AUSTRALIAN CURRICULUM, ASSESSMENT AND REPORTING AUTHORITY

NAP Sample Assessment Science Literacy 2015

Technical Report

NATIONAL ASSESSMENT PROGRAM

2015 NAP-SL Project Staff

Dr Sofia Kesidou and Nick Connolly from Educational Assessment Australia (EAA) were the Project Directors of 2015 NAP-SL. Jennifer Cowing was the EAA Project Manager. The EAA test development and marking teams were led by Frans Boot. The EAA NAP-SL Helpdesk was run by Sally Matheson and Jessica Fountain.

The sampling and data analysis tasks were led by Dr Rassoul Sadeghi, Dr Andrew Stephanou and Dr Kelvin Gregory from EAA with Mark Dulhunty and Leanne Calvitto from Educational Measurement Solutions (EMS).

The Technical Report was written by Nick Connolly with contributions from Mark Dulhunty (EMS), Dr Sofia Kesidou, Dr Andrew Stephanou, Frans Boot, Melissa Lennon and text from the 2012 NAP-SL Technical Report.

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

ACARA Copyright Administration
Level 13, 280 Elizabeth Street Sydney NSW 2000
email: info@acara.edu.au

Contents

List of Tables

List of figures

CHAPTER 1 2015 NAP SAMPLE ASSESSMENTS – SCIENCE LITERACY: OVERVIEW

1.1. Introduction

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

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

NAP—SL was the first assessment program designed specifically to provide information about performance against the National Goals for Schooling in the Twenty-First Century (now the Educational Goals for Young Australians).

In July 2001, the Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA, now superseded by the Education Council) agreed to the development of assessment instruments and key performance measures for reporting on student skills, knowledge and understandings in primary science. It directed the newly established Performance Measurement and Reporting Taskforce (PMRT), a nationally representative body, to undertake the national assessment program. The PMRT commissioned the assessment in July 2001 for implementation in 2003. The Primary Science Assessment Program (PSAP) – as it was then known – tested a sample of Year 6 students in all states and territories.

Subsequently, Ministers for Education also endorsed similar sample assessment programs to be conducted for civics and citizenship (CC) and information and communication technology literacy (ICTL). Each sample assessment program is repeated every three years so that performance in these areas of study can be monitored over time. The first cycle of each program was intended to provide the baseline data against which future performance could be compared. These assessment programs are collectively known as the National Assessment Program sample assessments.

The second cycle of the science literacy assessment was conducted as the National Assessment Program – Science Literacy (NAP—SL) in October 2006. The third cycle was conducted in October 2009 and the fourth cycle was conducted in October 2012. In 2014 work began on the fifth cycle of NAP—SL scheduled for 2015. The development and implementation of this national assessment in science literacy is undertaken by the Australian Curriculum, Assessment and Reporting Authority (ACARA). The assessment is repeated with a new sample of Year 6 students every three years in order to identify trends over time.

Unlike previous cycles of NAP—SL , the 2015 cycle was conducted as a computer-based test via an online test delivery system. Details of how this was accomplished are described in this report.

1.2. Purposes of the Technical Report

This Technical Report aims to provide detailed information with regard to the conduct of the 2015 NAP—SL assessment so that valid interpretations of the 2015 results can be made, and future cycles can be implemented with appropriate linking information from past cycles. Further, a fully documented set of NAP—SL procedures can also provide information for researchers who are planning assessments of this kind.

The methodologies used in the 2015 NAP—SL assessment can inform researchers of the current developments in large-scale assessments. They can also highlight the limitations and suggest possible improvements in the future. Consequently, it is of great importance to provide technical details on all aspects of the assessment.

1.3. Organisation of the Technical Report

This report is divided into twelve chapters.

Chapter 2 is an overview of the assessment domain and an outline of the item development, test development and test design processes.

Chapter 3 describes the piloting and trialling processes and the development of the final assessment including item selection.

Chapter 4 looks at the sampling procedures used across jurisdictions, schools and students.

Chapter 5 describes how the tests were administered and marked. It also provides an explanation of how student results were reported to schools.

Chapter 6 details the processes involved in computing the sampling weights.

Chapter 7 details the processes undertaken to analyse the data obtained from the final test.

Chapter 8 provides an outline of the scaling procedures followed as part of the data analysis.

Chapter 9 explains the equating procedures which were followed so that the 2015 results could be reported against the baseline established in 2006.

Chapter 10 provides a brief overview of the cut-points at each proficiency level and information on the performance of the items on the proficiency scale.

Chapter 11 describes how the results of the survey were used to develop a psychometric scale.

Chapter 12 describes the use of a multilevel modelling process to examine relationships between the survey data and student achievement data.

Appendices 1–11 provide further elaboration and exemplification of the information in the body of the Technical Report.

CHAPTER 2 TEST DEVELOPMENT AND TEST DESIGN

2.1. Assessment domain

The NAP sample assessments – science literacy (NAP—SL) measures the science literacy of primary school students in Australian schools. NAP—SL assesses the ability to think scientifically in a world in which science and technology are increasingly shaping children's lives. Specifically, it assesses students' ability to apply broad conceptual understandings of science in order to make sense of the world; to understand natural phenomena; and to interpret media reports about scientific issues. It also includes the ability to ask investigable questions; conduct investigations; collect and interpret data; and make informed decisions.

