy 2023 school summary reports included:

- descriptions of each item in the test
- details of which students were administered each item
- the level of credit students received for each item they were administered
- summary information of the weighted (with preliminary weights) percentage of students receiving different levels of credit for each item.

The school summary reports were developed as interactive Microsoft Excel reports, which were generated through the R open-source software program. These reports allowed schools to undertake detailed interrogation of the data using existing Excel features many would be familiar with.

The school summary reports were hosted on the school administration website, allowing schools to access the reports on the same website used for other NAP–Science Literacy administrative tasks and using existing login credentials.

Schools were advised to read their report in conjunction with the NAP–Science Literacy School Report Instructions provided in the appendices to this report (Appendix D). These instructions provided a description of each of the fields shown in the report and outlined how to interpret the data provided. An example of a school summary report is shown in Appendix E to this report.

Chapter 5: Scaling procedures

Chapter 5: Scaling procedures

Both cognitive and questionnaire items were scaled using Item Response Theory (IRT) scaling methodology. The cognitive items were used to derive a one-dimensional NAP–Science Literacy achievement scale, while a number of scales were constructed based on different sets of questionnaire items.

The scaling model

Test items were scaled with the one-parameter model (Rasch 1960). In the case of dichotomous items, the model predicts the probability of selecting a correct response (value of one) instead of an incorrect response (value of zero) and is modelled as:

$$P_i(\theta_n) = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)}$$

where $P_i(\theta_n)$ is the probability of person *n* scoring 1 on item *i*, θ_n is the estimated ability of person *n* and δ_i is the estimated location of item *i* on this dimension. For each item, item responses are modelled as a function of the latent trait θ_n.

For items with more than 2 (*k*) categories (as for example with Likert-type items) the more general Rasch partial credit model (Masters and Wright 1997) was applied, which takes the form of:

$$P_{x_i}(\theta_n) = \frac{exp \sum_{k=0}^{x}(\theta_n - \delta_i + \tau_{ik})}{\sum_{h=0}^{m_i} exp \sum_{k=0}^{h}(\theta_n - \delta_i + \tau_{ik})} \quad x_i = 0,1,\ldots,m_i$$

where $P_{x_i}(\theta_n)$ denotes the probability of person *n* scoring *x* on item *i*, θ_n denotes the person's ability, the item parameter δ_i gives the location of the item on the latent continuum, τ_{ik} denotes an additional step parameter for each step *k* between adjacent categories and m_i denotes the maximum score attainable on item *i*.

The analysis of item characteristics and the estimation of model parameters were carried out with the ACER ConQuest software package (Version 5.31 software: see Adams, Wu, Cloney, Berezner and Wilson 2020).

Scaling cognitive items

This section outlines the procedures for analysing and scaling the cognitive test items measuring science literacy. The procedures are somewhat different from scaling the student questionnaire items, which will be discussed in the following section.

The model fit of cognitive test items were assessed using a range of item statistics. The weighted mean-square statistic (infit), which is a residual-based fit statistic, was used as a global indicator of item fit. Infit statistics were reviewed both for item and step parameters. In addition to this, item characteristic curves (ICCs) were also used to review item fit. ICCs provide a graphical representation of item fit across the range of student abilities for each item (including dichotomous and partial credit items). The functioning of the partial credit items was further analysed by reviewing the proportion of responses in each response category and the correct ordering of mean abilities of students across response categories. Of the 361 items in the test, one item was removed at Year 10 (x00197367). Consequently, the item was not used to estimate student performance.

Final decisions on retaining test items were based on a range of different criteria. Generally, items were flagged for review if first item calibrations showed a considerably higher infit statistic (e.g. infit > 1.2) as well as low item–rest correlation (0.2 or lower). The ACER project team considered both item-fit criteria, as well as the content of the item, prior to a decision about removing or retaining flagged items for scaling.

Differential item functioning

The quality of the items was also explored by assessing differential item functioning (DIF) by gender. DIF occurs when groups of students with the same ability have different probabilities of responding correctly to an item. For example, if boys have a higher probability of success than girls with the same ability on an item, the item shows DIF in favour of boys. This constitutes a violation of the Rasch model, which assumes that the probability is only a function of ability and not of any other variable. Substantial item DIF with respect to gender may result in bias of performance estimates across gender groups.

An example item that advantages boys is presented in . The graph shows that at any ability (the horizontal axis), the probability of responding correctly is somewhat higher for boys (blue line) than for girls (green line). The DIF was in general consistent over the range of student ability for the item. Three out of 217 items in Year 6, and 9 out of 245 items in Year 10, with a statistically significant, absolute difference of 0.80 logits or greater, were flagged to show potential gender DIF. After a thorough review of the flagged items, and as a result of the balance between items favouring female and male students, it was decided not to delete any items due to significant gender DIF.

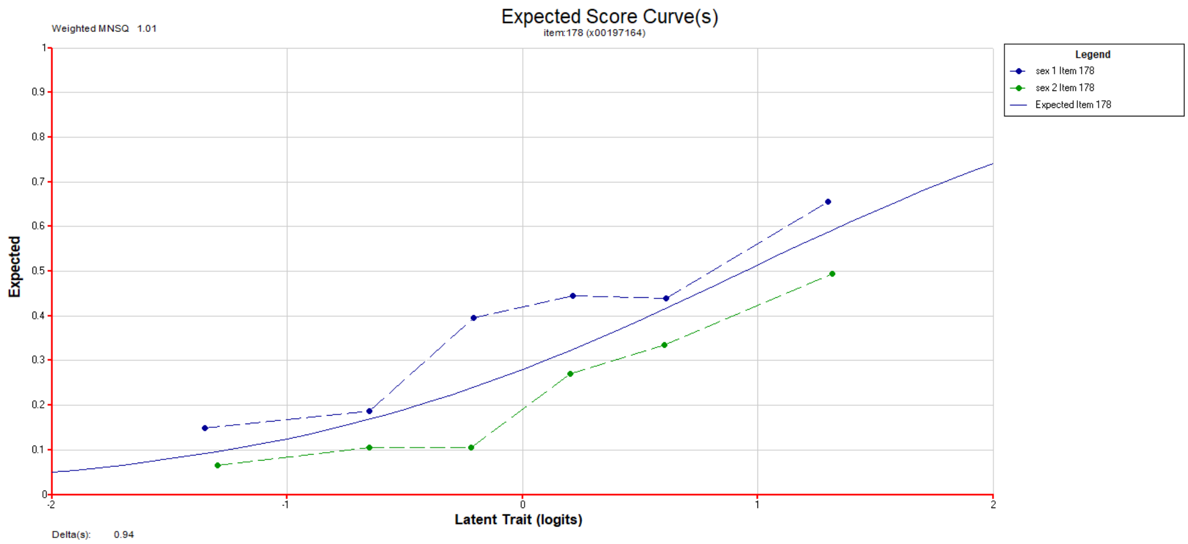

Figure 5.1: Example of item that advantages boys in Year 6

Another form of DIF used to evaluate the items was DIF related to the year level of students. Items with substantial year-level DIF were not used as link items between the Year 6 and the Year 10 assessments. Of the 102 common items between Year 6 and Year 10, 88 were used as link items and 14 were treated as different items for the 2 year levels with year-level-specific item parameters.

Item calibration

Missing student responses, likely caused by issues with test length (“Not reached” items)[11], were omitted from the calibration of item parameters but were treated as incorrect for the scaling of

[11] “Not reached” items were defined as all consecutive missing values at the end of the test except the first missing value of the missing series, which was coded as “embedded missing”, like other items that were presented to the student but did not receive a response.

student responses. All other missing responses were included as incorrect responses for the calibration of items (except for the ones that were not administered).

Item parameters were calibrated excluding the records with less than 5 non-missing responses for the cognitive test items, as well as records that were “exempt” or “left school”. The student weights were rescaled to ensure that each state or territory was equally represented in the sample. In the first stage of the scaling procedures, the items were calibrated separately for Year 6 and Year 10. There were 361 items that were included in total, of which 115 were Year 6 only items and 144 were Year 10 only items. The other 102 items were used for both year levels. Of the 102 common items, 88 were used as vertical link items and 14 were regarded as different items in the 2 year levels.

The difficulties of these 88 link items are plotted in Figure 5.2, with Year 6 estimates on the horizontal axis and Year 10 estimates on the vertical axis. For each set of 88 items, their respective difficulties were centred to a mean of zero for this graph. The black broken lines represent the boundaries of the confidence intervals around differences from zero (the identity line indicating that there are no differences in item difficulty). The green broken line is the identity line. The pink broken line is the best fit line of the scatter plot. The largest difference between the 2 relative difficulties was approximately 0.7 logits.

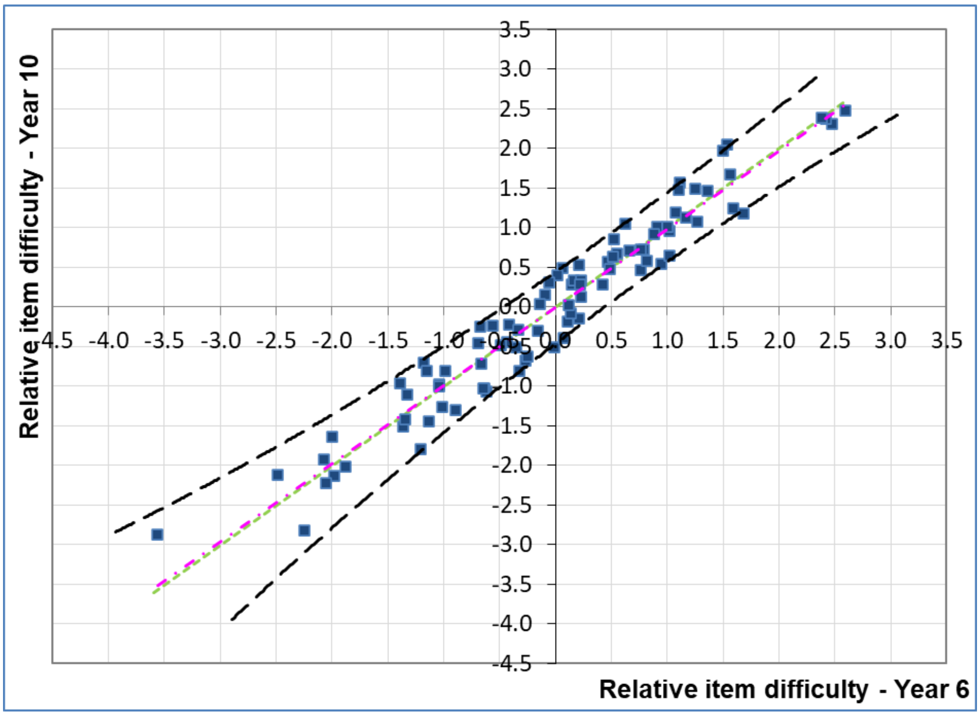

Figure 5.2: Scatter plot of relative item difficulties for Year 6 and Year 10

Figure 5.3 presents item maps for the 2 year levels. The crosses represent students; the numbers represent items. In the case of a partial credit item, the threshold is included. The vertical line represents the measured science literacy scale with high-performing students and difficult items at the top, and low-performing students and easy items at the bottom. The 2 scales are not directly comparable because they have been calibrated separately, but they have been lined up approximately for this report. The response probability in this figure is 0.5, which means that students with an ability equal to the difficulty (or threshold) of an item have a 50% chance of responding correctly to that item. The figure shows that the test was well targeted at each year level.

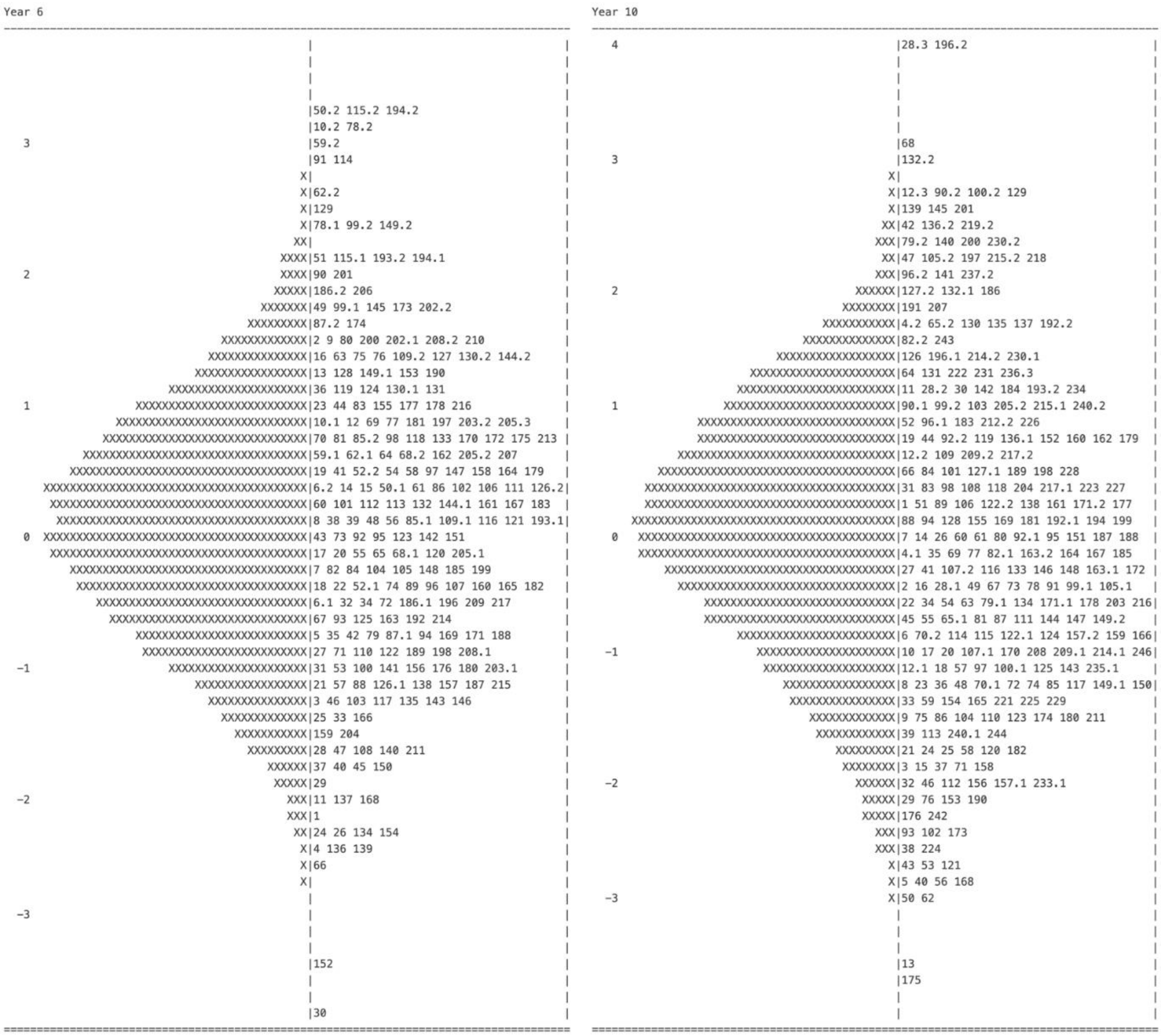

Figure 5.3: Item maps for Year 6 and Year 10

The overall reliability of the test, as obtained from the scaling model, was 0.85 for Year 6 and 0.89 for Year 10 (ACER ConQuest EAP/PV reliability estimate). Appendix F to this report shows the item difficulties in logits on the NAP–Science Literacy scale. It additionally presents those difficulties with a response probability of 0.62[12], required for reporting against the proficiency standards. It also shows the respective percentages of correct responses for each year sample (giving equal weight to each jurisdiction). The weighted fit statistics are included in the last column. In addition, column 4 indicates if an item was used as a horizontal link (trend) item.

In the second stage scaling procedures, Year 10 data was scaled, anchoring the estimates of the 88 vertical link items to their Year 6 item parameter estimates in order to place both year levels on the same scale.

Dimensionality

Science Literacy is a particularly well-suited construct within which to consider CCT. During the development of the NAP–Science Literacy Assessment Framework and the conceptual stages of item development, an extensive exercise in mapping the CCT continuum to the cognitive dimensions

[12] This means that a student with a scale score equal to the item difficulty parameter has 62% probability of giving a correct response to the test question.

was undertaken to illustrate the representation and possible coverage of CCT within and across the cognitive areas.

The NAP–Science Literacy 2023 assessment consisted of CCT items and non-CCT items, with 47% of items explicitly mapped to the CCT assessable elements. The correlations between these 2 latent variables were estimated by fitting a 2-dimensional model using a quadrature method in ACER ConQuest. The latent correlation for Year 6 was 0.981 and for Year 10 was 0.976, which supports the unidimensional structure of the NAP–Science Literacy scale.

Horizontal equating

Test forms at both year levels consisted of newly developed items and trend items. The trend items were developed for and used in previous cycles. As they had been kept confidential, they could be used as horizontal link items to equate the results of the 2023 assessment with the established NAP–Science Literacy scale. To ensure that the link items had the same measurement properties across cycles, their relative difficulties in 2023 and 2018 were compared.

shows a scatter plot of item difficulties for the 41 horizontal link items in 2018 and 2023. The average difficulty of each set of link items was set to zero and each dot represents one link item. The black broken lines represent the boundaries of the confidence intervals around differences from zero (the identity line indicating that there are no differences in item difficulty). The green broken line is the identity line. The pink broken line is the best fit line of the scatter plot. Items outside of these lines had statistically significant deviations from the identity line. The original standard errors provided by ACER ConQuest were adjusted by multiplying them by the square root of 4, the approximate design effect in 2023. This correction was made because data were collected from a cluster sample design, whereas the scaling software assumes simple random sampling of data (see Chapter 3 for more information on sampling). Historical items were not used as link items if the difference between relative item difficulties was significant and more than 0.5 logits.