The construct evolved from the definition of scientific literacy used by the Organisation for Economic Co-operation and Development (OECD) – Programme for International Student Assessment (PISA):

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

This definition has been adopted for the purpose of monitoring primary school science in NAP—SL (Ball et al. 2000). The science items and instruments assess outcomes that contribute to scientific literacy, including conceptual understandings, rather than focusing solely on scientific knowledge. They also assess student competence in carrying out investigations in realistic situations.

2.1.1. The historic 2012 NAP—SL Assessment Domain

A science literacy progress map (see Appendix 1) was developed during the first assessment cycle based on this construct of science literacy and on an analysis of the state and territory curriculum and assessment frameworks. The progress map describes the development of science literacy across three strands of knowledge which are inclusive of Ball et al.'s concepts and processes and the elements of the OECD–PISA definition.

In the previous four cycles of NAP—SL , three main areas of scientific literacy were assessed:

> Strand A: formulating or identifying investigable questions and hypotheses; planning investigations; and collecting evidence.
>
> Strand B: interpreting evidence and drawing conclusions from students' own or others' data; critiquing the trustworthiness of evidence and claims made by others; and communicating findings.
>
> Strand C: using science understandings for describing and explaining natural phenomena, and for interpreting reports about phenomena.

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

2.1.2. Transitioning to the Australian Curriculum: science

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

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

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

Table 2.1 Strands and sub-strands in the Australian Curriculum: science

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

Previous cycles of NAP—SL were developed in the absence of any science curriculum common across the states and territories. With the implementation of the Australian Curriculum: science in all states and territories in 2014, it is important that the NAP—SL construct was described in terms of the new Australian Curriculum: science.

Table 2.2 NAP—SL progress map strands mapped onto the strands/sub-strands of the Australian Curriculum: science

The NAP—SL progress map strands	Australian Curriculum: science strands/sub-strands
Strand A: formulating or identifying investigable questions and hypotheses; planning investigations; and collecting evidence.	Science inquiry skills – Questioning and predicting Science inquiry skills – Planning and conducting Science as a human endeavour

The NAP—SL progress map strands	Australian Curriculum: science strands/sub-strands
Strand B: interpreting evidence and drawing conclusions from students' own or others' data; critiquing the trustworthiness of evidence and claims made by others; and communicating findings.	Science inquiry skills – Processing and analysing data and information Science inquiry skills – Evaluating Science inquiry skills – Communicating Science as a human endeavour
Strand C: using science understandings for describing and explaining natural phenomena; and for interpreting reports about phenomena.	Science understanding Science as a human endeavour

There is a high degree of alignment between the NAP—SL Strand A: experimental design and data gathering and Strand B: interpreting experimental data and the science inquiry skills strand of the Australian Curriculum: science.

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

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

Table 2.3 Relation between NAP—SL major scientific concept areas and curriculum sub-strands

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

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

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

> **Literacy:** aspects of the literacy capability are found within the reading comprehension demands of both the stimuli and the items of NAP—SL .
>
> **Numeracy:** aspects of the numeracy capability are found within NAP—SL , including the reading and construction of graphs and tables, calculations and measurement, as well as some elements of spatial reasoning.
>
> **Information and communication technology (ICT):** aspects of the ICT capability will arise from online delivery.
>
> **Critical and creative thinking:** aspects of the critical and creative thinking capability arise from important cognitive skills inherent in scientific inquiry.

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

2.1.3. The 2015 NAP—SL Assessment Domain

The progress map was the key reference for test development for the 2003, 2006, 2009 and 2012 cycles of testing and was retained for the 2015 cycle to provide a direct connection with earlier assessments. Table A1.6 in Appendix 1 includes the version of the progress map that informed test development for the three most recent cycles.

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

The focus of the 2015 assessment was on concepts and skills from the Australian Curriculum: science Years 4–6 (version 7.5). However, as the Australian Curriculum represents a continuum, concepts and skills from Foundation through to Year 6 were also considered. This is consistent with the approach taken in previous NAP—SL cycles, in which the progress map articulates a progression in development of understanding and skills.

2.1.4. Framework review processes

The assessment framework underwent a series of review processes.

Following the 2012 NAP—SL assessment cycle, the progress map was extensively reviewed by subject matter experts in light of the assessment results. These revisions were brought forward into the 2015 cycle.

Subject matter experts were consulted in the early stages of framework development to identify curriculum issues.

Initial drafts of the framework were reviewed by subject matter experts.

The draft assessment framework was then reviewed by the Australian Curriculum Assessment and Reporting Authority (ACARA).

The revised version of the assessment framework was then sent to the Science Literacy Working Group (SLWG) for review by key stakeholders.

The assessment framework then underwent a series of reviews and refinement by Educational Assessment Australia (EAA) and ACARA before final agreement.

2.2. Item development process

The item development team had extensive experience in the testing of science literacy and significant experience with past NAP—SL item development processes. The team's objective during initial item development was to create a pool of items which covered the wide range of student science literacy performance of Australian Year 6 students.

The distribution of items across the assessment domain of science literacy was informed by both the new assessment framework and also the previous distribution (across strands and major scientific concept areas) used in 2012 NAP—SL .

Early discussion was held with ACARA on the range of innovation in item type and interaction that would best balance the opportunities provided by the new mode of delivery with the need for continuity in the NAP—SL assessments.

2.2.1. The item development team and recruitment of item writers

Leading members of the item development team had been involved in researching the item design implications of online delivery both via EAA's own testing and via its contracted research work. The team utilised this experience to help design test materials that used technology to improve students' experience of assessment.

The item development team included both full-time EAA staff and experienced external item writers with direct past experience with NAP—SL .

All external item writers had a solid science background, in-depth knowledge of Year 6 student understanding of science, and extensive expertise in writing science assessment items.

A total of ten item writers (including EAA staff) contributed items to ensure that no one individual had a disproportionate influence on the total pool of items developed.

2.2.2. Item writer training and initial item writing

2.2.2.1. Item writing workshop

Item development began with item writer training including a one-day item writing workshop. Item writers were briefed on the assessment framework, previous iterations of NAP—SL and priorities for the 2015 NAP—SL cycle (including online delivery).

The focus of the workshop included:

- writers' understanding of the test specifications
- valid interpretations of the assessment framework/progress map and key concept areas
- opportunities and constraints of online delivery on item writing.

2.2.2.2. *Initial item writing*

Item writers first identified and developed a range of stimulus topics using the major scientific concept areas and the Australian Curriculum: science as a guide for suitable topics. Then item writers developed associated items in static form in Word documents. These items and their associated metadata were reviewed.

In the first instance, an item set submitted by an item writer was reviewed by an experienced member of the item development team who would discuss any necessary changes with the item writer. The item writer would then make changes (if required) and re-submit the item set.

A pool of item sets in the form of Word documents was collated with initial metadata, including curriculum reference and descriptors. This pool of items then underwent further internal review.

This process was conducted in batches of development, to allow item sets to be reviewed and the item development process to be monitored. This helped ensure a balance of items was developed that would meet the needs of the assessment.

2.2.3. Internal reviews

Once a batch of item sets had been developed, they underwent a review for alignment to the specifications and assessment framework, content accuracy, context, literacy demand, technical qualities and fairness.

During internal review items were judged against a range of criteria that were also used to review items at later stages of item development. Criteria included:

- Alignment with assessment framework: The purpose was to ensure that items fit within the assessment framework and matched the specified strands, levels and concept areas.
- Suitability for online delivery: The purpose was to ensure that stimulus and items were suitable in terms of size and layout for online delivery and that opportunities to enhance the effectiveness of the item through online delivery had been thoroughly explored.
- Technical correctness of scientific content: The purpose was to ensure that stimulus content was factual, correct and that relevant data was derived from reliable and authoritative sources.
- Context: The purpose was to ensure that the context/stimulus of the items were likely to be accessible and of interest to Year 6 students.
- Clarity: The purpose was to ensure that the question or task was clearly stated, the graphics were clear and the wording in the stem and options was clear and concise.

- Literacy demand: The purpose was to ensure that the language used in the items was accessible to all students and that the use of unfamiliar and difficult vocabulary was avoided, except where such use was for subject-specific outcomes.
- Correctness of spelling and grammar.
- Fairness: The purpose was to ensure that items were free of cultural stereotyping, sensitive topics or offensive language and that items did not include irrelevant characteristics that might give a student group a particular advantage.
- Key check (for automated marking): The purpose was to ensure that the item has one, and only one, correct response or that the full range of correct response forms could be captured by an automated system.
- Suitability of marking guides (for expert marking): The purpose was to ensure that the criteria for awarding score points accurately reflected the intent of the question and discriminated sufficiently, that the marking guide appropriately rewarded a range of student responses, and that the description of the criteria for awarding score points was clear and sufficient to ensure consistency of marking.
- Standard item writing criteria: The purpose was to ensure all items follow best practice in item development including independent options for selected response items that reflect common errors.

In addition, item metadata was reviewed and amended during each review stage. Metadata was contained in EAA's item tracking system.

2.2.4. Transition to online delivery

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

Once authored into the online system, the new version of the items underwent an additional round of internal review.

2.2.5. New item types

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

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

These item types were still present in 2015 NAP—SL but online delivery permitted a greater range of item types to be used. Each item type is described in terms of an 'interaction' to describe the mechanism used by the test delivery system to capture a student's response. Item types included:

- Extended text: a text box is presented in which students can type text that is then typically marked by expert markers. This was used for items equivalent to the long constructed response and some items equivalent to short constructed response.
- Text entry: a text box is presented in which students can type text. The length of the answer is typically shorter than for extended text, and this item type is often designed for machine marking. This was used for items where students needed to give a numerical response.
- Multiple choice: a set of options is presented, preceded by a ‘radio button’ that is, a small circle which students can click to select their response. Only one response can be selected.
- Multiple choices: similar in layout to multiple choice but with a square box in front of each response. Clicking on the box displays a ‘tick’ to show that a response has been selected. Students can select multiple responses. Clicking on a response a second time de-selects it.
- Hotspot: a graphic divided into regions is displayed. Students choose their response by selecting a region. The interaction can be programmed to accept either a single correct answer (equivalent to multiple choice) or multiple answers (equivalent to multiple choices). This was used to present items equivalent to multiple choice or multiple choices with more complex graphical options and layouts.
- Interactive gap match: presents a set of words that can be dragged into marked gaps, such as a space in a sentence, a table or a diagram. In 2015 NAP—SL this type was used for completing tables and diagrams.
- Interactive graphic gap match: presents a set of pictures that can be dragged onto a larger graphic which has pre-defined regions or ‘gaps’. This was used for completing diagrams and graphs.
- Select point: students can click on any point on a graphic. An invisible zone defines a correct response. This type was used for plotting points on graphs or diagrams.
- Position object: similar to the interactive graphic gap match but allowed students to drag a graphic anywhere on a larger graphic. Pre-defined ‘gaps’ were not displayed. Only a small number of items were developed and none proceeded past the trial stage.
- Composite: combines more than one interaction. For example, an item requiring two responses in two separate text boxes. In particular the composite type was used to help present some historical items in a way that closely resembled their original paper layout.
- Match: a special item type used for the survey questions which allows several multiple choice items with identical choices to be displayed in a grid format.