Out of 45 common items in Year 6, 4 showed large DIF. The remaining 41 items were retained and used for equating between the 2023 and 2018 cycles. For both assessments, this set of selected link items showed similar average discrimination (item–rest correlation was 0.38 in 2018 and 0.34 in 2023).

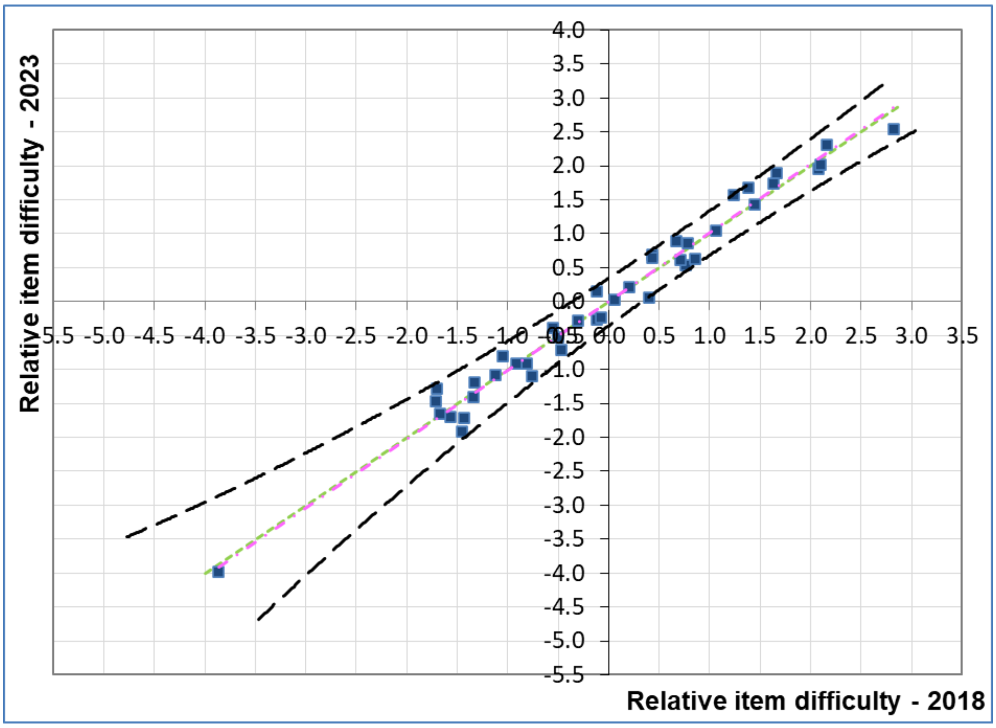

Figure 5.4: Relative item difficulties in logits of horizontal link items between 2018 and 2023

Item–rest correlation is an index of item discrimination, which is computed as the correlation between the scored item and the raw score of all other items in a booklet. It indicates how well an item discriminates between high- and low-performing students. The 2018 and 2023 values of these discrimination indices are plotted in Figure 5.5.

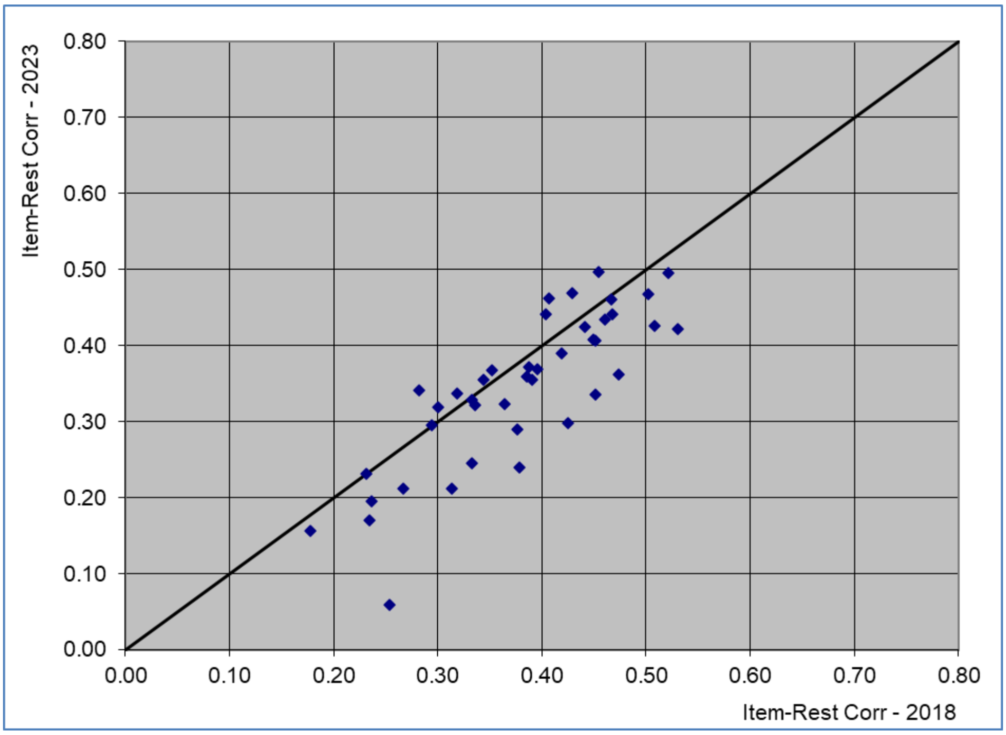

Figure 5.5: Discrimination of link items in 2018 and 2023

After the selection of link items, common item equating was used to shift the 2023 scale onto the historical scale. The value of the shift is the difference in average difficulty of the link items between 2018 and 2023 (0.588). After applying this shift, the same transformation was applied as in 2018. Original scale scores (logits) were converted as:

$$\theta_n^* = \{(\theta_n + 0.588 - 0.507 + 0.131 - \bar{\theta}_{06})/\sigma_{06}\} \times 100 + 400$$

where θ_n^* is the ability estimate for student *n*, θ_n is the original ability estimate for student *n* in logits, the numeric terms are the equating shift to 2018 and then to the historical scale, incorporating the horizontal shift from 2018 to 2015 and subsequently from 2015 to 2006 (-0.535 and 0.028 logits, respectively) and the adjustment of 0.131 logits to correct for the effect of switching from paper- to computer-based testing, $\bar{\theta}_{06}$ is the mean ability in logits of the Year 6 students in 2006 (0.201) and σ_{06} is the standard deviation in logits of the Year 6 students in 2006 (0.955).

Uncertainty in the link

The shift that equates the 2023 data with the 2018 data depends upon the change in difficulty of each of the individual link items. As a consequence, the sample of link items that have been chosen will influence the estimated shift. This means that the resulting shift could be slightly different if an alternative set of link items had been selected. As a result, there is an uncertainty associated with the equating that is due to the choice of link items, similar to the uncertainty associated with the sampling of schools and students.

The uncertainty that results from the selection of a subset of link items is referred to as a linking or equating error. This error should be taken into account when making comparisons between the results from different data collections across time. Just as with the error that is introduced through the process of sampling students, the exact magnitude of this equating error cannot be determined. We can, however, estimate the likely range of magnitudes for this error and take this error into account when interpreting results. This likely range of magnitude for the combined errors is represented as a standard error of each reported statistic.

The following approach has been used to estimate the equating error. Suppose we have a total of *L* score points in the link items in *K* clusters. Use *i* to index items in a unit and *j* to index units so that $\hat{\delta}_{ij}^{y}$ is the estimated difficulty of item *i* in unit *j* for year *y*, and let:

$$c_{ij} = \hat{\delta}_{ij}^{2023} - \hat{\delta}_{ij}^{2018}$$

The size (number of score points) of unit *j* is m_j so that:

$$\sum_{j=1}^{K} m_j = L \quad \text{and} \quad \bar{m} = \frac{1}{K}\sum_{j=1}^{K} m_j$$

Further, let:

$$c_{\bullet j} = \frac{1}{m_j}\sum_{i=1}^{m_j} c_{ij} \quad \text{and} \quad \bar{c} = \frac{1}{N}\sum_{j=1}^{K}\sum_{i=1}^{m_j} c_{ij}$$

Then the link error, taking into account the clustering, is as follows:

$$LinkError_{2023,2018} = \sqrt{\frac{\sum_{j=1}^{K} m_j^2 (c_{\bullet j} - \bar{c})^2}{K(K-1)\bar{m}^2}} = \frac{\sum_{j=1}^{K} m_j^2 (c_{\bullet j} - \bar{c})^2}{L^2}\frac{K}{K-1}$$

The link error between 2023 and 2018 is 6.03 scale score points.

The equating error between 2023 and 2015 is the square root of the sum of the squares of the equating errors between adjacent cycles.

$$error_{2023,2015} = \sqrt{6.03^2 + 4.39^2} = 7.46$$

The equating error between 2023 and each of the previous cycles are presented in Table 5.1.

Table 5.1: Equating error between 2023 and each of the previous cycles

Comparison cycle	Equating error on reporting scale
2018	6.03
2015	7.46
2012	8.99
2009	9.56
2006	10.24

Plausible values

Plausible values methodology was used to generate estimates of students' science literacy abilities. Using item parameters anchored at their estimated values from the calibration process, plausible values were randomly drawn from the marginal posterior of the latent distribution (Mislevy 1991; Mislevy and Sheehan 1987; von Davier, Gonzalez and Mislevy 2009). During this process, "not reached" items were marked as incorrect responses, in the same way as embedded missing responses were scored in the item calibration. Estimations are based on the conditional item response model and the population model, which includes the regression on background and questionnaire variables used for conditioning (Adams and Wu 2002). The ACER ConQuest Version 5.31 software was used for drawing plausible values.

Some variables were used as direct regressors in the conditioning model for drawing plausible values. The variables included school mean performance adjusted for the student's own performance[13] and dummy variables for the school-level variables of state/territory, sector, geographic location of the school, and the student-level variables of gender, Indigenous status, language background other than English, parental school education, and parental occupation group. Principal component analysis (PCA) was used to extract component scores from all other student-background variables and responses to the student questionnaire items. The principal components were estimated separately for each year level. Subsequently, the components that explained 99% of the variance in the original variables were included as regressors in the final conditioning model for each state or territory. Details of the coding of variables included directly in the conditioning model or included in the PCA are listed in Appendix G: Variables for conditioning.

Scaling questionnaire items

Before estimating student scores on the questionnaire scales, exploratory and confirmatory factor analyses were conducted in MPlus Version 8.5 (Muthen and Muthen 2020) with questionnaire data[14].

Exploratory factor analyses were carried out on all the questions to provide evidence of the factor structure (suggesting a one-factor solution to all of the questionnaire scales). Confirmatory factor analyses were carried out for all scales. For example, there are 8 items designed to measure student

[13] So-called *weighted likelihood estimates* (WLEs) were used as ability estimates in this case (Warm 1989).

[14] Not all questionnaire material was included for the purpose of developing scales, but rather for use in single-item reporting. This was particularly true for some trend material. This includes items QN01.1-QN01.7 and QN07.

perceptions of the scientific process (WORKSCI) and 6 items reflecting students' attitudes to equality in science (EQUALITY). The analyses confirmed the expected one-dimensional factor structure of each of these item sets.

Student and item parameters were estimated using the ACER ConQuest Version 5.31 software. Items were scaled using the Rasch partial credit model (Masters and Wright 1997). Item parameters and student scores were jointly estimated, giving equal weight to jurisdictional samples. Records with at least 2 non-missing responses per scale were used for the scaling of the questionnaire items. Weighted likelihood estimation was used to obtain the individual student scores (Warm 1989). The scales were converted to a metric with a mean score of 50 and a standard deviation of 10 based on the Year 6 sample and applied to both Year 6 and Year 10.

Table 5.2 describes the main characteristics of the questionnaire scales, including the scale reliabilities (Cronbach's alpha) and their respective correlation with science literacy scores (using PV1). The CCTEFF (Student self-efficacy to apply critical and creative thinking to problem solving tasks) scale has been developed for Year 10 only and the metric is based on the Year 10 sample.

Table 5.2: Description of the questionnaire scales

Name	Index name	Question numbers	Number of items	Cronbach's alpha		Correlation with achievement	
				Year 6	Year 10	Year 6	Year 10
Student perception of the nature of science	NATURE	QN02.1-QN02.6	6	0.83	0.84	0.13	0.28
Student experience of science-related activities outside of school	EXPOUT	QN03.1-QN03.8	8	0.84	0.88	0.07	0.29
Student experience of science-related activities at school	EXPSCH	QN04.1-QN04.3	3	0.71	0.78	0.15	0.33
Student perception of the influence of science	INFLUENCE	QN01.8-QN01.10, QN05.1-QN05.7	10	0.89	0.92	0.38	0.47
Science topics studied at school	TOPICS	QN06.1-QN06.9	9	0.78	0.85	0.04	0.24
Student perception of the scientific process	WORKSCI	QN08.1-QN08.6	6	0.92	0.95	0.38	0.46
Student attitude to equality in science	EQUALITY	QN09.1-QN09.6	6	0.92	0.91	0.30	0.34

Name	Index name	Question numbers	Number of items	Cronbach's alpha		Correlation with achievement	
				Year 6	Year 10	Year 6	Year 10
Exposure to activities conducive to critical and creative thinking	CCTEXPOS	QN10.1-QN10.8	8	0.90	0.91	0.17	0.18
Family support for critical and creative thinking	CCTSUPP	QN11.1-QN11.6	6	0.93	0.94	0.21	0.32
Participation in activities outside of school related to critical and creative thinking	CCTPART	QN12.1-QN12.5	5	0.80	0.86	0.16	0.23
Student self-efficacy to apply critical and creative thinking to problem solving tasks	CCTEFF	QN13.1-QN13.8	8	-	0.95	-	0.40

Chapter 6: Proficiency levels and the proficient standard

Chapter 6: Proficiency levels and the proficient standard

One of the key objectives of NAP–Science Literacy is to monitor and report trends in science literacy achievement. As is standard practice in NAP sample assessments, 2 summary measures of student achievement are used in addition to reporting students' science literacy solely in terms of average scale score achievement. These are:

1. Proficiency level achievement: for NAP–Science Literacy, a set of 5 proficiency levels were developed, each representing a range on the scale, which was accompanied by descriptions of the science skills and capabilities associated with that level. The percentage of students performing at each proficiency level provided a measure of student achievement.

2. Proficient standard achievement: the proficient standards represent points on the NAP–Science Literacy scale indicating a "challenging but reasonable" proficiency level that Year 6 and Year 10 students would be expected to reach. The percentage of students who had attained (i.e. reached or exceeded) the proficient standard presented an additional measure of student performance. The proportion of students achieving at or above the proficient standard is also the national Key Performance Measure (KPM) for science literacy specified in the Measurement Framework for Schooling in Australia 2020 (ACARA 2020b).

This chapter describes the development of these 2 measures for NAP–Science Literacy.

Proficiency levels

The NAP–Science Literacy scale forms the basis for the empirical comparison of student achievement. In addition to the scale, 5 proficiency levels with substantive descriptions were established in 2006. These described levels were syntheses of the item contents within each level. Scale level descriptions have been reviewed following each cycle of the assessment, including most recently in 2023, to ensure they accurately reflect the NAP–Science Literacy test content and continue to adequately cover the definition of science literacy.

Comparison of student achievement against the proficiency levels provides an empirically and substantively convenient way of describing profiles of student achievement. 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 possess the understandings and skills defined as applying to lower proficiency levels.

Creating the proficiency levels

The proficiency levels were established in 2006 and were based on an approach developed for the OECD's Programme for International Student Assessment (PISA). PISA made use of a method that ensured that the notion of *being at a level* could be interpreted consistently and in line with the fact that the achievement scale is a continuum. It provides a common understanding about what *being at a level* means and that the meaning of *being at a level* is consistent across levels. Similar to the approach taken in the PISA study (OECD 2005:255), this method took the following 3 variables into account:

- the expected success of a student at a particular level on a test containing items at that level
- the width of the levels in that scale
- the probability that a student in the middle of a level would correctly answer an item of average difficulty for that level.

To achieve this for NAP–Science Literacy, the following 2 parameters for defining proficiency levels were adopted:

- setting the response probability for the analysis of data at $p = 0.62$
- setting the width of the proficiency levels at 1.25 logits.

Once these parameters had been established, it was possible to make the following statements about the achievement of students relative to the proficiency levels:

- A student at the lowest possible point of the proficiency level is likely to get 50% correct on a test made up of items spread uniformly across the level, from the easiest to the most difficult.
- A student at the lowest possible point of the proficiency level is likely to get 62% correct on a test made up of items similar to the easiest items in the level.
- A student at the top of the proficiency level is likely to get 82% correct on a test made up of items similar to the easiest items in the level.

It should be acknowledged that it would have been possible to choose other solutions with different parameters defining the proficiency levels. The approach used in PISA, and adopted for NAP–Science Literacy, attempted to balance the notions of mastery and "pass" in a way that is likely to be understood by the community.

Proficiency level cut-points

With the addition of Year 10 content to the scale in 2018, as well as the implementation of a standard-setting exercise in the following year to establish the Year 10 proficient standard, adjustments to the width of the proficiency levels were made so that they adequately covered the breadth of scale scores across the 2 year-level cohorts.

The scale score cut-points for the proficiency levels remained unchanged for 2023 and are shown in Table 6.1. As can be seen from the figure, the width of each level is slightly over 100 scale score points.