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

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

2.2.6. Use of multimedia

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

In developing multimedia stimulus, the following factors were considered:

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

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

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

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

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

2.2.7. Adapting historical paper-based items to online delivery

In each cycle of NAP—SL after 2003, the final test forms have included historical link items. These are test items taken from previous test cycles that are used to help calibrate the difficulty of the test and to provide a way to measure changes of achievement over time.

The best practice for historical link items is to ensure that the items chosen have strong psychometric properties, that they include a range of difficulties and that they are presented in exactly the same way as they were originally. Test developers avoid making even small changes in historical link items because of the potential of introducing changes that affect the difficulty of the item.

NAP—SL 2015 also needed to include historical link items. However, because of the shift from a paper-based test to a computer-based test, each of the historical items had to be adapted for display online. Consequently, it was not possible to ensure that there were no changes to the historical link items. All of the chosen items had been designed to be displayed in paper booklets on A4 pages arranged in a portrait format using conventional paper-based item formats. In particular, past NAP—SL stimulus could be up to a page and a half in length when diagrams were included. Simply repeating the existing layout in the online environment would mean students would need to scroll through stimulus to find relevant information.

It was also felt that the historical link items should not look too different from the new items. As the historical link items had black and white or grayscale graphics and the new items had full colour graphics, it was decided to include some use of colour in the historical link item graphics.

Some minor adjustments were also needed to the way students responded to items. Cases where students had to circle a word in the paper-based version were changed into online multiple choice items. Cases where the paper-based version had a multiple-choice item with graphical options in an unusual layout were changed to graphical hotspot items with the same layout.

All changes were reviewed by subject specialists and EAA's Senior Psychometrician.

2.3. Assessment construction and delivery

2.3.1. Test specifications

Item and test development was based on the following specifications:

- develop/select approximately 110 items in total for the final test forms (including link items from previous cycles)
- provide sufficient assessment items for up to two hours of testing for each student in the national sample
- provide sufficient assessment items to form School Release Materials for subsequent teacher use and items to be held secure for 2018
- develop three core types of items
    - multiple choice items and other selected response formats
    - short constructed response items (requiring one or two word responses from students)
    - long constructed response items requiring students to provide an extended response of the order of one or two sentences – up to a short paragraph

- enhance and adapt the three core item types to online use by use of QTI standard items provided by the online delivery system
- balance the core item types within the trial item pool to be approximately
    - 50 per cent multiple-choice/selected response
    - 10 per cent short constructed response
    - 40 per cent long constructed response
- the balance between process items (Strands A and B) and conceptual items (Strand C) would be approximately in the proportion half process and half conceptual items.

The assessment itself would be split into two parts:

- an objective test consisting of a mix of items gathered into thematically related item sets
- a set of inquiry tasks consisting of sets of items organised into a sequence that mimic the stages of a science investigation.

2.3.2. Objective test design

In order to cover a wide range of content areas in science, but at the same time not to place too much burden on each student, the Balanced Incomplete Block rotation design was implemented. A rotation design allows a greater number of items to be assessed by using several forms with different items rotated across them. It minimises the effect of biased item parameters caused by varying item positions arising from the placement of an item in a test form. Items were placed in 'clusters' and the clusters were rotated through the test forms, each appearing three times, each time in a different location ('block') in the test form. Seven test forms were developed for the final assessment. Students were required to sit only one of the test forms.

Table 2.4 Rotation design used in the 2015 NAP—SL final assessment

Form	Block 1	Block 2	Block 3
1	Cluster 1	Cluster 2	Cluster 4
2	Cluster 2	Cluster 3	Cluster 5
3	Cluster 3	Cluster 4	Cluster 6
4	Cluster 4	Cluster 5	Cluster 7
5	Cluster 5	Cluster 6	Cluster 1
6	Cluster 6	Cluster 7	Cluster 2
7	Cluster 7	Cluster 1	Cluster 3

2.3.3. Inquiry task design

Students sat one of two inquiry tasks. Tasks were allocated so that half of the students who sat any given objective test form would sit inquiry task 1 and half would sit inquiry task 2.

In total, there were 14 pairings of objective test forms and inquiry tasks.

2.3.4. Inquiry task development

In addition to the transition to online delivery, 2015 NAP—SL also marked a major change in the way science inquiry skills were assessed. Previous NAP—SL cycles included a practical component in which students completed a practical task in groups of three and then answered items individually.

While this approach had been relatively successful in the past it was not one that could be replicated using the test delivery system that was available. Numerous options were considered but discounted either because the technical demands would be too great for schools or because of issues of assessment validity.

After discussion with ACARA it was agreed to develop a series of inquiry tasks. Each task used the model derived from the science inquiry skills strand of the Australian Curriculum: science in a simulated science investigation:

1. Questioning and predicting.
2. Planning.
3. Conducting.
4. Processing and analysing.
5. Evaluating.

With the over-arching skill of ‘Communicating’ running through the whole task, these steps were formulated into a proposed structure for each inquiry task.

Table 2.5 Proposed structure for the inquiry task

Task stage		Science inquiry sub-strands	Description (Australian Curriculum elaborations are included as examples)
Intro		N/A	Written and video stimulus that sets up the topic that will be investigated. The video will present a question or a phenomenon to be investigated further. There may be a basic comprehension question at this stage to encourage engagement with the initial stimulus and to provide an easy first question.
1		Questioning & predicting	Based on their understanding of the stimulus presented, students will be asked to make predictions about what will occur when some variable is changed. Students may also be asked to suggest questions that could be posed or aspects that could be investigated.
2	2.1	Planning	Students may be presented with the choices made at Stage 1 by a fictional character (e.g. “Amy wanted to investigate…”, “Caleb predicted that…”). They will then be asked to plan various aspects of the investigation. This may include open-ended items and selected-response items (e.g. selecting equipment).

Task stage		Science inquiry sub-strands	Description (Australian Curriculum elaborations are included as examples)
	2.2	Conducting	Students may be presented with some of the choices made by the fictional character as they have planned their investigation. Students will then be asked questions related to conducting the experiment. These may include safety issues, use of measurement equipment and use of appropriate units. Students may be presented with a video or a graphical stimulus of the experiment being conducted from which they may need to collect relevant data.
3		Processing & analysing	Students may be presented with data collected from the experiment by the fictional character (which could be the same as the data the students collected). They will be asked to process these data in various ways. This may include completing tables, producing graphs and organising data. Students will be asked to consider predictions made about the experiment.
4		Evaluating	Students may be given a summary of the experiment and the choices that were made by the fictional characters. They will be asked questions related to improvements that might be made to the experiment or problems with the way it was conducted. [This stage may be the final stage in some tasks.]
5 (optional)		Communicating	Students may be asked to describe or illustrate a key idea or aspect of the task. This stage may occur earlier or aspects of communicating may be assessed at other stages and this stage omitted altogether.

This structure was given to experienced science educators as the basis for developing a set of tasks. Outlines of these tasks were passed onto ACARA and key stakeholders for review before full item development work began.

After feedback and approval of the outlines, the inquiry tasks were authored into the online item authoring and review system. Because of the nature of the inquiry task it was felt that a greater emphasis on interactive items and multimedia stimulus should be placed on the inquiry task.

Initial video materials were developed that involved school aged students (one boy and one girl) conducting a science investigation. A separate audio track was recorded that narrated the investigations. These materials were developed to be of sufficient quality to do an initial pilot of the inquiry task items.

Each of the inquiry task items was then piloted online in a number of schools. Each pilot was observed by test development staff actively involved in the development of the tasks. After each task was completed observers discussed the experience with the students involved (see section *3.1.1 Inquiry task pilot*).

All student feedback was incorporated into further revisions of the inquiry task materials. The major changes in presentation were:

- the voices of the characters and the overall narration were replaced
- two professional voice actors recorded new scripts from the point of view of the characters conducting the investigation
- the footage of actual children was replaced with animated characters in all but one task.

These changes allowed for greater flexibility in editing the short stimulus videos included in the inquiry task.

Four inquiry tasks were included in the trial of all the items. The two tasks with the best overall performance psychometrically and best balance of skills were then selected for the main study.

2.3.5. Assessment review processes

All items and stimulus materials underwent several review steps:

1. An item writer wrote an initial version of an item set in Microsoft Word.
2. This initial version was reviewed by an experienced member of EAA staff who would discuss any necessary changes with the item writer.
3. A pool of item sets was collated with initial metadata, including curriculum reference and descriptors.
4. This pool of items then underwent further internal review by panels of EAA staff.
5. Selected items from this pool were authored into the online authoring and review system along with graphics.
6. Online versions of these items then underwent further internal review and edits.
7. Edited versions of the items were then released to ACARA for review by key stakeholders.
8. Items were then edited in response to feedback from ACARA and key stakeholders.
9. Items were further edited using feedback from the pilots.
10. Items were collated into trial test forms.
11. Trial test forms were reviewed internally and by ACARA.
12. Post-trial items were reviewed by EAA against their psychometric performance and recommendations were made which were passed onto ACARA and key stakeholders.
13. ACARA, EAA and key stakeholders met to review the post-trial recommendations and to propose additional changes.
14. Final forms were constructed which underwent further review by EAA and ACARA.

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

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

At major review stages, all comments on items were collated and tracked against subsequent edits. Records of the items as they appeared at each review stage were kept to allow comparisons to be made of how the items changed during the review and editing processes.

2.3.6. Items delivered

A total of 260 items were released for review prior to the trial, including 43 historical link items used in the 2012 NAP—SL assessment. The new items were reviewed in two batches by the Science Literacy Working Group (SLWG).

Table 2.6 Composition of the pre-trial item pool (new items only)

Australian Curriculum: science strand	Objective test	Inquiry task	Total
Science as a human endeavour	19	1	20
Science inquiry skills	66	41	107
Science understanding	86	4	90
Concept area			
Earth and space	37		37
Energy and force	51	25	76
Living things	48	21	69
Matter	35		35
2012 NAP—SL strand			
A	21	16	37
B	40	21	61
C	110	9	119
Item type			
Long constructed response	49	20	69
Short constructed response	24	6	30
Multiple choice	73	14	87
Other online types	25	6	31

Type			
Cloze	14	4	18
Extended text	45	17	62
Hotspot	8		8
Interactive gap match	13	1	14
Interactive graphic gap match	9	3	12
Multiple choice	67	16	83
Multiple choices	9	3	12
Position object	2		2
Select point	2	2	4
Text entry	2		2
Total	**171**	**46**	**217**

The 'cloze' item type was a non-standard item type that was superseded in the item authoring and review system by the composite item type. All the cloze type items were re-authored as composite types prior to the trial.