Table 6.1: Proficiency level cut-points and percentage of Year 6 and Year 10 students in each level in 2023

Proficiency level	Cut-points		Percentage			
	Logits	Scale	Year 6		Year 10	
Level 5 and above	2.13	602	2	(±0.6)	20	(±2.2)
Level 4	1.13	497	16	(±1.5)	34	(±2.7)
Level 3	0.13	393	39	(±2.3)	30	(±2.4)
Level 2	-0.87	288	32	(±1.9)	13	(±1.8)
Level 1 and below			12	(±1.9)	3	(±1.3)

*Confidence intervals (1.96 * SE) are reported in brackets.*
Results are rounded to the nearest whole number so some totals may appear inconsistent.

Describing proficiency levels

The description of each level on the NAP–Science Literacy scale provides a synthesised overview of the knowledge, skills and understandings that a student working within the level can demonstrate. The levels are set so that any student is likely to respond correctly to at least half of the items in their proficiency level. As outlined previously, a student with an achievement scale score at the bottom of a level has a 62% chance of correctly answering any question at the bottom of that level. Conversely, that student would have a 38% chance of correctly answering any question at the top of that level.

Overall, higher levels on the scale refer to more complex applications of knowledge, skills and comprehension. The scale is developmental in the sense that students are assumed to be typically able to demonstrate achievement of the skills and cognition described in the level below as well as at their measured level of achievement. The descriptions for each proficiency level are provided in the appendices to this report (Appendix H).

While the level descriptions have been updated over the cycles of NAP–Science Literacy to reflect new scientific contexts and refreshed assessment frameworks, the underlying conceptualisation of science literacy measured in NAP–Science Literacy has remained constant. This principle is important in assessments that extend over several cycles and are concerned with measuring change. It is accepted that changes in methods and content are necessary for assessments to remain relevant, but that maintaining the meaning of the construct is a necessary condition for measuring change (von Davier and Mazzeo 2009).

Setting the proficient standards

The process for setting standards in science literacy, information and communications technologies, civics and citizenship, and secondary (15-year-old) reading, mathematics and science was endorsed by the Performance Measurement and Reporting Taskforce (PMRT) of the Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) at its meeting on 6 March 2003 and is described in the paper "Setting National Standards" (PMRT 2003). This process, referred to as the empirical judgmental technique, requires stakeholders to examine the test items and the results from the national assessments, and agree on a proficient standard for each of the year levels an assessment is aimed at.

A proficient standard represents a "'challenging but reasonable' expectation of student achievement at a year level with students needing to demonstrate more than elementary skills expected at that year level" (ACARA 2020a:6). Importantly, a proficient standard is different from either a benchmark or a national minimum standard, which both refer to a level of minimum competence.

The proficient standard for Year 6 was established in 2006 and for Year 10 in 2018. Both standards were established through a standard-setting process that brought together expert science educators, including practising primary and secondary teachers, from all states and territories across all 3 education sectors. It was also inclusive and reflective of teaching experiences across major cities, regional and remote locations, as well as high and low socio-educational communities.

The proficient standard for Year 6 is 393 scale score points, which is the boundary between levels 2 and 3 on the NAP–Science Literacy scale. The proficient standard for Year 10 is 497 scale score points, which is the boundary between levels 3 and 4 on the scale. Year 6 students performing at level 3 or higher and Year 10 students performing at level 4 or higher above have consequently met or exceeded their relevant proficient standard.

Chapter 7: Reporting of results

Chapter 7: Reporting of results

The students assessed in NAP–Science Literacy 2023 were selected using a 2-stage cluster sampling procedure. In the first stage, schools were sampled from a sampling frame with a probability proportional to their size as measured by student enrolments in the relevant year level. In the second stage, 20 students at each year level were randomly sampled within schools (see Chapter 3 on sampling and weighting).

Applying cluster sampling techniques is an efficient and economical way of selecting students in educational research. However, as these samples were not obtained through (one-stage) simple random sampling, standard formulae to obtain sampling errors of population estimates are not appropriate. In addition, NAP–Science Literacy estimates were obtained using plausible value methodology (see Chapter 5 on scaling procedures), which allows for estimating and combining the measurement error of achievement scores with their sampling error.

Reporting of results by subgroups of interest becomes more limited as group sizes decrease, due to the increase in error that accompanies this. For this cycle of NAP–Science Literacy, the gender category “other” is not reported because there are fewer than 30 students with valid data.

This chapter describes the method applied for estimating sampling as well as measurement error. In addition, it contains a description of the types of statistical analyses and significance tests that were carried out for reporting of results in this report.

Computation of sampling and measurement variance

Unbiased standard errors from studies should include both sampling variance and measurement variance. One way of estimating sampling variance on population estimates from cluster samples is by using the application of replication techniques (Wolter 1985). The sampling variances of population means, differences, percentages and correlation coefficients in NAP–Science Literacy studies were estimated using the jackknife repeated replication technique (JRR). The other component of the standard error of achievement test scores, the measurement variance, can be derived from the variance among the 5 plausible values for NAP–Science Literacy. In addition, for comparing achievement test scores with those from previous cycles (2006, 2009, 2012, 2015 and 2018), an equating error was added as a third component of the standard error.

Replicate weights

When applying the JRR method for stratified samples, primary sampling units (PSUs) – in this case schools – are paired into pseudo-strata, also called sampling zones. The assignment of schools to these sampling zones needs to be consistent with the sampling frame from which they were sampled (to obtain pairs of schools that were adjacent in the sampling frame) and zones are always constructed within explicit strata of the sampling frame. This procedure ensures that schools within each zone are as similar to each other as possible.[15] For NAP–Science Literacy 2023, 191 sampling zones were constructed in Year 6 and 114 in Year 10.

Within each sampling zone, a jackknife indicator variable was created by randomly assigning a value of 2 for one school and assigning a value of zero to the other one. To create replicate weights for each of these sampling zones, the jackknife indicator variable was multiplied by the original sampling weights of students within the corresponding zone so that one of the paired schools had a contribution of zero and the other school a double contribution, whereas schools from all other sampling zones remained unmodified.

[15] In the case of an odd number of schools within an explicit stratum on the sampling frame, the remaining school is randomly divided into 2 halves and each half assigned to the 2 other schools in the final sampling zone to form *pseudo-schools*.

At each year level, 193 replicate weights were computed. The last 2 replicates in Year 6 and the last 79 replicates in Year 10 were equal to the final sampling weight. This was done to have a consistent number of replicate weight variables in the final database.

Standard errors

To compute the sampling variance for a statistic *t*, *t* is estimated once for the original sample *S* and then for each of the jackknife replicates J_h. The JRR variance is computed using the formula:

$$Var_{jrr}(t) = \sum_{h=1}^{H} \left[t(J_h) - t(S) \right]^2$$

where *H* is the number of replicate weights, *t(S)* the statistic *t* estimated for the population using the final sampling weights, and $t(J_h)$ the same statistic estimated using the weights for the h^{th} jackknife replicate. For all statistics that are based on variables other than student test scores (plausible values) the standard error of *t* is equal to:

$$\sigma(t) = \sqrt{Var_{jrr}(t)}$$

The computation of JRR variance can be obtained for any statistic. However, many standard statistical software packages like SPSS® do not generally include any procedures for replication techniques. Therefore, specialist software, the SPSS® Replicates add-in, was used to run tailored SPSS® macros to estimate JRR variance for means and percentages.[16]

Population statistics for NAP–Science Literacy scores were always estimated using all 5 plausible values with standard errors reflecting both sampling and measurement error. If *t* is any computed statistic and t_i is the statistic of interest computed on one plausible value, then:

$$t = \frac{1}{M} \sum_{i=1}^{M} t_i$$

with *M* being the number of plausible values.

The sampling variance *U* is calculated as the average of the sampling variance over all plausible values (U_i):

$$U = \frac{1}{M} \sum_{i=1}^{M} U_i$$

Using 5 plausible values for data analysis allows the estimation of the error associated with the measurement of NAP–Science Literacy due to the lack of precision of the test instrument. The measurement variance or imputation variance B_m was computed as:

$$B_m = \frac{1}{M-1} \sum_{i=1}^{M} (t_i - t)^2$$

[16] Conceptual background and application of macros with examples are described in the PISA Data Analysis Manual SPSS®, Second Edition (OECD 2009).

To obtain the final standard error of NAP–Science Literacy statistics, the sampling variance and measurement variance were combined as:

$$SE = \sqrt{U + \left(1 + \frac{1}{M}\right) B_m}$$

with *U* being the sampling variance.

The 95% confidence interval, as used in this report, was computed as 1.96 times the standard error. The actual 95% confidence interval of a statistic is between the value of the statistic *minus* 1.96 times the standard error and the value of the statistic *plus* 1.96 times the standard error.

Reporting of mean differences

Chapter 4 of the public report includes comparisons of achievement test results across states and territories; that is, means of scales and percentages are compared in graphs and tables. Each population estimate is accompanied by its 95% confidence interval. In addition, tests of significance for the difference between estimates are provided, to flag results that are significant at the 5% level ($p < 0.05$), which indicates a 95% probability that these differences are not a result of sampling and measurement error.

The following types of significance tests for achievement mean differences in population estimates were reported:

- between states and territories
- between student subgroups such as male and female students
- between this assessment cycle and previous ones in 2018, 2015, 2012, 2009 and 2006 for Year 6, and between this assessment cycle and the previous one in 2018 for Year 10.

Mean differences between states and territories and year levels

Pairwise comparison tables allow the comparison of population estimates between one state or territory and another, or between Year 6 and Year 10. Differences in means were considered significant when the test statistic *t* was outside the critical values ±1.96 ($\alpha = 0.05$). The *t* value is calculated by dividing the difference in means by its standard error, which is given by the formula:

$$SE_{dif\_ij} = \sqrt{SE_i^2 + SE_j^2}$$

where SE_{dif_ij} is the standard error of the difference and SE_i and SE_j are the standard errors of the 2 means *i* and *j*. This computation of the standard error was only applied for comparisons between 2 samples that had been drawn independently from each other (for example, jurisdictions or year levels).

In this report, differences were also estimated between percentages attaining the proficient standards in states and territories. The method for estimating the standard error of the difference between percentages is identical to the procedure described for mean differences.

Mean differences between dependent subgroups

The formula for calculating the standard error described in the previous section is not appropriate for subgroups from the same sample (see OECD 2009 for more detailed information). Here, the covariance between the 2 standard errors for subgroup estimates needs to be considered and JRR should be used to estimate correct sampling errors of mean differences. Standard errors of differences between statistics for subgroups from the same sample (for example, groups classified according to student background characteristics) were derived using the SPSS® Replicates add-in. Differences between subgroups were considered significant when the test statistic *t* was outside the critical values ±1.96 (α = 0.05). The value *t* was calculated by dividing the mean difference by its standard error.

Mean differences between assessment cycles (2006, 2009, 2012, 2015, 2018 and 2023)

Chapter 4 of the public report also includes comparisons of achievement results across assessment cycles. The process of equating tests across different achievement cycles introduces a new form of error when comparing population estimates over time: the equating or linking error. When computing the standard error, equating error as well as sampling and measurement error was taken into account. The computation of equating errors is described in Chapter 5 of this report.

The value of the equating error between 2023 and the previous assessment in 2018 is 6.03 score points on the NAP–Science Literacy scale for both year levels. When testing the difference of a statistic between these 2 assessment cycles, the standard error of the difference was computed as follows:

$$SE(t_{23} - t_{18}) = \sqrt{SE_{23}^2 + SE_{18}^2 + EqErr_{23\_18}^2}$$

where *t* can be any statistic in units on the NAP–Science Literacy scale (mean, percentile, gender difference, but *not* percentages), SE_{23}^2 is the respective squared standard error of this statistic in 2023, SE_{18}^2 the corresponding squared standard error in 2018 and $EqErr_{23\_18}^2$ the squared equating error for comparing 2023 with 2018 results.

When comparing population estimates between 2023 and the assessment in 2015, 2 equating errors (between 2023 and 2018 and between 2018 and 2015) had to be considered. This was achieved by applying the following formula for the calculation of the standard error for differences between statistics from 2023 and 2015:

$$SE(\mu_{23} - \mu_{15}) = \sqrt{SE_{23}^2 + SE_{15}^2 + EqErr_{23\_15}^2}$$

where $EqErr_{23\_15}^2$ reflects the uncertainty associated with the equating between the assessment cycles of 2023 and 2018 (6.03 score points) as well as between 2018 and 2015 (4.39 score points). This combined equating error was equal to 7.46 score points and was calculated as:

$$EqErr_{23\_15} \quad = \sqrt{EqErr_{23\_18}^2 \quad + EqErr_{18\_15}^2}$$

Similarly, for comparisons between 2023 and the NAP–Science Literacy assessment in 2006, the equating errors between each adjacent pair of assessments had to be taken into account and standard errors for differences were computed as:

$$SE(\mu_{23} - \mu_{06}) = \sqrt{SE_{23}^2 + SE_{06}^2 + EqErr_{23\_06}^2}$$

$\mathrm{EqErr}_{23\_06}^2$ reflects the uncertainty associated with the equating between the assessment cycles of 2023 and 2018 (6.03 score points) and between 2018 and 2006 (8.28 score points). The combined equating error was equal to 10.24 score points and was calculated as:

$$EqErr_{23\_06} \quad = \sqrt{EqErr_{23\_18}^2 \quad + EqErr_{18\_06}^2}$$

To report the significance of differences between percentages at or above proficient standards, the corresponding equating error had to be estimated using a different approach. To obtain an estimate, the following replication method was applied to estimate the equating error for percentages at the proficient standards.

For the cut-point that defines the corresponding proficient standard at each year level (393 for Year 6 and 497 for Year 10), a number of *n* replicate cut-points were generated. This was achieved by adding a random error component with a mean of 0 and a standard deviation equal to the estimated equating error of 6.03 score points for comparisons between 2023 and 2018, 7.46 score points for comparisons between 2023 and 2015, 8.99 score points for comparisons between 2023 and 2012, 9.56 score points for comparisons between 2023 and 2009, and 10.24 score points for comparisons between 2023 and 2006. Percentages of students at or above each replicate cut-point (ρ_n) were computed and the equating error was estimated as:

$$EqErr(\rho) = \sqrt{\frac{\sum(\rho_n - \rho_o)^2}{n}}$$

where ρ_o is the percentage of students at or above the (reported) proficient standard. The standard errors of the differences in percentages at or above proficient standards between 2023 and 2018 were calculated as:

$$SE(\rho_{23} - \rho_{18}) = \sqrt{SE(\rho_{23})^2 + SE(\rho_{18})^2 + EqErr(\rho_{23\_18})^2}$$

where ρ_{23} is the percentage at or above the proficient standard in 2023 and ρ_{18} in 2018, $SE(\rho_{23})$ and $SE(\rho_{18})$ their respective standard errors, and $EqErr(\rho_{23\_18})$ the equating error for comparisons. For estimating the standard error of the corresponding differences in percentages at or above proficient standards between 2023 and 2015, the following formula was used:

$$SE(\rho_{23} - \rho_{15}) = \sqrt{SE(\rho_{23})^2 + SE(\rho_{15})^2 + EqErr(\rho_{23\_15})^2}$$

Likewise, for estimating the standard error of the corresponding differences in percentages at or above proficient standards between 2023 and 2009 and between 2023 and 2006, the following formulas were used:

$$SE(\rho_{23} - \rho_{09}) = \sqrt{SE(\rho_{23})^2 + SE(\rho_{09})^2 + EqErr(\rho_{23\_09})^2}$$

$$SE(\rho_{23} - \rho_{06}) = \sqrt{SE(\rho_{23})^2 + SE(\rho_{06})^2 + EqErr\left(\rho_{23\_06}\right)^2}$$

For NAP–Science Literacy 2023, 5000 replicate cut-points were created. Equating errors on percentages were estimated for each sample or subsample of interest. Table 7.1 and Table 7.2 show the values of these equating errors for Year 6 and Year 10 respectively.

Table 7.1: Year 6 equating errors for comparisons between percentages

Group	2023/2018	2023/2015	2023/2012	2023/2009	2023/2006
Aust	2.28	2.83	3.42	3.64	3.90
NSW	2.14	2.67	3.24	3.45	3.71
VIC	2.56	3.13	3.75	3.97	4.24
QLD	2.22	2.78	3.39	3.61	3.88
SA	2.43	3.02	3.62	3.84	4.11
WA	2.17	2.69	3.25	3.45	3.70
TAS	2.04	2.58	3.17	3.38	3.65
NT	2.31	2.84	3.40	3.61	3.86
ACT	2.41	2.91	3.44	3.63	3.87
Female	2.41	2.99	3.60	3.83	4.10
Male	2.15	2.68	3.24	3.45	3.71
Non-Indigenous students	2.30	2.85	3.45	3.67	3.93
Indigenous students	1.92	2.42	2.94	3.13	3.36
English only	2.25	2.81	3.41	3.63	3.89
Language other than English	2.40	2.93	3.50	3.71	3.96
Major cities	2.24	2.78	3.36	3.57	3.83
Regional	2.39	2.98	3.60	3.83	4.10
Remote	2.02	2.49	2.97	3.15	3.35
Senior managers and professionals	1.91	2.37	2.86	3.05	3.27
Other managers and associate professionals	2.29	2.87	3.48	3.71	3.98
Tradespeople & skilled office, sales and service staff	2.54	3.16	3.83	4.08	4.37
Machine operators, labourers, hospitality, and related staff	2.82	3.45	4.12	4.36	4.65
Not in paid work in last 12 months	2.31	2.82	3.36	3.55	3.79
Bachelor’s degree or above	2.09	2.98	3.61	3.85	4.13
Advanced diploma/diploma	2.38	3.20	3.88	4.14	4.44
Certificate I to IV (inc trade cert)	2.56	3.07	3.69	3.92	4.18
Year 12 or equivalent	2.45	2.11	2.52	2.67	2.85
Year 11 or equivalent	1.73	2.67	3.13	3.30	3.51
Year 10 or equivalent	2.23	2.58	3.07	3.25	3.47
Year 9 or equivalent or below	2.15	3.52	4.19	4.42	4.70

Table 7.2: Year 10 equating errors for comparisons between percentages

Group	2023/2018
Aust	2.15
NSW	2.20
VIC	2.17
QLD	2.16
SA	1.82
WA	2.23
TAS	1.85
NT	3.62
ACT	1.83
Female	2.19
Male	2.10
Non-Indigenous students	2.19
Indigenous students	1.40
English only	2.15
Language other than English	2.15
Major cities	2.12
Regional	2.14
Remote	4.73
Senior managers and professionals	2.08
Other managers and associate professionals	2.27
Tradespeople & skilled office, sales and service staff	2.26
Machine operators, labourers, hospitality, and related staff	2.03
Not in paid work in last 12 months	2.58
Bachelor's degree or above	2.05
Advanced diploma/diploma	2.20
Certificate I to IV (inc trade cert)	2.49
Year 12 or equivalent	2.02
Year 11 or equivalent	2.14
Year 10 or equivalent	2.10
Year 9 or equivalent or below	1.13

References

ACARA (Australian Curriculum, Assessment and Reporting Authority) (2020a) NAP–ICT Literacy Assessment Framework 2020, Sydney: ACARA.