2.4. Trial form construction

From the approved pool of items trial test forms were developed.

Selected items for the objective test forms were grouped in eight clusters, C1–C8. Each objective test form consisted of two clusters. This allowed eight test forms to be constructed. The inquiry task items made up four inquiry task units, each of which was contained in an independent form.

Table 2.7 Structure of trial test forms by cluster

Trial Forms	Cluster 1	Cluster 2	Source of items
Objective form 1	C1	C2	2006, 2009 & 2012
Objective form 2	C2	C3	2012 & 2015
Objective form 3	C3	C4	2015
Objective form 4	C4	C5	2015
Objective form 5	C5	C6	2015
Objective form 6	C6	C7	2015
Objective form 7	C7	C8	2015
Objective form 8	C8	C1	2006, 2009 & 2015
Inquiry task 1			2015
Inquiry task 2			2015
Inquiry task 3			2015
Inquiry task 4			2015

Each cluster contained approximately 20 items. Based on the assumption that each item takes 1.5 minutes, each cluster contained approximately 30 minutes of material. The trial tests contained 204 different items, including inquiry tasks and associated items. In each class, the eight objective forms were randomly assigned.

2.4.1. Mode-effect forms

Objective form 1 of the trial contained only historical link items. This was done so that this form could be used to study the impact of the adaptation of these items to the online system.

A paper-based version of this form was also developed that contained the historical link items in their original format and layout. The item order for both the computer-based and paper-based versions was the same.

2.4.2. Items selected for trial

The composition of items selected for trial, including the historical items, is presented here.

Table 2.8 Characteristics of items selected for trial

Australian Curriculum: science strand	Objective items	Inquiry task items	Total
Science as a human endeavour	14	1	15
Science inquiry skills	65	40	105
Science understanding	79	5	84
2012 NAP—SL concept area			
Earth and space	37		37
Energy and force	53	25	78
Living things	36	21	57
Matter	32		32
2012 NAP—SL strand			
A	20	17	37
B	41	21	62
C	97	8	105
Type			
Composite	20	10	30
Extended text	46	17	63
Hotspot	6		6
Interactive gap match	8	1	9
Interactive graphic gap match	7	3	10
Multiple choice	60	11	71
Multiple choices	7	2	9
Position object	2		2
Select point	1	2	3
Text entry	1		1
Total	**158**	**46**	**204**

2.5. Student survey development

In 2009, a student survey about students' attitudes to and interests in science and science experiences in school was introduced into NAP—SL . This addition to the testing program was continued in 2012. The survey was conducted following completion of the practical task. It was decided that the survey would also be included in the 2015 cycle and would be conducted online after the inquiry task.

2.5.1. Historical survey questions

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

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

2.5.2. The role of the science as a human endeavour strand

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

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

- Science involves testing predictions by gathering data and using evidence to develop explanations of events and phenomena (ACSHE081/98).
- Science involves making predictions and describing patterns and relationships (ACSHE061).
- Important contributions to the advancement of science have been made by people from a range of cultures (ACSHE082/99 v7.5).

2.5.3. Survey review processes

All survey items underwent internal review by EAA staff. The original survey items were circulated to key stakeholders for discussion and the two new clusters were then reviewed by ACARA and key stakeholders.

All survey items were trialled online and the results compared with past performance. The survey items were then reviewed again by ACARA and key stakeholders and as a consequence some minor changes were made to the two new clusters of items.

The final survey can be seen in Appendix 9.

CHAPTER 3 PILOTS, TRIALS AND FINAL ASSESSMENT DEVELOPMENT

3.1. Trial of test items and student survey

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

- a pilot of the new inquiry task section of the assessment, designed to evaluate the effectiveness of the items and to gain early insights into running the assessment online
- a field trial comprising historical link items to ensure that these items remained effective test items after modification
- a large-scale trial designed to evaluate the psychometric properties of the items.

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

3.1.1. Inquiry task pilot

Educational Assessment Australia (EAA) conducted pilots of the 2015 NAP sample assessments – science literacy (NAP—SL) inquiry tasks in November 2014. Four tasks were piloted: Bouncing Balls, Mustard Seeds, Pendulums and Sunscreen.

Eight sessions were conducted with each task piloted in two sessions. These sessions were conducted in five schools:

- an independent Baptist school
- two Catholic Education Office (CEO) schools
- two NSW Department of Education (DET) public schools.

The pilots had multiple purposes:

- to study student engagement with the tasks
- to study the practicalities of running the tasks online
- to study student interaction with the test platform.

The observations from the pilot were intended to inform the following:

- modifications and edits to items and stimulus in the inquiry tasks
- advice to schools involved in the NAP—SL trials and main study on how to conduct the tests
- advice to the Australian Curriculum, Assessment and Reporting Authority (ACARA) and Education Services Australia (ESA) on issues relating to the test delivery platform.

3.1.2. Mode-effect study

One of the main objectives in 2015 for NAP–SL was to develop an equating design to place the 2015 results on the science literacy scale established after the 2006 scale. Since this is the fifth cycle of NAP–SL, there was an opportunity to draw link items from previous test cycles using the common-item method with multiple-linking linkage.

The items drawn from previous test cycles had to be adapted for online delivery. This process included modified layouts for stimulus and the use of colour graphics.

Given these changes, it was important to investigate the impact of the test delivery mode on the performance of the historically linked items to ensure that these items functioned in a similar fashion in an online environment.

A mixed-mode science literacy assessment (paper-based versus computer-based) was administered to a sample of students during the 2015 NAP–SL trial to directly determine the effect of mode of delivery in the NAP–SL context. The comparison of item locations of historically linked items obtained from paper-based and computer-based versions was intended to inform:

- the selection of historical links for the final assessment; and
- the work needed to place 2015 NAP–SL onto the historical scale.

Methodology

3.1.2.1. Sample

The mode-effect sample was selected from the trial sample. The trial sample included 50 schools across New South Wales, Queensland and Victoria. Ten trial schools were selected randomly to participate in the mode-effect study. In each selected school, approximately 50 students were selected randomly to participate. Students were then randomly assigned to either the paper-based test (PBT) or the identical computer-based test (CBT). The number of participants for CBT and PBT were 245 and 254, respectively. Consequently, the students in the two test conditions could be considered to be equivalent samples drawn from the same cohort of students.

3.1.2.2. Instruments

The historically linked items from 2006 to 2012 NAP—SL (a total of 43 items) were compiled into one test form with two clusters. This single test form was delivered online to the CBT sample while the PBT sample was given a physical test booklet. The final test consisted of 23 multiple-choice items and 20 constructed-response items. Students were given one mark for the correct answer, except for item 36 where the possible marks were 0, 1 or 2 (that is, the total possible mark for the assessment was 44).

3.1.2.3. Procedure

Delivering a test paper which includes most past NAP—SL link items required a highly secure test administration process. For the PBT, a time-limited study in the selected schools was conducted. Trained observers were employed to deliver the hardcopy materials, monitor the invigilation, and collect and return the materials securely to EAA.

The mode effect study was conducted at the same time as the NAP—SL trial.

3.1.2.4. Marking and data processing

Computer-based test: A team of five experienced markers was engaged for a 7-day period following the trial. The marking centre was located on site at EAA. Markers were briefed on the process, with a focus on the task of using the criteria provided in the marking guides to mark items with extended text responses.

All items were marked online using a marking system designed to work with the test delivery system. EAA staff, in collaboration with the technology providers, set up the marking system prior to the marking centre. Markers were then trained in the use of the marking system. To maintain quality control, double marking was used initially to ensure consistency of marking. Any inconsistency between marks was escalated in the system to senior staff. In addition to this, marks were sampled by senior staff to ensure that marking was correct and consistent throughout. Data were exported from the test-delivery/marking system and then processed by EAA.

Paper-based test: Constructed response items were marked at EAA. Special consideration was given to ensure the consistency of markers. Data obtained from the PBT was entered manually and checked using the standard parallel processing procedure.

3.1.2.5. Key findings - results and discussion

To investigate differential item function (DIF) of items across two modes the Rasch measurement model was applied to calibrate items simultaneously. At item level, two out of 43 items displayed significant uniform DIF: item 1 and item 26 favoured PBT and CBT respectively. Further examination of item 1 revealed that the item graphics did not appear on the screen in the same way as they appeared on paper due to a different resolution in the online environment. This might explain why the PBT group had a higher chance of responding to this item correctly. Closer inspection of item 26 showed that answering correctly was easier on CBT as all the required information was presented on one screen and students did not need to turn the page to obtain extra information.

For the purposes of estimating the impact of the change in difficulty of the historical link items, the mode-effect study data was re-analysed. Two scales were constructed with the data collected with the PBT and CBT tests using Quest. The case reliability of the two tests was 0.87 and 0.85 respectively. Blank responses were treated as missing. The 43 items in each of the two tests showed satisfactory fit to the Rasch model.

The facility of most items in each of the two tests was lower for the CBT version, suggesting that the CBT test was more difficult than the PBT test for the same ability samples. Equating of the CBT scale on the PBT scale allowed all items to be located on the same scale and thus the two tests to be compared in difficulty. The table below shows the results of the equating.

The estimation of abilities was performed by treating all blank responses as incorrect.

Table 3.1

Comparison of difficulty of CBT and PBT after equating

Scale	Mean item difficulty (logit)	Mean ability (logit)	Standard deviation ability (logit)
PBT	0.000	–0.1740	1.08
CBT	0.000	–0.5076	1.00
Difference CBT-PBT		0.334	

The ratio of the standard deviation of abilities is 1.08 supporting the equating of the two scales. As it can be seen the difference in mean student achievement is 0.334 logits. Given that two groups of students are equivalent in term of their ability, the difference in mean indicates that interaction of students with the online version of historical link items was different to student interactions with the same items in the paper format and therefore the difference needs to be taken into account when finalising the longitudinal equating of 2015 NAP-SL outcomes.

It is important to note that these differences are only applicable to the 2015 NAP–SL tests cycle. In addition, and equally important, the observed differences in mode of testing are not expected to occur in other NAP assessments including NAPLAN tests as they have different items and test design compared with NAP–SL tests.

3.1.3. Large scale trial

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

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

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

Each student completed one of the eight trial objective test forms and one of the four trial inquiry tasks. Test forms and tasks were allocated randomly before the trial period so that all test forms and all inquiry tasks were undertaken at each participating school.

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

3.1.4. Data analysis

Data were exported from the test-delivery/marking system by ESA and then processed by EAA. The trial scores were analysed by EAA using Conquest and then RUMM software. The data was also sent to an external contractor, Educational Measurement Solutions (EMS), for parallel processing. The results of the parallel analyses were consistent. The results of these analyses were compiled onto a psychometric input data sheet supplied by ACARA.