ACARA (Australian Curriculum, Assessment and Reporting Authority) (2020b) Measurement Framework for Schooling in Australia 2020, Sydney: ACARA. https://acara.edu.au/reporting/measurement-framework-for-schooling-in-australia

ACARA (Australian Curriculum, Assessment and Reporting Authority) (2022) Data Standards Manual: Student Background Characteristics, Sydney: ACARA. https://acara.edu.au/reporting/data-standards-manual-student-background-characteristics

ACARA. (n.d.) Science Glossary (Version 8.4). https://www.australiancurriculum.edu.au/f-10-curriculum/science/glossary/?letter=S

Adams RJ and Wu ML (2002) PISA 2000 Technical Report, Paris: OECD.

Adams RJ, Wu ML, Cloney D, Berezner A and Wilson M (2020) ACER ConQuest: Generalised Item Response Modelling Software (Version 5.29) [Computer software], Australian Council for Educational Research. https://www.acer.org/au/conquest

Masters GN and Wright BD (1997) "The partial credit model" in WJ Van der Linden and RK Hambleton (eds) *Handbook of Modern Item Response Theory*, pp 101–122, New York/Berlin/Heidelberg: Springer.

Mislevy RJ (1991) "Randomization-based inference about latent variables from complex samples" *Psychometrika*, vol 56, pp 177–196.

Mislevy RJ and Sheehan KM (1987) "Marginal estimation procedures" in AE Beaton (ed) *The NAEP 1983–1984 Technical Report*, pp 293–360, Princeton, NJ: Educational Testing Service.

Muthén, L.K. and Muthén, B.O. (2020). Mplus User's Guide. Version 8.5. Los Angeles, CA: Muthén & Muthén

National Research Council. (2000). Inquiry and the National Science Education Standards: A Guide for Teaching and Learning. Washington, DC: The National Academies. https://doi.org/10.17226/9596

Olson JF, Martin MO and Mullis IVS (eds) (2013) Methods and Procedures in TIMSS & PIRLS 2011, Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Organisation for Economic Cooperation and Development (OECD) (2005) PISA 2003 Technical Report, Paris: OECD.

Organisation for Economic Cooperation and Development (OECD) (2009) PISA Data Analysis Manual SPSS® (2nd edn) Paris: OECD.

Performance Measurement and Reporting Taskforce (PMRT) (2003) "Setting National Standards" paper presented at the March 2003 meeting of the Performance Measurement and Reporting Taskforce.

Rasch G (1960) *Probabilistic Models for Some Intelligence and Attainment Tests*, Copenhagen: Nielsen and Lydiche.

Von Davier M, Gonzalez E and Mislevy R (2009) "What are plausible values and why are they useful?" *IERI Monograph Series*, vol 2, pp 9–36, Hamburg and Princeton: IERI Institute and ETS.

von Davier M and Mazzeo J (2009) Review of PISA test design: Recommendations for fostering stability in assessment results, paper presented at the PISA Research Conference, Kiel, Germany.

Warm TA (1989) "Weighted likelihood estimation of ability in Item Response Theory" *Psychometrika*, vol 54, pp 427–450.

Wolter KM (1985) *Introduction to Variance Estimation*. New York: Springer-Verlag.

Appendices

Appendix A: Student questionnaire

All questions were presented to both Year 6 and Year 10 unless otherwise stated.

Question 1: Year 6 version

1. How much do you agree with the statements below?

Select one choice in each row.

	Strongly agree	Agree	Disagree	Strongly Disagree
I would like to learn more science at school.	○	○	○	○
I think it would be interesting to be a scientist.	○	○	○	○
I enjoy doing science.	○	○	○	○
I enjoy learning new things in science.	○	○	○	○
I learn science topics quickly.	○	○	○	○
I can understand new ideas about science easily.	○	○	○	○
Science is part of my everyday life.	○	○	○	○
Science is important for lots of jobs.	○	○	○	○
Science is important because it changes how we live.	○	○	○	○
Scientific information helps people make good decisions.	○	○	○	○

Question 1: Year 10 version

	Strongly agree	Agree	Disagree	Strongly Disagree
I want to study one or more science subjects in Years 11 and 12.				
I am considering a science-related career.				
I enjoy doing science.				
I enjoy learning new things in science.				
I learn science topics quickly.				
I can understand new ideas about science easily.				
Science is part of my everyday life.				
Science is important for lots of jobs.				
Science is important because it changes how we live.				
Scientific information helps people make good decisions.				

2. How much do you agree with the statements below?

Select one choice in each row.

	Strongly agree	Agree	Disagree	Strongly Disagree
Science is about remembering facts.				
Science is about doing experiments.				
Science is finding out about how things work.				
Science is about solving problems.				
Science is about collaborating with others.				
Science is about making enquiries.				

3. How often do you do these things outside of school?

Select one choice in each row.

	Frequently (more than 2 times a week)	Often (1 or 2 times a week)	Sometimes (less than once a week)	Never
Watch television or stream content about science-related topics				
Read physical or digital books, newspapers or articles about science				
Listen to podcasts, audiobooks or radio on science-related topics				
Talk about science with my friends				
Talk about science with my family				
Post or share content about science-related topics on the internet or social media				
Contribute to existing discussions about science-related topics on the internet or social media				
'Like' someone else's content on science-related topics on the internet or social media				

4. How often do you do these things at school?

Select one choice in each row.

	Frequently (more than 2 times a week)	Often (1 or 2 times a week)	Sometimes (less than once a week)	Never
Watch television or stream content about science-related topics				
Read physical or digital books, newspapers or articles about science				
Talk about science with my friends				

5. How much do you agree with the statements below?

Select one choice in each row.

	Strongly agree	Agree	Disagree	Strongly Disagree
Scientific information helps people make informed decisions.	○	○	○	○
Our scientific knowledge is constantly changing.	○	○	○	○
Science can help us understand global issues that impact on people and the environment.	○	○	○	○
I follow the advice of the scientific community when making decisions related to health crises (e.g. during the COVID-19 pandemic).	○	○	○	○
Government decisions should be based on scientific evidence where available.	○	○	○	○
I know where to find scientific information about local and global issues.	○	○	○	○
I know how to decide whether to trust online information about a science topic.	○	○	○	○

6. Which of these science topics have you studied at school?

Select one choice in each row.

	Yes	No
Earth sciences - for example, weather, soil, rocks, using Earth's resources	○	○
Space (astronomy) - for example, galaxies, objects in space including the planets, Sun and Moon	○	○
Forces and motion - for example, how toys and other machines move and work	○	○
Energy, forms and transfer - for example, electricity, heat, light, sound, magnets	○	○
Living things - for example, how animals and plants survive in their environment, food chains and webs, ecosystems	○	○
Multicellular systems - for example, the human body, cells, tissues, organs, body systems	○	○
Diversity and evolution - for example, how living things change over time	○	○
States of matter - for example, changes to materials (solids, liquids and gases), processes of change such as melting, evaporation	○	○
Properties of matter - characteristics of materials such as density, mass, volume, melting point, hardness, elasticity	○	○

Question 6b: Year 6 only

6b. How often do you have science lessons at school?

Note: A science lesson is a lesson with any teacher where you explore how and why things happen. In science lessons, you do experiments, collect information, or talk about scientific ideas.

Select one choice only.

- More than once a week
- Once a week
- Less than once a week, but more than once a month
- Once a month or less
- Never

Question 7: first bullet point only shown to Year 6

7. Do you agree with the statements below?

Select one choice in each row.

	Yes	No
My classroom teacher teaches science to our class.		
My teacher invites visitors to school to talk about science topics.		
Our class goes on excursions related to the science topics we are learning about.		
My teacher can explain scientific concepts clearly.		

8. How much do you agree with the statements below?

Science is about ...

Select one choice in each row.

	Strongly agree	Agree	Disagree	Strongly Disagree
making observations about the world.				
asking questions about objects and events.				
making predictions and testing them.				
describing patterns and relationships.				
using evidence to develop explanations.				
building knowledge by trial and error.				

9. How much do you agree with the statements below?

Select one choice in each row.

	Strongly agree	Agree	Disagree	Strongly Disagree
People from many different countries have made important contributions to science.	○	○	○	○
Women and men are both involved in science.	○	○	○	○
People from all cultural backgrounds in Australia are involved in science.	○	○	○	○
People of all ages are involved in science.	○	○	○	○
Women and men are equally skilled in science.	○	○	○	○
Female scientists get as much recognition as male scientists.	○	○	○	○

10. How often do the following activities take place in your science lessons?

Select one choice in each row.

	Never	Sometimes	Mostly	Always
My teacher asks us to brainstorm ideas.	○	○	○	○
My teacher helps me identify patterns between different pieces of information.	○	○	○	○
My teacher encourages me to explain the reasons why I did something.	○	○	○	○
My teacher encourages me to think through all the different options when making decisions.	○	○	○	○
I get to plan and carry out my own investigations.	○	○	○	○
I use a computer or tablet for research into science-related topics.	○	○	○	○
Our class has in-depth discussions about science ideas.	○	○	○	○
We work in groups to carry out investigations.	○	○	○	○

11. How much do you agree with the following statements?

My family encourage me to ...

Select one choice in each row.

	Strongly agree	Agree	Disagree	Strongly Disagree
come up with creative solutions to solving problems.	○	○	○	○
question information I find on the internet or TV.	○	○	○	○
consider situations from different perspectives.	○	○	○	○
consider the source of information.	○	○	○	○
explain my reasons for doing something.	○	○	○	○
look at the different parts of a problem to help me solve it.	○	○	○	○

12. How often do you do the following activities outside of school?

Select one choice in each row.

	Never	Sometimes	Mostly	Always
Do activities which require creative solutions	○	○	○	○
Participate in problem solving activities	○	○	○	○
Come up with my own activities to entertain myself	○	○	○	○
Develop new ways to solve problems	○	○	○	○
Debate topics with my family or friends	○	○	○	○

Question 13: Year 10 only

13. How confident are you in undertaking the following activities?

Select one choice in each row.

	Not at all confident	Not very confident	Somewhat confident	Very confident
Making predictions based on prior evidence	○	○	○	○
Identifying what I don't know about a topic, so I understand what I need to learn	○	○	○	○
Identifying patterns and making connections between different pieces of information	○	○	○	○
Testing different options and monitoring the outcomes	○	○	○	○
Thinking about problems from different perspectives	○	○	○	○
Working on tasks that require creative thinking	○	○	○	○
Questioning the accuracy of the source of information I am receiving	○	○	○	○
Explaining where my ideas came from	○	○	○	○

Appendix B: Technical checks – Excerpts from the STSO Manual

The nominated School Technical Support Officer (STSO) at each school was tasked with completing a number of technical checks in order to ensure the school's technical set-up for the assessment was "test-ready". STSOs were asked to complete these tasks in the weeks leading up to the scheduled assessment at their school. The instructions reproduced below are excerpts from the STSO Manual that have been modified slightly to improve readability as an Appendix.

Run a bandwidth test

You must ensure that your school's bandwidth capabilities meet the minimum requirements for the NAP–SL assessment delivery system. Please make a note of the upload and download speed of the bandwidth test you complete so you can include the results in the STSO technical preparations questionnaire. If possible, do more than one bandwidth test and take an average across all tests.

To conduct the bandwidth test, please navigate to any free online speed test tool. There are many bandwidth tests available online but 2 are provided below:

https://speedof.me/
http://www.speedtest.net/

The bandwidth test should be done on a student computer that will be used for the assessment. For accuracy, you should also conduct the bandwidth test during normal school hours, if possible.

If your school's internet connection **does not** meet the following **minimum** requirements:

- 2–3 Mbps download and
- 100–150 Kbps upload

we may need to contact you to discuss running 2 or more smaller test sessions.

Install the locked down browser (LDB) on student devices

Students access the NAP–SL assessment via the locked down browser (LDB), so this must be installed on all devices used by students to take the assessment. The LDB is the same application that is used to sit NAPLAN online testing. Please ensure the LDB is installed on devices used by Year 6 and Year 10 students sitting NAP–SL.

IMPORTANT NOTE: The most recent version of the LDB is needed to access the NAP–SL assessment. If any device already has the LDB installed, you should check that it is not out of date. You can do this by launching the LDB. If the system alerts you that your LDB is out of date, you will need to download a new version.

It is also imperative that the locked down browser is installed on a profile that students will be able to access on the day of the test. The device check must also be conducted using this profile while accessing the internet connection available to students.

1. Open a browser and navigate to https://www.assessform.edu.au/
2. Click on the locked down browser link (Figure 1).

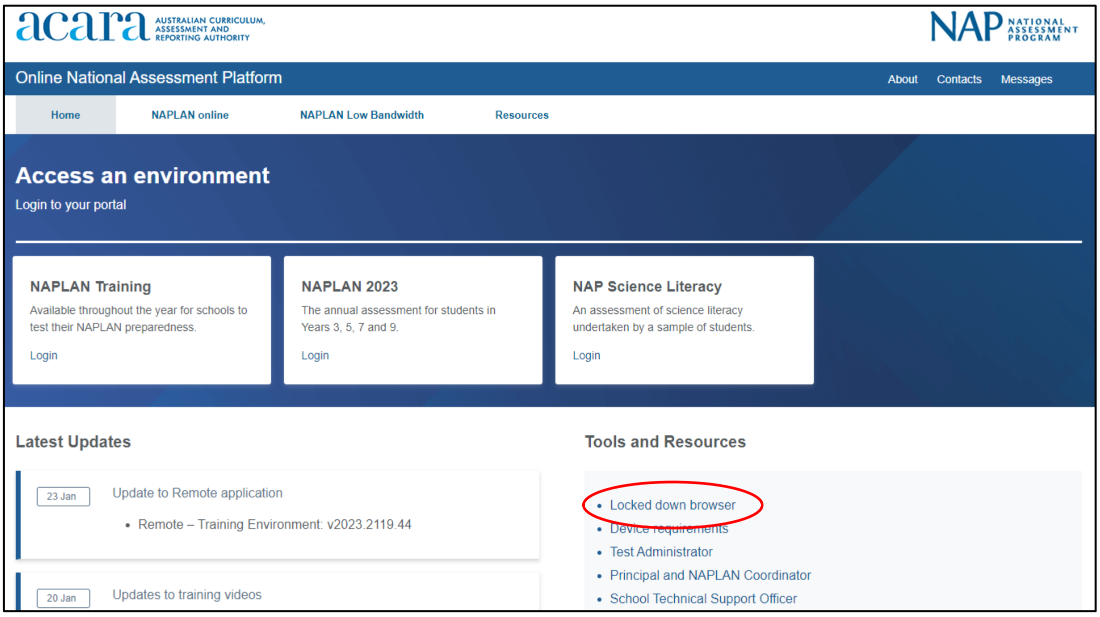

Figure 1: Assessform home page

3. On the locked down browser page (Figure 2), you will find links to the locked down browser user guides and device requirements information. The LDB user guides provide detailed instructions for installing the LDB on a range of different devices. The device requirements page, accessed via the minimum device requirements link, outlines the minimum specifications a device must meet to interact successfully with the online assessment platform. You should check that student devices meet these requirements before downloading the locked down browser onto them.

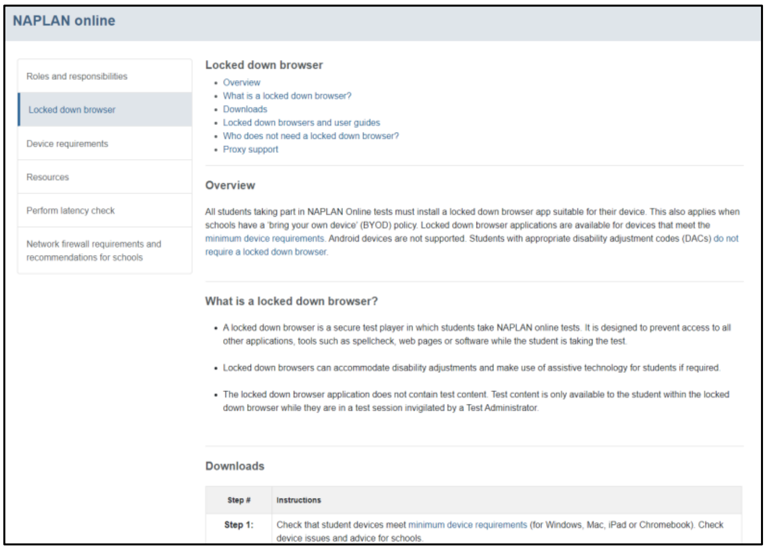

Figure 2: Locked down browser page

4. On the locked down browser page, you will also find download links to the LDB installation files. Click on the appropriate link for the device you are using and install the LDB. Ensure you install the LDB on <u>all</u> devices students will use to take the assessment.

Perform the device check on student devices

To ensure that all student devices will be able to successfully run the assessment, you must perform a device check on each machine. Note: the platform offers several ways to perform a device check. For all student computers used in the NAP–SL assessment, the device check must be performed <u>via the locked down browser</u>. If the check is not performed this way, there is a risk that computers/devices may not be able to access the test event on test day.

IMPORTANT NOTE: even if a machine already had the LDB installed and you did not need to download a new version, you must still perform the device check.

1. Launch the locked down browser.
2. Select *Device check* from the list of options provided (Figure 3). Do not select *Device check (without login)*.

Figure 3: Starting the device check

3. Select *NAP Science Literacy* from the list of server options provided.

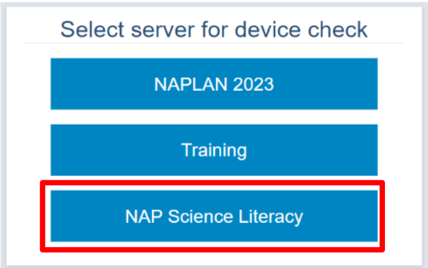

Figure 4: Server selection

4. In the login boxes that appear (Figure 5), enter your STSO username and password (provided in the email sent to you with this document).

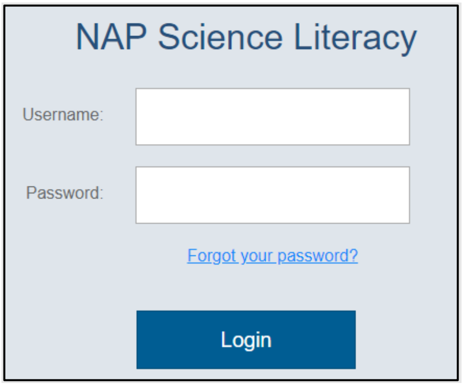

Figure 5: Device check login screen

5. The device check will now run for one to 2 seconds. Once it is complete, you will see a screen similar to the one below (Figure 6).

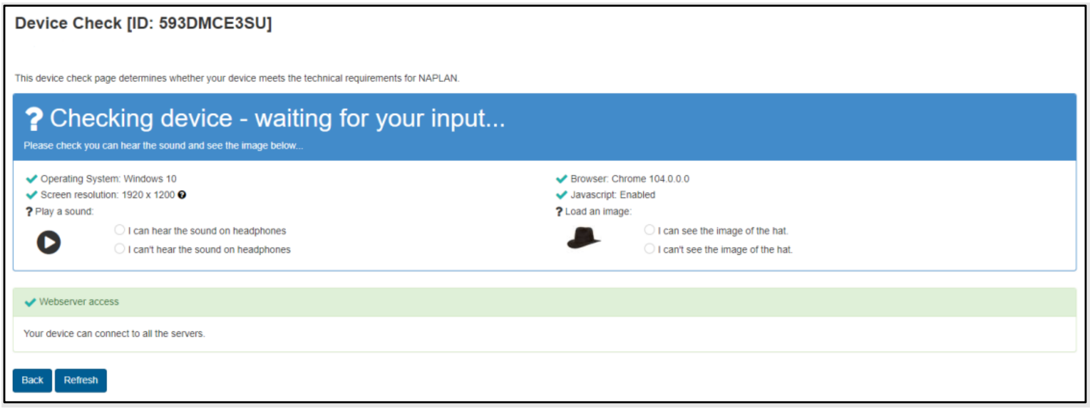

Figure 6: Device check

6. As previously stated, there is no audio component to the NAP–SL assessment. However, to complete the device check, please click on *I can hear the sound on headphones* (Figure 7).

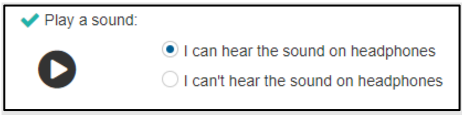

Figure 7: Device check for sound

7. Indicate whether the device can load an image by selecting the appropriate response (Figure 8).

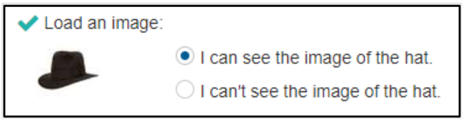

Figure 8: Device check for images

8. The device check is now complete. The device and browser you are using have been checked against the technical requirements for NAP assessments. You should see one of the 2 screens below (Figure 9 and Figure 10).

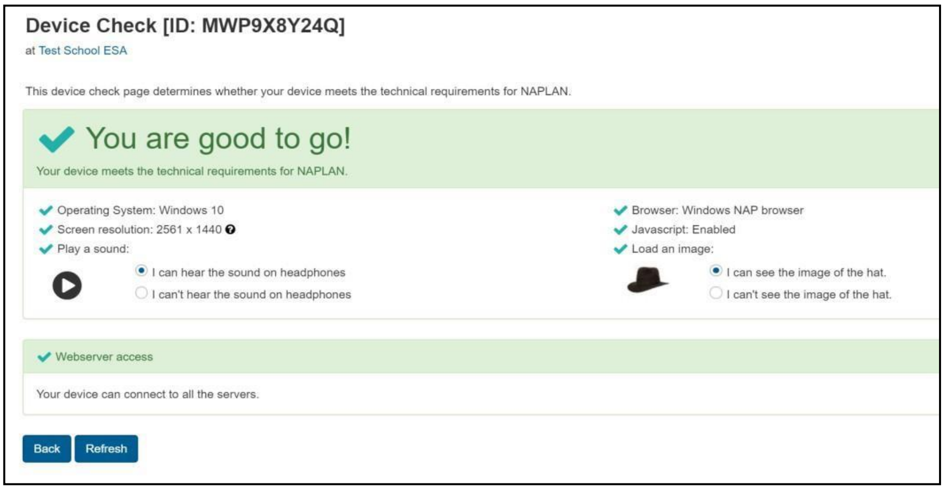

Figure 9: Device check pass

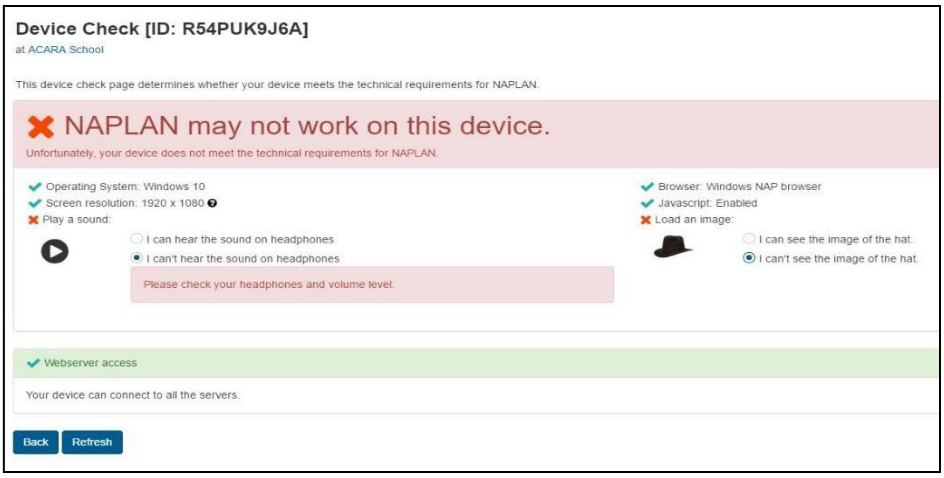

Figure 10: Device check fail

If your device check was successful, please proceed to step 9.

If you receive a fail against an element of the test, please see Section 3 – *Getting help* to assist you in rectifying the problem. Once the device, network or LDB has been updated, please run the device check again.

9. Click the Back button on the device check screen (Figure 11). Your result will be saved. Please note: if you exit the device check by using the grey X in the bottom right corner, your device check will not register in the device list for your school.

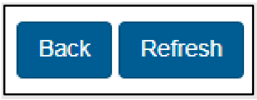

Figure 11: Finishing the device check

10. Exit the app.
11. Repeat steps 1–10 **for every device** that will be used for the NAP–SL assessment.

Ensure a device for the Test Administrator (TA) is prepared

You will need to ensure a device has been set aside for the Test Administrator to use on assessment day. The device can be one the Test Administrator already uses (i.e. a work issued laptop) or a spare computer in the testing room. This device does not need to have the LDB installed, and the device check should be performed outside the LDB.

To run the device check on the TA machine, open the home page of the Assessform website assessform.edu.au and click on *Device Check (no results stored)* in the Tools and resources section on the right (Figure 12) and follow the instructions.

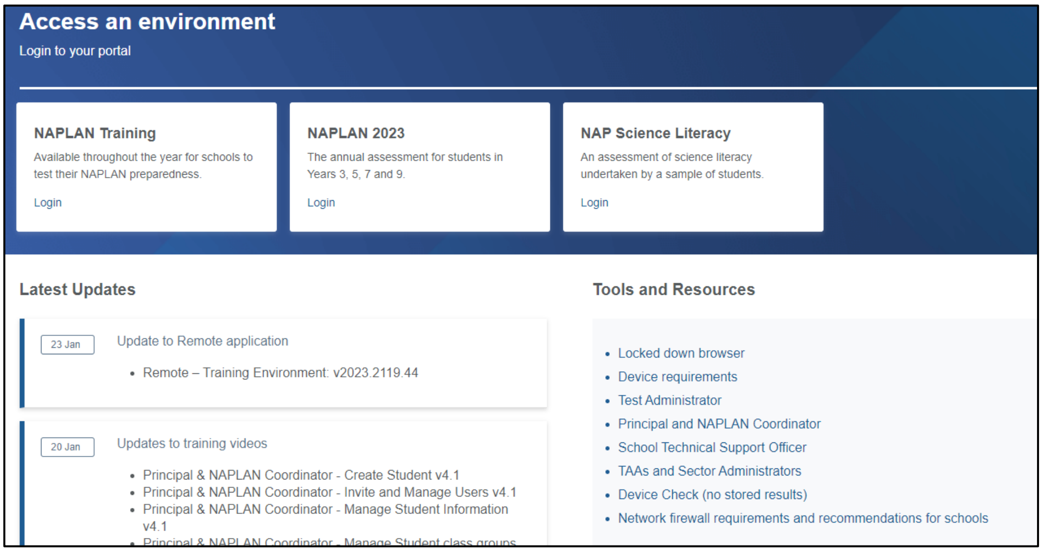

Figure 12: Device check for TA device

Complete a short STSO technical preparations questionnaire to let us know your school's devices are ready

Once you have performed all technical readiness steps (speed test; download, install and checking of the LDB on all student devices; student device check; TA device check) please complete the "STSO technical preparations questionnaire". The specific link to your school's questionnaire can be found in the email that also contained your login details to the Assessform website.

Appendix C: Quality monitor report template

NAP–Science Literacy Main Study 2023
QUALITY MONITOR REPORT

Quality Monitor			
School Name			
State/Territory		**Sector**	
Year Level		**Date**	
Number of Students Present			

1. Staff Present

Who was present for the assessment session? (Please check all that apply and indicate whether they were present for all or part of the test session)

Staff Member	Present for all of session (X)	Present for part of session (X)
Test Administrator		
School Contact		
School Technical Support Officer		
Principal		
Other (please specify) ______________________________		

Were the School Contact and Test Administrator roles held by the same person?

☐ Yes, same person ☐ No, different people

2. Timing

Room Set Up and Logging in

How long did it take for the computers to be switched on and logged into? _____ (mins)

Did the STSO or other school staff member assist the TA in setting up the computers?

☐ No ☐ Yes

Was the room suitably set up for the assessment and for students' optimal participation?

☐ No ☐ Yes

If No, please provide further comment.

____
____
____

Introductory TA script

How long did it take the TA to lead students through the initial assessment instructions, before the practice questions? ______ (mins)

Please detail any issues that were experienced during the introductory process.

____
____
____

How long did it take the students to complete the practice questions, on average?

Please provide further comment if actual time for any student was **significantly** different to expected time of 5 mins.

____
____
____

Assessment Session

Students are given a set time allowance to complete the assessment (60 mins for Y6, 75 mins for Y10). For the majority of students in this test session, was this time allowance:

☐ Too generous ☐ Just right ☐ Too short

How many students were able to complete the assessment in the allocated time?

☐ No students were able to complete in time.
☐ A minority of students were able to complete in time.
☐ The majority of students were able to complete in time.
☐ All students were able to complete in time.

Please provide further comment on test time, if needed.

____
____
____

***Survey* (untimed, but suggested time of 20 mins)**

How long did it take most of the students to complete the survey? ______ (mins)

How long did it take the slowest student to complete the survey? ______ (mins)

Please provide further comment on survey timing, if needed.

____
____
____

3. Test Instructions

Was the script followed according to the Test Administrator Handbook?

☐ No ☐ Yes

If changes were made, were they:

☐ Major ☐ Minor

Why do you think the TA made changes to the script?

____
____
____

Do you think the variation to the script affected the performance of students?

☐ No ☐ Yes

If Yes, please provide further comment.

____
____
____

4. Assistance Given

Were there any particular test questions that students asked for clarification about?

☐ No ☐ Yes

Please provide a general description of the item and a brief description of the issue/clarification requested:

____
____
____

In your opinion, did the Test Administrator follow the instructions in the TA Handbook when assisting students with their questions?

☐ No ☐ Yes

If No, please provide further comment.

____
____
____

Was any extra assistance given to any students with special needs?

☐ No ☐ Yes

If Yes, please provide further comment.

____
____
____

5. Technical Matters

What devices did students use to sit the assessment? (Check all that apply)

☐ Desktop computers
☐ Laptop computers
☐ iPads
☐ Chromebooks

If iPads were used, did students use an external keyboard?

☐ Yes, all iPad users had an external keyboard.
☐ No, no iPad users had an external keyboard. They used the onscreen ('pop up') keyboard instead.
☐ Amongst iPad users, there was a mix of external keyboards and onscreen ('pop up') keyboard use.

Were any technical issues experienced at this school before or during the assessment session?

☐ No ☐ Yes

If Yes, were they:

☐ Major ☐ Minor

If technical issues were experienced, please describe what they were.

____
____
____

Do you think the technical issues affected the performance of students?

☐ No ☐ Yes

If Yes, please provide further comment.

____
____
____

6. Student Behaviour	No students	Some students	Most students
a) How many students appeared to be engaged in the test material?	☐	☐	☐
b) How many students made noise or moved around, causing disruption to other students during the session?	☐	☐	☐
c) How many students attempted to navigate to other websites or access their mobile phones during the session?	☐	☐	☐
d) How many students appeared to struggle with understanding how to navigate the test interface?	☐	☐	☐

7. Outside Interruptions

Were the students distracted or impacted by any outside interruptions? For example:

- Announcements over the PA or intercom system
- Noise from other classes in the school
- Distractions from other students not participating in the test session within the classroom
- Students or teachers visiting the testing room

☐ No ☐ Yes

If yes, please specify the disruption:

____
____
____

8. School Receptiveness

How receptive was the school towards participating in NAP–Science Literacy? What do you perceive to be the school's overall attitude and level of commitment towards the assessment?

____
____
____

As a visitor, were you made to feel welcome by the school?

____
____
____

9. Other Comments

Please provide any other comments that you feel would help us improve this assessment and its administration.

____
____
____
____
____
____

Appendix D: School summary report instructions

The NAP Science Literacy 2023 report for your school is provided on the Report tab of this spreadsheet.

Below is a brief description of the contents of each of the fields shown in this report.

Unit Name	The name of the unit the question belongs to. Each unit contains between 2 and 15 questions that are developed around a single theme or stimulus. Students' test forms are made up of a selection of these units.
Descriptor	A brief description of what students need to do in order to complete a question. Each row refers to a single question in the assessment.
AC Science Strand	Refers to the Australian Curriculum: Science strand that the question aligns with. Hovering over the data cell will display the full strand name.
AC Science Sub-Strand	Refers to the Australian Curriculum: Science sub-strand that the question aligns with. Hovering over the data cell will display the full sub-strand name.
AC Science Code	Refers to the Australian Curriculum: Science (v9) code that the question aligns with.
Cognitive Dimension	Refers to one of the three NAP Science Literacy Assessment Framework cognitive dimensions assessed by each question. Hovering over the data cell will display the dimension.
CCT Element	Refers to one of the four elements of the Critical and Creative Thinking learning continuum that the question aligns with, where applicable.
Percent Correct	Shows an estimate of the national percentage of students who responded to the question correctly. For questions with a maximum score of more than 1, you will see more than one percentage. Each percentage denotes the number of students that reached each score or higher. For example, if a task has a maximum score of 2, the first number is the percentage of students that received a score of 1 or 2, the second number is the percentage of students that received a score of 2.
Max Score	Shows the maximum score possible for each question.

The score for each question is listed under the name of each student. There are four possible score values for each question:

i. Blank: The question was not in a unit assigned to the student.

ii. Red (0): The student responded to the question incorrectly.

iii. Green (1, 2, 3): The student responded to the question correctly (or partially correctly). The number refers to the score the student received for their response to the question. This can be compared to the maximum score for that task.

iv. Grey (N): The question was assigned to that student, but the student did not provide a response.

The report has a set of clickable filters, so you can manipulate how you would like to view the data. For example, view students grouped by unit, content area or score awarded.

Appendix E: School summary report (excerpt from a sample report)

acara AUSTRALIAN CURRICULUM, ASSESSMENT AND REPORTING AUTHORITY

ACER

Unit Name	Descriptor	AC Science Strand	AC Science Sub-Strand	AC Science Code	Cognitive Dimension	CCT Element	Percent Correct	Max Score	Student 01 F 113000204	Student 02 F 113000206	Student 03 M 113000207	Student 04 F 113000208	Student 05 M 113000209	Student 06 F 113000210	Student 07 M 113000211	Student 08 F 113000212	Studen M 113000
Fleas	Explains an advantage of a parasite not killing its host.	SU	BS	AC9S4U01	RAE		69	1		1					1		
Fleas	Extracts information from a life cycle diagram.	SU	BS	AC9S3U01	RAE		81	1		1					1		
Fleas	Identifies a parasitic relationship between two organisms.	SU	BS	AC9S4U01	KUP		75	1		1					1		
Floating top	Understands that a magnet can repel another magnet.	SU	PS	AC9S4U04	RAE		54	1	1				1				1
Floating top	Identifies the opposing forces of gravity and air resistance acting on an object.	SU	PS	AC9S4U03	RAE	Inquiring	17	1	1				0				1
Floating top	Identifies the aim of an investigation by looking at the outcomes.	SI	QP	AC9S4I01	RAE		73	1	1				1				1
Floating top	Evaluates conclusions from an investigation to identify and explain an incorrect conclusion.	SI	E	AC9S5I05	RAE	Analysing	72,40	2	1				1				2
Generating electricity	Explains why the Sun is essential for human survival.	SU	ESS	AC9S1U01	KUP	Reflecting	87	1			1						
Generating electricity	Identifies the labels for a pie graph using information from a text.	SI	PMA	AC9S5I04	KUP		78	1			0						
Generating electricity	Analyses a graph to identify the renewable energy source that generates the most electricity.	SI	PMA	AC9S4I04	KUP		96	1			1						
Generating electricity	Identifies the energy transformations that occurs in solar panels.	SU	PS	AC9S6U03	KUP		87	1			0						
Generating electricity	Orders the steps to explain how solar panels work.	SHE	UIS	AC9S6H02	RAE	Inquiring	40	1			0						
Glacier	Explains the impact of an investigation on a real-world application.	SHE	UIS	AC9S10H03	SC	Analysing	72,30,3	3		2				2			
Glacier inquiry	Sequences the steps in the formation of an artificial glacier.	SHE	UIS	AC9S10H04	RAE	Analysing	42	1		1				0			
Glacier inquiry	Identifies the scientific question for an investigation.	SI	QP	AC9S6I01	SC	Inquiring	59	1		1				0			
Glacier inquiry	Selects the most accurate piece of equipment to measure volume.	SI	PC	AC9S8I03	SC		85	1		1				1			
Glacier inquiry	Interprets data from a table to draw a conclusion.	SI	PMA	AC9S4I04	SC		91	1		1				1			
Glacier inquiry	Identifies a way to improve reliability in an investigation.	SI	E	AC9S6I05	SC	Analysing	71	1		1				1			
Glacier inquiry	Identifies the independent variable in an investigation.	SI	PC	AC9S6I02	KUP	Analysing	51	1		1				1			
Glacier inquiry	Identifies a reason for multiple trials in an investigation.	SI	E	AC9S6I05	SC	Inquiring	77	1		1				1			
Glacier inquiry	Calculates the missing average in a data table.	SI	PMA	AC9S6I04	RAE		75	1		1				1			
Glacier inquiry	Suggests the cause for an abnormal result in a trial.	SI	E	AC9S6I05	RAE	Analysing	71	1		1				1			
Glacier inquiry	Identifies conclusions that can be drawn from information in a graph.	SI	PMA	AC9S6I04	RAE	Analysing	27	1		1				1			
Glaciers	Recognises the effect of freezing water within rocks.	SU	ESS	AC9S5U02	RAE	Reflecting	57	1					1				

Appendix F: Item difficulties

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00118614	1	Year 6	Yes	-2.16	-1.67	226	-2.16	226					87		0.92	
x00118666	1	Year 6	Yes	1.51	2.00	611	1.51	611					23		0.92	
x00118667	1	Year 6	Yes	-1.22	-0.73	324	-1.22	324					75		0.95	
x00118697	1	Year 6	Yes	-2.42	-1.93	199	-2.42	199					90		0.81	
x00118710	1	Year 6	Yes	-0.79	-0.30	370	-0.79	370					68		0.90	
x00119001	2	Year 6	Yes	-0.04	0.45	448	-0.46	404	0.37	492			53		0.99	
x00119003	1	Year 6	Yes	-0.29	0.20	422	-0.29	422					58		1.11	
x00119010	1	Year 6	Yes	0.13	0.62	466	0.13	466					49		1.00	
x00119011	1	Year 6	Yes	1.46	1.95	605	1.46	605					23		0.99	
x00119016	2	Year 6	Yes	2.03	2.52	665	0.94	551	3.12	779			18		1.07	
x00121241	1	Year 6	Yes	-2.22	-1.73	220	-2.22	220					88		1.02	
x00121243	1	Year 6	Yes	-1.42	-0.93	304	-1.42	304					79		0.96	
x00121244	1	Year 6	Yes	-2.20	-1.71	222	-2.20	222					88		0.87	
x00121245	1	Year 6	Yes	-0.89	-0.40	359	-0.89	359					70		0.97	
x00121255	1	Year 6	Yes	-1.61	-1.12	284	-1.61	284					81		1.05	
x00121259	1	Year 6	Yes	-1.91	-1.42	252	-1.91	252					85		1.08	
x00121260	1	Year 6	Yes	-4.48	-3.99	-17	-4.48	-17					98		1.04	
x00121263	1	Year 6	Yes	-1.04	-0.55	344	-1.04	344					72		0.86	

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00121264	1	Year 6	Yes	-0.47	0.02	403	-0.47	403					62		0.93	
x00121338	1	Year 6	Yes	-1.42	-0.93	304	-1.42	304					78		0.98	
x00121341	1	Year 6	Yes	-0.45	0.04	405	-0.45	405					61		1.00	
x00121342	1	Year 6	Yes	-0.73	-0.24	375	-0.73	375					67		1.02	
x00121586	1	Year 6	Yes	-1.31	-0.82	315	-1.31	315					77		0.91	
x00121588	1	Year 6	Yes	-1.59	-1.10	286	-1.58	286					81		0.83	
x00121590	1	Year 6	Yes	0.10	0.59	463	0.10	463					50		1.06	
x00195293	1	Year 6	No	0.57	1.06	513	0.57	513					40		0.89	
x00195309	1	Year 6	No	-0.11	0.38	441	-0.11	441					54		1.00	
x00195318	1	Year 6	No	-2.51	-2.02	190	-2.51	190					91		0.97	
x00195324	1	Year 6	No	-0.56	-0.07	394	-0.56	394					63		0.85	
x00195331	2	Year 6	No	0.25	0.74	479	-0.07	445	0.58	513			45		1.02	
x00195332	1	Year 6	No	0.89	1.38	546	0.89	546					34		1.08	
x00195333	1	Year 6	No	0.76	1.25	532	0.75	531					36		0.93	
x00195334	1	Year 6	No	-0.83	-0.34	366	-0.83	366					69		0.85	
x00195340	1	Year 6	No	-0.54	-0.05	396	-0.54	396					63		1.05	
x00195343	1	Year 6	No	0.04	0.53	457	0.04	457					51		0.87	
x00195382	1	Year 6	No	-0.43	0.06	408	-0.43	408					61		1.02	
x00195392	1	Year 6	No	1.42	1.91	601	1.42	601					25		1.01	
x00195462	1	Year 6	No	0.69	1.18	525	0.69	525					38		0.88	

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00195465	1	Year 6	No	-0.24	0.25	427	-0.24	427					57		0.94	
x00195466	1	Year 6	No	0.95	1.44	552	0.95	552					32		0.83	
x00195467	1	Year 6	No	-0.21	0.28	430	-0.21	430					57		0.82	
x00195468	2	Year 6	No	0.40	0.89	495	0.10	463	0.71	527			42		1.30	
x00195469	1	Year 6	No	0.40	0.89	494	0.40	494					44		0.95	
x00195470	2	Year 6	No	0.43	0.92	497	-0.81	368	1.66	626			44		1.00	
x00195471	1	Year 6	No	-1.06	-0.57	341	-1.06	341					72		0.96	
x00195472	1	Year 6	No	-0.32	0.17	419	-0.32	419					58		0.94	
x00195476	1	Year 6	No	2.00	2.49	662	2.00	662					16		0.97	
x00195477	1	Year 6	No	2.86	3.35	752	2.86	752					8		0.94	
x00195834	2	Year 6	No	2.10	2.59	672	1.76	636	2.44	708			9		0.98	
x00195835	1	Year 6	No	-0.98	-0.49	350	-0.98	350					70		1.01	
x00195841	1	Year 6	No	0.30	0.79	484	0.30	484					45		1.17	
x00195842	1	Year 6	No	0.36	0.85	490	0.36	490					44		0.96	
x00195843	1	Year 6	No	-1.28	-0.79	318	-1.28	318					75		1.00	
x00195844	1	Year 6	No	-0.29	0.20	423	-0.29	423					57		1.03	
x00195845	1	Year 6	No	-0.20	0.29	432	-0.20	432					55		0.90	
x00195846	1	Year 6	No	0.41	0.90	495	0.41	495					43		0.87	
x00196596	1	Year 6	No	0.10	0.59	463	0.10	463					50		1.03	
x00196622	1	Year 6	No	-1.21	-0.72	326	-1.21	326					75		0.98	

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00196645	1	Year 6	No	0.70	1.19	526	0.70	526					38		1.16	
x00196646	1	Year 6	No	1.19	1.68	577	1.19	577					29		0.98	
x00196647	1	Year 6	No	-0.13	0.36	439	-0.13	439					55		1.09	
x00196648	1	Year 6	No	0.12	0.61	465	0.12	465					50		1.03	
x00196725	1	Year 6	No	-1.30	-0.81	316	-1.30	316					76		1.00	
x00196728	1	Year 6	No	-2.36	-1.87	205	-2.36	205					89		1.00	
x00196729	1	Year 6	No	-2.02	-1.53	241	-2.02	241					85		0.89	
x00196730	1	Year 6	No	-1.06	-0.57	341	-1.06	341					72		1.11	
x00196732	1	Year 6	No	-2.37	-1.88	205	-2.37	205					89		0.98	
x00196733	1	Year 6	No	-1.66	-1.17	279	-1.66	279					81		0.86	
x00196737	1	Year 6	No	-1.04	-0.55	344	-1.04	344					72		1.00	
x00196738	1	Year 6	No	-0.06	0.43	447	-0.06	447					53		1.03	
x00196739	1	Year 6	No	-1.29	-0.80	317	-1.29	317					76		0.89	
x00196741	2	Year 6	No	0.77	1.26	533	0.20	473	1.34	593			34		0.99	
x00196772	1	Year 6	No	1.72	2.21	632	1.72	632					20		0.90	
x00196774	1	Year 6	No	-1.25	-0.76	322	-1.25	321					75		1.04	
x00196775	1	Year 6	No	0.40	0.89	494	0.40	494					43		1.00	
x00196776	1	Year 6	No	-0.19	0.30	432	-0.19	432					55		0.96	
x00196779	2	Year 6	No	1.83	2.32	645	1.27	586	2.39	703			15		0.93	
x00196851	1	Year 6	No	-2.25	-1.76	216	-2.25	216					88		0.99	

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00196879	1	Year 6	No	0.98	1.47	555	0.98	555					32		0.97	
x00196880	1	Year 6	No	-0.95	-0.46	353	-0.95	353					71		1.08	
x00196881	1	Year 6	No	-1.11	-0.62	336	-1.11	336					74		0.99	
x00196882	1	Year 6	No	0.32	0.81	486	0.32	486					46		1.04	
x00196939	1	Year 6	No	-1.53	-1.04	292	-1.53	292					80		1.00	
x00196940	1	Year 6	No	-0.40	0.09	411	-0.40	411					60		1.23	
x00196941	1	Year 6	No	0.23	0.72	476	0.23	476					47		1.01	
x00196942	1	Year 6	No	0.59	1.08	514	0.59	514					39		1.02	
x00196944	1	Year 6	No	-0.65	-0.16	385	-0.65	385					65		1.02	
x00196945	1	Year 6	No	0.56	1.05	511	0.56	511					40		0.83	
x00196985	1	Year 6	No	-0.35	0.14	416	-0.34	416					60		1.14	
x00196987	1	Year 6	No	-1.33	-0.84	314	-1.33	314					78		1.00	
x00196988	1	Year 6	No	0.24	0.73	478	0.25	478					47		1.04	
x00196989	1	Year 6	No	-1.99	-1.50	244	-1.99	244					86		0.92	
x00197147	1	Year 6	No	1.57	2.06	617	1.57	617					22		1.01	
x00197153	1	Year 6	No	0.79	1.28	535	0.79	535					36		1.12	
x00197154	1	Year 6	No	-1.04	-0.55	344	-1.04	344					73		1.01	
x00197155	1	Year 6	No	0.94	1.43	551	0.94	551					33		1.19	
x00197164	1	Year 6	No	1.06	1.55	563	1.06	563					30		1.03	
x00197165	1	Year 6	No	0.34	0.83	488	0.34	488					45		0.93	

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00197420	1	Year 6	No	-0.99	-0.50	348	-0.99	348					71		1.05	
x00197421	1	Year 6	No	0.85	1.34	541	0.85	541					34		1.01	
x00197423	1	Year 6	No	-0.36	0.13	415	-0.36	415					59		0.99	
x00197424	1	Year 6	No	0.32	0.81	486	0.32	486					45		0.98	
x00197438	1	Year 6	No	-1.10	-0.61	337	-1.10	337					74		0.90	
x00197439	1	Year 6	No	-0.75	-0.26	374	-0.75	374					67		0.89	
x00197447	1	Year 6	No	0.40	0.89	495	0.40	495					43		1.14	
x00197448	1	Year 6	No	-0.65	-0.16	384	-0.65	384					65		0.98	
x00204859	2	Year 6	No	1.62	2.11	622	1.48	608	1.75	636			13		1.16	
x00204907	1	Year 6	No	-1.50	-1.01	295	-1.50	295					80		0.90	
x00204988	1	Year 6	No	1.94	2.43	656	1.94	656					17		0.95	
x00207573	1	Year 6	No	-0.50	-0.01	400	-0.50	400					63		0.96	
x00207574	1	Year 6	No	1.56	2.05	616	1.56	616					22		1.04	
x00207584	1	Year 6	No	-1.59	-1.10	286	-1.59	286					81		0.96	
x00207585	1	Year 6	No	0.37	0.86	491	0.37	491					44		1.15	
x00207604	1	Year 6	No	0.76	1.25	532	0.76	532					36		1.06	
x00207605	1	Year 6	No	-0.60	-0.11	389	-0.60	389					65		1.07	
x00197142_Y6	1	Year 6	No	0.78	1.27	534	0.78	535					35		1.20	
x00195629_Y6	1	Year 6	No	-0.60	-0.11	390	-0.60	390					63		1.14	

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00194782_Y6	1	Year 6	No	0.20	0.69	474	0.20	474					47		1.12	
x00121360_Y6	1	Year 6	Yes	1.18	1.67	575	1.18	575					28		1.17	
x00120785_Y6	1	Year 6	Yes	1.24	1.73	582	1.24	582					27		0.98	
x00121411_Y6	1	Year 6	Yes	0.54	1.03	509	0.54	509					40		0.91	
x00195456_Y6	1	Year 6	No	-0.74	-0.25	375	-0.74	375					67		0.98	
x00195634_Y6	1	Year 6	No	-0.43	0.06	407	-0.43	407					60		1.09	
x00196708_Y6	1	Year 6	No	2.54	3.03	718	2.54	718					10		0.95	
x00195632_Y6	1	Year 6	No	0.06	0.55	459	0.06	459					50		1.01	
x00194751_Y6	1	Year 6	No	0.09	0.58	462	0.09	462					49		1.02	
x00158739_Y6	1	Year 6	No	2.16	2.65	679	2.16	679					14		0.87	
x00197427_Y6	1	Year 6	No	0.32	0.81	486	0.32	486					44		1.10	
x00195628_Y6	1	Year 6	No	0.05	0.54	457	0.05	457					50		0.98	
x00120408	1	Year 10	Yes	1.31	1.80	589	1.31	589						43		1.05
x00120414	1	Year 10	Yes	0.55	1.04	510	0.55	510						59		0.98
x00120453	1	Year 10	Yes	-0.84	-0.35	364	-0.84	364						82		0.94

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00120504	2	Year 10	Yes	1.85	2.34	646	0.92	549	2.79	744				33		1.07
x00120512	1	Year 10	Yes	-1.87	-1.38	257	-1.87	257						92		0.99
x00120514	1	Year 10	Yes	0.21	0.70	474	0.21	474						65		0.82
x00120517	1	Year 10	Yes	1.03	1.52	560	1.03	560						49		0.95
x00120519	1	Year 10	Yes	-0.22	0.27	430	-0.22	430						73		0.84
x00120520	1	Year 10	Yes	-0.46	0.03	404	-0.46	404						77		1.05
x00120524	1	Year 10	Yes	0.01	0.50	453	0.01	453						69		1.02
x00120536	1	Year 10	Yes	2.26	2.75	689	2.26	689						25		1.16
x00120540	3	Year 10	Yes	1.84	2.33	645	-0.01	452	1.68	629	3.84	855		36		0.81
x00120784	1	Year 10	Yes	-0.85	-0.36	364	-0.85	364						82		0.83
x00120785_ Y10	1	Year 10	Yes	0.59	1.08	514	0.59	514						58		0.97
x00120940	1	Year 10	Yes	0.75	1.24	531	0.75	531						57		0.92
x00120954	3	Year 10	Yes	2.78	3.27	744	0.61	516	2.17	680	5.56	1035		29		1.03
x00120956	1	Year 10	Yes	-1.14	-0.65	333	-1.14	334						86		1.01
x00120957	1	Year 10	Yes	2.23	2.72	686	2.23	686						28		0.97
x00120965	1	Year 10	Yes	1.42	1.91	601	1.42	601						43		0.87
x00120970	1	Year 10	Yes	-1.03	-0.54	345	-1.03	345						85		0.94
x00120977	1	Year 10	Yes	-0.28	0.21	423	-0.28	423						75		1.23
x00120985	1	Year 10	Yes	0.45	0.94	500	0.45	500						61		0.99

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00120986	1	Year 10	Yes	0.91	1.40	548	0.92	548						51		1.03
x00120987	1	Year 10	Yes	-0.24	0.25	427	-0.24	427						74		1.03
x00120988	1	Year 10	Yes	-0.83	-0.34	366	-0.83	366						83		1.13
x00121048	1	Year 10	Yes	-1.51	-1.02	295	-1.51	295						90		1.04
x00121051	1	Year 10	Yes	-0.66	-0.17	383	-0.66	383						81		0.96
x00121052	1	Year 10	Yes	-1.87	-1.38	256	-1.87	256						93		1.31
x00121056	1	Year 10	Yes	0.69	1.18	525	0.69	525						58		0.93
x00121057	1	Year 10	Yes	3.48	3.97	817	3.48	817						11		1.10
x00121083	1	Year 10	Yes	-1.59	-1.10	286	-1.59	286						90		0.92
x00121094	1	Year 10	Yes	1.89	2.38	650	1.89	650						34		1.14
x00121095	1	Year 10	Yes	0.35	0.84	489	0.35	489						64		1.14
x00121097	1	Year 10	Yes	-0.94	-0.45	354	-0.94	354						84		1.01
x00121099	1	Year 10	Yes	3.38	3.87	806	3.38	806						13		1.14
x00121166	1	Year 10	Yes	-0.28	0.21	423	-0.28	424						76		1.10
x00121174	1	Year 10	Yes	0.60	1.09	516	0.60	516						61		1.04
x00121177	1	Year 10	Yes	-1.98	-1.49	245	-1.98	245						94		1.13
x00121178	1	Year 10	Yes	1.33	1.82	591	1.33	591						46		1.19
x00121360_Y10	1	Year 10	Yes	1.89	2.38	650	1.89	650						31		1.19
x00121411_Y10	1	Year 10	Yes	-0.10	0.39	442	-0.10	442						70		0.95

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00121525	1	Year 10	Yes	0.99	1.48	556	0.99	556						50		1.06
x00121623	1	Year 10	Yes	0.56	1.05	511	0.56	511						60		0.95
x00158739_Y10	1	Year 10	No	1.58	2.07	618	1.58	618						39		0.95
x00158894	1	Year 10	No	0.62	1.11	518	0.62	518						60		0.98
x00158926	1	Year 10	No	4.21	4.70	893	4.21	893						6		1.03
x00158938	1	Year 10	No	0.92	1.41	549	0.92	549						54		1.04
x00159081	1	Year 10	No	-0.22	0.27	430	-0.22	430						73		0.84
x00194751_Y10	1	Year 10	No	-0.48	0.01	402	-0.48	402						78		0.85
x00194756	1	Year 10	No	0.57	1.06	512	0.57	512						60		1.00
x00194782_Y10	1	Year 10	No	0.99	1.48	556	0.99	556						51		1.20
x00195404	1	Year 10	No	-0.13	0.36	439	-0.13	439						74		0.95
x00195405	1	Year 10	No	-0.45	0.04	406	-0.45	406						79		1.02
x00195406	1	Year 10	No	0.38	0.87	492	0.38	492						65		0.91
x00195409	1	Year 10	No	1.09	1.58	567	1.09	567						51		1.02
x00195410	1	Year 10	No	1.24	1.73	582	1.24	582						48		0.99
x00195411	2	Year 10	No	2.96	3.45	762	2.08	671	3.83	853				17		0.94
x00195412	1	Year 10	No	0.62	1.11	517	0.62	517						61		1.11
x00195413	2	Year 10	No	1.41	1.90	600	1.04	561	1.78	639				43		0.88
x00195432	1	Year 10	No	-1.46	-0.97	300	-1.46	300						90		0.83

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00195434	1	Year 10	No	1.16	1.65	574	1.16	574						50		0.87
x00195456_Y10	1	Year 10	No	-0.10	0.39	442	-0.10	442						71		1.07
x00195459	2	Year 10	No	1.25	1.74	584	0.51	506	1.99	661				44		1.06
x00195557	2	Year 10	No	1.84	2.33	645	-0.10	443	3.77	847				41		0.96
x00195558	1	Year 10	No	1.59	2.08	619	1.59	619						40		0.84
x00195560	1	Year 10	No	-1.36	-0.87	310	-1.36	310						89		0.77
x00195576	1	Year 10	No	2.07	2.56	669	2.07	669						31		1.06
x00195577	1	Year 10	No	-0.47	0.02	404	-0.47	403						79		0.89
x00195578	2	Year 10	No	1.97	2.46	659	0.60	516	3.33	802				36		0.86
x00195581	1	Year 10	No	1.24	1.73	582	1.24	582						47		1.06
x00195582	2	Year 10	No	0.36	0.85	490	0.07	460	0.65	521				68		0.94
x00195583	1	Year 10	No	1.36	1.85	595	1.36	594						45		0.99
x00195586	1	Year 10	No	1.65	2.14	625	1.65	625						39		0.95
x00195628_Y10	1	Year 10	No	-0.44	0.05	406	-0.44	406						76		1.02
x00195629_Y10	1	Year 10	No	0.42	0.91	496	0.42	496						61		1.23
x00195632_Y10	1	Year 10	No	-0.52	-0.03	398	-0.52	398						77		0.92
x00195634_Y10	1	Year 10	No	0.26	0.75	480	0.27	480						64		1.07
x00195840	1	Year 10	No	-0.21	0.28	430	-0.21	430						73		1.09

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00196008	1	Year 10	No	1.42	1.91	601	1.42	601						41		1.17
x00196010	1	Year 10	No	1.78	2.27	639	1.78	639						34		1.02
x00196036	1	Year 10	No	-0.12	0.37	440	-0.12	440						72		0.90
x00196038	1	Year 10	No	2.44	2.93	708	2.44	708						23		0.98
x00196039	2	Year 10	No	2.30	2.79	693	1.54	613	3.07	773				24		1.14
x00196557	1	Year 10	No	1.16	1.65	574	1.16	574						47		1.13
x00196558	1	Year 10	No	3.86	4.35	857	3.86	857						8		0.95
x00196559	1	Year 10	No	2.81	3.30	747	2.81	747						18		1.13
x00196560	1	Year 10	No	2.29	2.78	692	2.29	692						26		0.92
x00196562	2	Year 10	No	3.55	4.04	825	3.03	770	4.07	879				7		0.87
x00196574	1	Year 10	No	2.77	3.26	742	2.77	742						18		0.93
x00196575	1	Year 10	No	1.22	1.71	581	1.22	581						44		1.05
x00196577	1	Year 10	No	3.69	4.18	839	3.69	839						9		0.96
x00196578	1	Year 10	No	3.43	3.92	812	3.43	812						11		0.96
x00196581	1	Year 10	No	3.12	3.61	779	3.12	779						14		1.11
x00196583	1	Year 10	No	2.26	2.75	690	2.26	690						25		1.11
x00196590	1	Year 10	No	-0.03	0.46	450	-0.03	450						69		1.04
x00196593	1	Year 10	No	0.33	0.82	488	0.34	488						62		0.97
x00196594	1	Year 10	No	3.67	4.16	836	3.67	836						9		0.97
x00196655	1	Year 10	No	0.78	1.27	534	0.78	534						56		0.97

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00196656	1	Year 10	No	0.38	0.87	492	0.38	492						64		1.04
x00196660	1	Year 10	No	0.70	1.19	525	0.70	525						58		0.92
x00196664	2	Year 10	No	0.04	0.53	457	-0.21	430	0.30	484				76		1.08
x00196665	1	Year 10	No	-0.27	0.22	424	-0.27	424						76		0.94
x00196673	1	Year 10	No	0.97	1.46	554	0.97	554						52		0.95
x00196674	1	Year 10	No	1.82	2.31	643	1.82	643						35		1.10
x00196697	1	Year 10	No	-0.85	-0.36	364	-0.85	364						84		0.96
x00196704	1	Year 10	No	0.21	0.70	475	0.21	475						67		1.01
x00196708_Y10	1	Year 10	No	1.83	2.32	644	1.83	644						37		0.97
x00196712	1	Year 10	No	0.18	0.67	471	0.18	471						67		1.05
x00196718	1	Year 10	No	1.22	1.71	580	1.22	580						49		0.95
x00196719	1	Year 10	No	0.16	0.65	469	0.16	469						68		1.07
x00196749	2	Year 10	No	0.82	1.31	538	0.43	497	1.21	579				54		0.95
x00196809	1	Year 10	No	0.67	1.16	523	0.67	523						56		1.18
x00196846	1	Year 10	No	-1.32	-0.83	315	-1.32	315						88		0.94
x00196963	1	Year 10	No	0.59	1.08	514	0.59	514						60		1.00
x00196971	1	Year 10	No	1.78	2.27	639	1.78	639						36		0.93
x00197142_Y10	1	Year 10	No	1.91	2.40	652	1.91	652						34		1.18
x00197159	1	Year 10	No	0.66	1.15	522	0.67	522						59		1.03

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00197160	1	Year 10	No	2.99	3.48	766	2.99	766						16		1.07
x00197161	1	Year 10	No	0.80	1.29	536	0.80	536						56		1.28
x00197188	1	Year 10	No	0.94	1.43	550	0.94	550						52		1.13
x00197189	1	Year 10	No	1.54	2.03	614	1.54	614						39		0.91
x00197191	1	Year 10	No	-1.19	-0.70	328	-1.19	328						87		0.93
x00197192	1	Year 10	No	2.94	3.43	761	2.94	761						16		1.00
x00197193	2	Year 10	No	2.00	2.49	662	1.19	577	2.81	747				30		1.05
x00197361	2	Year 10	No	1.51	2.00	611	0.84	541	2.18	680				38		1.00
x00197365	1	Year 10	No	1.08	1.57	566	1.08	566						47		0.96
x00197368	2	Year 10	No	4.32	4.81	905	2.46	710	6.19	1100				11		0.99
x00197378	1	Year 10	No	3.28	3.77	796	3.28	796						12		0.85
x00197380	1	Year 10	No	1.57	2.06	616	1.57	616						38		0.85
x00197390	1	Year 10	No	0.99	1.48	556	0.99	556						50		0.89
x00197392	1	Year 10	No	3.42	3.91	811	3.42	811						11		1.02
x00197394	1	Year 10	No	3.69	4.18	839	3.69	839						8		1.29
x00197401	1	Year 10	No	1.06	1.55	564	1.06	564						48		1.08
x00197402	1	Year 10	No	0.54	1.03	509	0.54	509						59		1.00
x00197404	1	Year 10	No	1.25	1.74	583	1.25	583						44		1.15
x00197407	2	Year 10	No	1.38	1.87	597	0.73	529	2.03	665				41		1.17

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00197427_Y10	1	Year 10	No	0.81	1.30	538	0.81	538						56		1.00
x00197428	1	Year 10	No	2.82	3.31	748	2.82	748						19		0.92
x00197440	1	Year 10	No	0.93	1.42	550	0.93	550						52		1.05
x00197442	2	Year 10	No	1.42	1.91	601	0.84	540	2.00	661				41		0.96
x00197474	1	Year 10	No	0.41	0.90	495	0.41	495						64		1.10
x00197477	2	Year 10	No	1.49	1.98	608	1.22	581	1.75	635				40		0.95
x00197478	1	Year 10	No	3.22	3.71	790	3.22	790						14		1.14
x00197479	2	Year 10	No	2.36	2.85	700	1.17	575	3.56	825				29		0.97
x00197488	1	Year 10	No	2.25	2.74	688	2.25	688						26		1.12
x00197527	1	Year 10	No	1.59	2.08	619	1.59	619						40		0.96
x00197530	1	Year 10	No	-0.34	0.15	417	-0.34	417						77		1.00
x00197533	2	Year 10	No	2.96	3.45	762	2.53	717	3.39	807				13		0.87
x00197534	1	Year 10	No	2.32	2.81	696	2.32	696						27		0.97
x00197535	1	Year 10	No	1.10	1.59	568	1.10	568						50		1.32
x00204927	1	Year 10	No	2.28	2.77	691	2.28	691						25		1.20
x00204984	3	Year 10	No	1.26	1.75	584	0.14	468	1.21	579	2.42	706		48		1.11
x00204985	2	Year 10	No	2.18	2.67	681	1.22	580	3.14	782				30		0.94
x00204987	1	Year 10	No	0.95	1.44	552	0.95	552						52		1.14
x00206405	1	Year 10	No	0.87	1.36	543	0.87	543						53		1.14

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00207631	1	Year 10	No	2.61	3.10	725	2.61	725						20		0.98
x00207680	1	Year 10	No	-0.56	-0.07	394	-0.56	394						79		1.01
x00120773	1	Link	Yes	-1.98	-1.49	246	-1.98	246					84	95	0.88	0.67
x00120774	1	Link	Yes	0.93	1.42	550	0.93	550					32	49	1.05	1.16
x00120787	1	Link	Yes	0.38	0.87	492	0.38	492					43	66	1.04	0.75
x00120800	1	Link	Yes	0.35	0.84	489	0.35	489					45	74	0.98	0.77
x00120810	1	Link	Yes	1.39	1.88	598	1.39	598					24	35	0.96	1.13
x00120815	1	Link	Yes	-0.14	0.35	437	-0.14	437					54	70	0.99	1.14
x00120821	1	Link	Yes	-0.36	0.13	415	-0.36	415					59	82	0.94	0.95
x00120840	1	Link	Yes	0.51	1.00	505	0.51	505					41	63	1.03	0.94
x00120845	1	Link	Yes	-0.07	0.42	445	-0.07	445					53	74	1.01	0.84
x00120846	1	Link	Yes	-1.11	-0.62	336	-1.11	336					73	81	0.97	1.35
x00120848	1	Link	Yes	-0.37	0.12	414	-0.37	414					59	82	1.05	0.94
x00120850	1	Link	Yes	1.04	1.53	561	1.04	561					31	50	1.07	1.08
x00121368	1	Link	Yes	-1.80	-1.31	264	-1.80	264					84	90	0.94	1.03
x00121380	1	Link	Yes	0.18	0.67	472	0.18	472					48	60	0.97	1.02
x00121381	1	Link	Yes	0.14	0.63	467	0.14	467					49	62	1.01	1.00
x00121383	1	Link	Yes	-1.70	-1.21	274	-1.70	274					82	91	0.90	0.65
x00121420	1	Link	Yes	-0.76	-0.27	372	-0.76	372					67	79	1.07	0.94
x00121426	1	Link	Yes	0.03	0.52	455	0.03	455					51	74	0.96	0.85

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00121530	1	Link	Yes	1.06	1.55	563	1.06	563					30	48	1.09	1.03
x00121533	1	Link	Yes	-1.78	-1.29	266	-1.78	266					82	92	0.89	0.74
x00122348	1	Link	Yes	1.81	2.30	642	1.81	642					19	26	1.07	0.87
x00130077	2	Link	Yes	1.86	2.35	647	0.42	496	3.31	799			23	40	0.92	1.07
x00159079	2	Link	No	0.12	0.61	465	-0.32	419	0.55	510			49	73	0.86	0.98
x00159080	1	Link	No	-1.06	-0.57	342	-1.05	342					72	81	0.91	0.97
x00161795	1	Link	No	0.50	0.99	505	0.50	505					41	56	0.99	0.92
x00161796	1	Link	No	-0.16	0.33	436	-0.16	436					55	71	0.97	0.91
x00194752	1	Link	No	-1.07	-0.58	340	-1.07	340					72	86	0.89	0.76
x00194754	1	Link	No	0.48	0.97	503	0.48	503					41	55	0.97	1.03
x00194764	2	Link	No	1.84	2.33	645	0.66	522	3.02	769			20	36	0.94	0.84
x00194797	1	Link	No	0.40	0.89	494	0.40	494					43	65	1.09	1.08
x00194800	2	Link	No	1.64	2.13	624	0.63	518	2.65	730			21	36	1.02	1.18
x00194813	1	Link	No	1.44	1.93	603	1.44	603					23	43	1.06	1.11
x00195400	1	Link	No	1.35	1.84	593	1.35	593					25	41	1.14	1.09
x00195447	1	Link	No	0.83	1.32	539	0.83	539					35	52	1.03	1.03
x00195453	2	Link	No	2.75	3.24	740	2.36	700	3.13	780			5	19	1.01	1.11
x00195457	1	Link	No	1.54	2.03	614	1.54	614					22	44	1.02	0.90
x00195630	1	Link	No	-0.74	-0.25	375	-0.74	375					66	83	0.77	0.65
x00195635	1	Link	No	0.49	0.98	503	0.49	503					41	66	1.22	0.99

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00195638	1	Link	No	0.69	1.18	525	0.69	525					37	57	0.99	0.91
x00196021	1	Link	No	-0.35	0.14	416	-0.35	416					58	80	1.02	0.80
x00196023	1	Link	No	-1.61	-1.12	284	-1.61	284					80	90	1.01	0.89
x00196025	2	Link	No	0.76	1.25	532	0.16	469	1.35	594			34	54	0.89	0.94
x00196026	1	Link	No	-0.88	-0.39	361	-0.88	361					68	77	1.01	1.02
x00196030	1	Link	No	0.41	0.90	495	0.41	495					42	64	0.90	0.82
x00196567	1	Link	No	0.29	0.78	483	0.29	483					45	55	1.06	1.09
x00196568	1	Link	No	0.21	0.70	475	0.21	475					47	57	1.22	1.27
x00196570	1	Link	No	2.87	3.36	753	2.87	753					8	18	1.05	1.12
x00196571	2	Link	No	2.69	3.18	734	2.08	670	3.30	798			7	19	0.94	1.15
x00196685	1	Link	No	-0.86	-0.37	363	-0.86	363					69	87	0.88	0.77
x00196686	1	Link	No	0.01	0.50	453	0.01	453					52	78	0.94	0.81
x00196688	1	Link	No	1.16	1.65	573	1.16	573					29	48	1.03	0.98
x00196694	1	Link	No	-0.62	-0.13	388	-0.62	388					64	85	0.98	0.86
x00196695	2	Link	No	-0.39	0.10	411	-1.13	334	0.35	489			60	80	0.99	0.92
x00196706	1	Link	No	1.37	1.86	596	1.37	596					25	36	1.05	1.01
x00196707	1	Link	No	1.28	1.77	586	1.28	586					27	47	1.11	1.02
x00196709	2	Link	No	1.22	1.71	580	1.13	570	1.32	591			19	59	1.12	1.02
x00196710	1	Link	No	1.09	1.58	567	1.09	567					30	54	1.00	0.97
x00196711	1	Link	No	0.26	0.75	480	0.26	480					46	75	0.94	0.88

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00196715	1	Link	No	0.79	1.28	535	0.79	535					36	53	1.03	1.01
x00196716	1	Link	No	-2.21	-1.72	221	-2.21	221					88	93	0.98	1.53
x00196836	1	Link	No	-1.73	-1.24	271	-1.73	271					82	89	0.97	1.27
x00196837	1	Link	No	-0.05	0.44	447	-0.05	447					53	79	0.89	0.77
x00196843	1	Link	No	-3.29	-2.80	108	-3.29	108					95	96	0.94	1.56
x00196850	1	Link	No	1.29	1.78	588	1.29	588					26	46	1.01	1.13
x00197138	1	Link	No	-0.71	-0.22	378	-0.71	378					65	80	1.00	0.97
x00197139	1	Link	No	0.79	1.28	536	0.79	536					34	50	0.98	1.00
x00197140	1	Link	No	-0.77	-0.28	372	-0.77	372					66	83	0.98	1.06
x00197143	1	Link	No	1.78	2.27	639	1.78	639					18	28	1.07	1.01
x00197430	1	Link	No	-0.29	0.20	422	-0.29	422					57	71	0.98	1.02
x00197431	2	Link	No	0.75	1.24	530	-0.33	418	1.82	643			36	56	1.06	1.31
x00197441	1	Link	No	-0.91	-0.42	357	-0.91	357					70	77	1.00	1.26
x00197443	1	Link	No	1.29	1.78	588	1.29	588					26	52	1.07	1.05
x00197449	2	Link	No	1.18	1.67	576	0.18	472	2.18	681			29	46	0.99	0.92
x00197453	2	Link	No	2.66	3.15	731	2.11	674	3.21	788			7	17	1.01	0.96
x00197482	1	Link	No	0.44	0.93	498	0.44	498					43	59	1.15	1.19
x00197483	1	Link	No	-0.42	0.07	408	-0.42	408					61	73	1.02	1.11
x00197496	1	Link	No	0.90	1.39	547	0.90	547					33	44	0.96	1.04
x00197511	1	Link	No	-0.93	-0.44	355	-0.93	355					69	91	0.92	0.63

				Difficulty			Threshold 1		Threshold 2		Threshold 3					
Item	Scores	Vertical link	Horizontal link	RP=0.5	RP=0.62	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	RP=0.5	SL Scale	Correct Year 6	Correct Year 10	Weighted fit (MNSQ) Year 6	Weighted fit (MNSQ) Year 10
x00197513	1	Link	No	-0.21	0.28	430	-0.21	430					55	75	1.00	0.86
x00197515	1	Link	No	1.52	2.01	612	1.52	612					22	37	1.07	0.97
x00197520	1	Link	No	1.96	2.45	658	1.96	658					16	43	0.94	1.05
x00204885	2	Link	No	-0.06	0.43	446	-0.96	352	0.84	541			53	69	1.02	0.94
x00204981	3	Link	No	0.42	0.91	496	-0.15	437	0.50	505	0.94	551	39	62	1.09	1.14
x00206122	1	Link	No	0.49	0.98	504	0.49	504					41	60	1.03	1.03
x00206315	2	Link	No	0.34	0.83	488	-0.88	361	1.55	615			46	54	1.10	1.09
x00207629	1	Link	No	-1.09	-0.60	338	-1.09	338					72	86	0.89	0.73
x00207696	1	Link	No	1.04	1.53	561	1.04	561					30	56	1.13	1.12
x00207718	1	Link	No	-0.40	0.09	410	-0.40	410					60	70	0.98	1.17

Appendix G: Variables for conditioning

Variable	Name	Values	Coding	Regressor
Adjusted school mean achievement	sch_adj_mn	Adjusted school mean	Logits	Direct
State and territory by sector	State, Sector	ACT, Government	10000000000000000000000	Direct
		ACT, Catholic	01000000000000000000000	
		ACT, Independent	00100000000000000000000	
		NSW, Government	00000000000000000000000	
		NSW, Catholic	00010000000000000000000	
		NSW, Independent	00001000000000000000000	
		NT, Government	00000100000000000000000	
		NT, Catholic	00000010000000000000000	
		NT, Independent	00000001000000000000000	
		QLD, Government	00000000100000000000000	
		QLD, Catholic	00000000010000000000000	
		QLD, Independent	00000000001000000000000	
		SA, Government	00000000000100000000000	
		SA, Catholic	00000000000010000000000	
		SA, Independent	00000000000001000000000	
		TAS, Government	00000000000000100000000	
		TAS, Catholic	00000000000000010000000	
		TAS, Independent	00000000000000001000000	
		VIC, Government	00000000000000000100000	

Variable	Name	Values	Coding	Regressor
		VIC, Catholic	00000000000000000010000	
		VIC, Independent	00000000000000000001000	
		WA, Government	00000000000000000000100	
		WA, Catholic	00000000000000000000010	
		WA, Independent	00000000000000000000001	
School geographic location	Geolocation	Major Cities of Australia	0000	Direct
		Inner Regional Australia	1000	
		Outer Regional Australia	0100	
		Remote Australia	0010	
		Very Remote Australia	0001	
Gender	Gender	Male, Other	0	Direct
		Female	1	
Indigenous status indicator	INDIG	Indigenous	10	Direct
		Non-Indigenous	00	
		Missing	01	
LOTE spoken at home	LBOTE	Yes	1	Direct
		No, Missing	0	
Parental highest occupation group	POCC	Mode of year level	00000	Direct
		Other category 1	10000	
		Other category 2	01000	
		Other category 3	00100	
		Other category 4	00010	

Variable	Name	Values	Coding	Regressor
		Not stated or unknown	00001	
Highest level of parental education	PARED	Mode of year level	00000	Direct
		Other category 1	10000	
		Other category 2	01000	
		Other category 3	00100	
		Other category 4	00010	
		Not stated or unknown	00001	
Age	AGE	Value	Copy, Mean	Direct
I would like to learn more science at school.	QN01.1_Y6	Strongly agree	Four dummies for each variable with the highest frequency category as the reference category	PCA
I want to study one or more science subjects in Years 11 and 12.	QN01.1_Y10	Agree		
I think it would be interesting to be a scientist.	QN01.2_Y6	Disagree		
I am considering a science-related career.	QN01.2_Y10	Strongly disagree		
I enjoy doing science.	QN01.3	Missing		
I enjoy learning new things in science.	QN01.4			
I learn science topics quickly.	QN01.5			
I can understand new ideas about science easily.	QN01.6			
Science is part of my everyday life.	QN01.7			
Science is important for lots of jobs.	QN01.8			
Science is important because it changes how we live.	QN01.9			
Scientific information helps people make good decisions.	QN01.10			
Science is about remembering facts.	QN02.1	Strongly agree	Four dummies for each variable with the highest	PCA
Science is about doing experiments.	QN02.2	Agree		

Variable	Name	Values	Coding	Regressor
Science is finding out about how things work.	QN02.3	Disagree	frequency category as the reference category	
Science is about solving problems.	QN02.4	Strongly disagree		
Science is about collaborating with others.	QN02.5	Missing		
Science is about making enquiries.	QN02.6			
Watch television or stream content about science-related topics	QN03.1	Frequently (more than 2 times a week)	Four dummies for each variable with the highest frequency category as the reference category	PCA
Read physical or digital books, newspapers or articles about science	QN03.2	Often (1 or 2 times a week)		
Listen to podcasts, audiobooks or radio on science-related topics	QN03.3	Sometimes (less than once a week)		
Talk about science with my friends	QN03.4	Never		
Talk about science with my family	QN03.5	Missing		
Post or share content about science-related topics on the internet or social media	QN03.6			
Contribute to existing discussions about science-related topics on the internet or social media	QN03.7			
'Like' someone else's content on science-related topics on the internet or social media	QN03.8			
Watch television or stream content about science-related topics	QN04.1	Frequently (more than 2 times a week)	Four dummies for each variable with the highest frequency category as the reference category	PCA
Read physical or digital books, newspapers or articles about science	QN04.2	Often (1 or 2 times a week)		
Talk about science with my friends	QN04.3	Sometimes (less than once a week)		
		Never		
		Missing		
Scientific information helps people make informed decisions.	QN05.1	Strongly agree	Four dummies for each variable with the highest frequency category as the reference category	PCA
Our scientific knowledge is constantly changing.	QN05.2	Agree		
Science can help us understand global issues that impact on people and the environment.	QN05.3	Disagree		

Variable	Name	Values	Coding	Regressor
I follow the advice of the scientific community when making decisions related to health crises (e.g. during the COVID-19 pandemic).	QN05.4	Strongly disagree		
Government decisions should be based on scientific evidence where available.	QN05.5	Missing		
I know where to find scientific information about local and global issues.	QN05.6			
I know how to decide whether to trust online information about a science topic.	QN05.7			
Earth sciences - for example, weather, soil, rocks, using Earth's resources	QN06.1	Yes	One dummy for each variable with the highest frequency category as the reference category	PCA
Space (astronomy) - for example, galaxies, objects in space including the planets, Sun and Moon	QN06.2	No, Missing		
Forces and motion - for example, how toys and other machines move and work	QN06.3			
Energy, forms and transfer - for example, electricity, heat, light, sound, magnets	QN06.4			
Living things - for example, how animals and plants survive in their environment, food chains and webs, ecosystems	QN06.5			
Multicellular systems - for example, the human body, cells, tissues, organs, body systems	QN06.6			
Diversity and evolution - for example, how living things change over time	QN06.7			
States of matter - for example, changes to materials (solids, liquids and gases), processes of change such as melting, evaporation	QN06.8			
Properties of matter - characteristics of materials such as density, mass, volume, melting point, hardness, elasticity	QN06.9			
How often do you have science lessons at school?	QN06.10_Y6	More than once a week	Five dummies for each variable with the highest frequency category as the reference category	PCA
		Once a week		
		Less than once a week, but more than once a month		
		Once a month or less		

Variable	Name	Values	Coding	Regressor
		Never		
		Missing		
My classroom teacher teaches science to our class.	QN07.1_Y6	Yes	One dummy for each variable with the highest frequency category as the reference category	PCA
My teacher invites visitors to school to talk about science topics.	QN07.2	No, Missing		
Our class goes on excursions related to the science topics we are learning about.	QN07.3			
My teacher can explain scientific concepts clearly.	QN07.4			
making observations about the world.	QN08.1	Strongly agree	Four dummies for each variable with the highest frequency category as the reference category	PCA
asking questions about objects and events.	QN08.2	Agree		
making predictions and testing them.	QN08.3	Disagree		
describing patterns and relationships.	QN08.4	Strongly disagree		
using evidence to develop explanations.	QN08.5	Missing		
building knowledge by trial and error.	QN08.6			
People from many different countries have made important contributions to science.	QN09.1	Strongly agree	Four dummies for each variable with the highest frequency category as the reference category	PCA
Women and men are both involved in science.	QN09.2	Agree		
People from all cultural backgrounds in Australia are involved in science.	QN09.3	Disagree		
People of all ages are involved in science.	QN09.4	Strongly disagree		
Women and men are equally skilled in science.	QN09.5	Missing		
Female scientists get as much recognition as male scientists.	QN09.6			
My teacher asks us to brainstorm ideas.	QN10.1	Never	Four dummies for each variable with the highest	PCA
My teacher helps me identify patterns between different pieces of information.	QN10.2	Sometimes		

Variable	Name	Values	Coding	Regressor
My teacher encourages me to explain the reasons why I did something.	QN10.3	Mostly	frequency category as the reference category	
My teacher encourages me to think through all the different options when making decisions.	QN10.4	Always		
I get to plan and carry out my own investigations.	QN10.5	Missing		
I use a computer or tablet for research into science-related topics.	QN10.6			
Our class has in-depth discussions about science ideas.	QN10.7			
We work in groups to carry out investigations.	QN10.8			
come up with creative solutions to solving problems.	QN11.1	Strongly agree	Four dummies for each variable with the highest frequency category as the reference category	PCA
question information I find on the internet or TV.	QN11.2	Agree		
consider situations from different perspectives.	QN11.3	Disagree		
consider the source of information.	QN11.4	Strongly disagree		
explain my reasons for doing something.	QN11.5	Missing		
look at the different parts of a problem to help me solve it.	QN11.6			
Do activities which require creative solutions	QN12.1	Never	Four dummies for each variable with the highest frequency category as the reference category	PCA
Participate in problem solving activities	QN12.2	Sometimes		
Come up with my own activities to entertain myself	QN12.3	Mostly		
Develop new ways to solve problems	QN12.4	Always		
Debate topics with my family or friends	QN12.5	Missing		
Making predictions based on prior evidence	QN13.1	Not at all confident	Four dummies for each variable with the highest frequency category as the reference category	PCA
Identifying what I don't know about a topic, so I understand what I need to learn	QN13.2	Not very confident		
Identifying patterns and making connections between different pieces of information	QN13.3	Somewhat confident		

Variable	Name	Values	Coding	Regressor
Testing different options and monitoring the outcomes	QN13.4	Very confident		
Thinking about problems from different perspectives	QN13.5	Missing		
Working on tasks that require creative thinking	QN13.6			
Questioning the accuracy of the source of information I am receiving	QN13.7			
Explaining where my ideas came from	QN13.8			

Questionnaire items are presented in this table in the order shown to students. For this reason, some items on the same scale are not presented together.
Questionnaire items have been renamed to be consistent with main study ordering. For this reason, numbers for the removed field trial items are excluded.

Appendix H: Proficiency level descriptions

Proficiency level	Description
Level 5	At Level 5, students can apply scientific principles and abstract concepts to develop and evaluate scientific explanations for complex, multi-faceted phenomena in familiar and unfamiliar contexts. Students are able to propose and justify their own scientific solutions and critique solutions made by others to address personal, community and global issues. Students can design valid scientific investigations that would systematically generate reliable data and explain the purpose of an experimental design, including how equipment allows data to be collected accurately. They can explain the value of models to investigate scientific phenomena and evaluate their advantages and limitations. Students can critically evaluate the outcomes of scientific investigations to identify limitations and sources of error, and propose alternative strategies. They can explain relationships between variables, evaluate data and information presented in a variety of formats, and justify conclusions that are consistent with evidence.
Level 4	At Level 4, students can apply scientific principles and concepts to construct and evaluate scientific explanations for complex, related phenomena in familiar contexts. Students are able to explain how scientific knowledge informs decisions and actions, and propose scientific solutions to address personal, community and global issues. Students can select equipment to collect accurate data and explain how to control variables to obtain valid outcomes. Students are able to analyse data and information resulting from investigations presented in a variety of formats. They can draw conclusions using evidence and scientific explanations and can propose strategies to improve the reliability of investigations.
Level 3	At Level 3, students can draw on scientific principles and concepts to construct and interpret scientific explanations of phenomena of increasing complexity in familiar contexts. Students can explain how scientific knowledge influences strategies proposed to solve personal and community problems. Students are able to plan straightforward investigations including identifying equipment to collect accurate data and identify and classify variables in a fair test. They can identify a source of error in an investigation and analyse data and information presented in a variety of formats. Students are able to draw conclusions consistent with evidence and support or refute predictions using evidence.
Level 2	At Level 2, students can draw on basic scientific principles and concepts to identify, explain and classify phenomena in familiar contexts. Students are able to recognise how the application of scientific knowledge can be used to develop solutions in their personal and community contexts. In the context of scientific investigations, students can identify scientific questions and predictions, and understand how variables influence outcomes. They can select appropriate equipment for a scientific investigation, perform simple calculations and label simple scientific diagrams. They can interpret data and information presented in a variety of formats and identify information that supports a conclusion from simple investigations.

Proficiency level	Description
Level 1	At Level 1, students can draw on basic knowledge and personal experience to recognise and describe aspects of phenomena using science concepts in familiar contexts. Students can identify familiar issues relating to a scientific concept that may affect their daily life. Students are able to use basic science inquiry skills to identify suitable equipment and identify risk management strategies for an investigation, take measurements and label graphics in familiar contexts. They can analyse simple representations of data and information to identify patterns and draw basic conclusions.