Key criteria for judging the performance of items were measures of item fit statistics (weighted MNSQ) and performance illustrated by Item Characteristic Curves (ICC). Percentage correct and point-biserial correlation were noted, but only informed a decision to eliminate an item if other indices were poor. Based on the initial analysis (first run), four items were eliminated because of poor indices and flagged as 'Reject'. An extra item was flagged as 'Reject' due to technical issues with how the student response data were captured by the system. Table 3.2 below shows the five rejected items.

Table 3.2 Rejected items

Item	Item type	Unit	Objective/ Inquiry	Reason for rejecting
NSL15E_V004.2	Interactive gap match	Dissolving	Objective	Poor fit
NSL15E_V001.5	Position object	Light bulbs	Objective	Technical problems
NSL15E_C008.1	Multiple choice	The moon	Objective	Poor fit
NSL15E_Z004.4	Extended text	Sinking marble	Objective	Poor fit
NSL15Esi1-5.1	Multiple choice	Bouncing balls	Inquiry	Poor fit

The following diagram illustrates the distribution of all trialled items (indicated by item identifiers used during analysis). This diagram provides 'at a glance' the range of difficulty of the items, and how they align with the ability of students in the trial pool (each 'x' represents five students). The range of item difficulty was nearly 8 logits; the easiest item was I0113

(NSL15E_F001.1 Recycling) at –4.06 logits. The most difficult item was I0123 (NSL15E_C003.1 Plastic bottle) at 3.83 logits.

Figure 3.1 Item person map of 2015 items

```
--------------------------------------------------------------------------------------------------------
LOCATION              PERSONS   ITEMS [locations]
--------------------------------------------------------------------------------------------------------
  4.0                         |
                              | I0123
                              |
                              |
                              | I0194
  3.0                         |
                              | I0193 I0114 I0128 I0111
                              | I0133 I0035
                              | I0086 I0089 I0177 I0174 I0154 I0181
                              | I0192 I0083
  2.0                         | I0121 I0139
                            x | I0085 I0196 I0203 I0132 I0078 I0096 I0057
                            x | I0171 I0101 I0110
                            x | I0202 I0153 I0145 I0080 I0072 I0137 I0144 I0107
                           xx | I0106 I0074 I0156 I0094 I0098 I0122
  1.0                      xx | I0138 I0041 I0119 I0158 I0131 I0016 I0180 I0204 I0136  I0095
                      xxxxxxx | I0004 I0043 I0082 I0006 I0103 I0151 I0188
                     xxxxxxxx | I0161 I0147 I0109 I0176 I0069 I0125 I0190
                     xxxxxxxx | I0052 I0034 I0152 I0172 I0051 I0020 I0200 I0100 I0076  I0036 I0159 I0201
                  xxxxxxxxxxx | I0173 I0178 I0197 I0141 I0187 I0102 I0150 I0157 I0097  I0048 I0127 I0011
  0.0          xxxxxxxxxxxxxx | I0081 I0185 I0033 I0146 I0075
              xxxxxxxxxxxxxxx | I0084 I0149 I0093 I0092 I0019 I0073 I0116 I0120 I0066  I0195 I0191 I0170 I0155 I0104 I0091
             xxxxxxxxxxxxxxxx | I0198 I0015 I0169 I0143 I0162 I0077 I0012 I0058 I0087  I0070 I0126 I0056 I0118 I0088
               xxxxxxxxxxxxxx | I0071 I0032 I0026 I0049 I0160 I0046 I0014 I0067 I0002  I0063 I0060
           xxxxxxxxxxxxxxxxxx | I0028 I0186 I0029 I0105 I0068 I0148 I0134
 -1.0           xxxxxxxxxxxxx | I0042 I0013 I0001 I0009 I0040 I0023 I0047 I0065 I0030  I0005 I0175 I0062
                 xxxxxxxxxxxx | I0021 I0117 I0055 I0112 I0184 I0017 I0165 I0022 I0025  I0064 I0007 I0003
                 xxxxxxxxxxxx | I0124 I0168 I0140 I0179
                   xxxxxxxxxx | I0199 I0050 I0045 I0090 I0024 I0130 I0135
                     xxxxxxxx | I0039 I0164 I0183 I0059
 -2.0                 xxxxxxx | I0008 I0099 I0129 I0166 I0010 I0031
                        xxxxx | I0018 I0182 I0061 I0079 I0108 I0038
                           xx | I0115 I0054
                           xx | I0027 I0053 I0044 I0142
                            x | I0037
 -3.0                       x | I0167
                              |
                              |
                              | I0189 I0163
                              |
 -4.0                         | I0113
--------------------------------------------------------------------------------------------------------
                 x = 5 Persons
--------------------------------------------------------------------------------------------------------
```

As can be seen from the diagram, the 2015 trial assessment achieved an excellent spread of item difficulties but did contain many difficult items for the Year 6 cohort. There were a number of items that all students found to be very easy, a number of items that were challenging (even for the most able students), and many items in the middle range.

Differential Item Functioning (DIF) analyses for gender were carried out for all remaining items. However, DIF analysis for language (LBOTE) could not be considered due to small sample size and the lack of information about specific language background provided by students who participated in the trial.

The DIF analyses were carried out using Conquest by fitting a facets model, where the interaction between an item and the gender group is estimated. In cases where items exhibited large DIF, content experts inspected the reasons for the observed bias. The items were flagged but not automatically removed simply based on statistical evidence of bias.

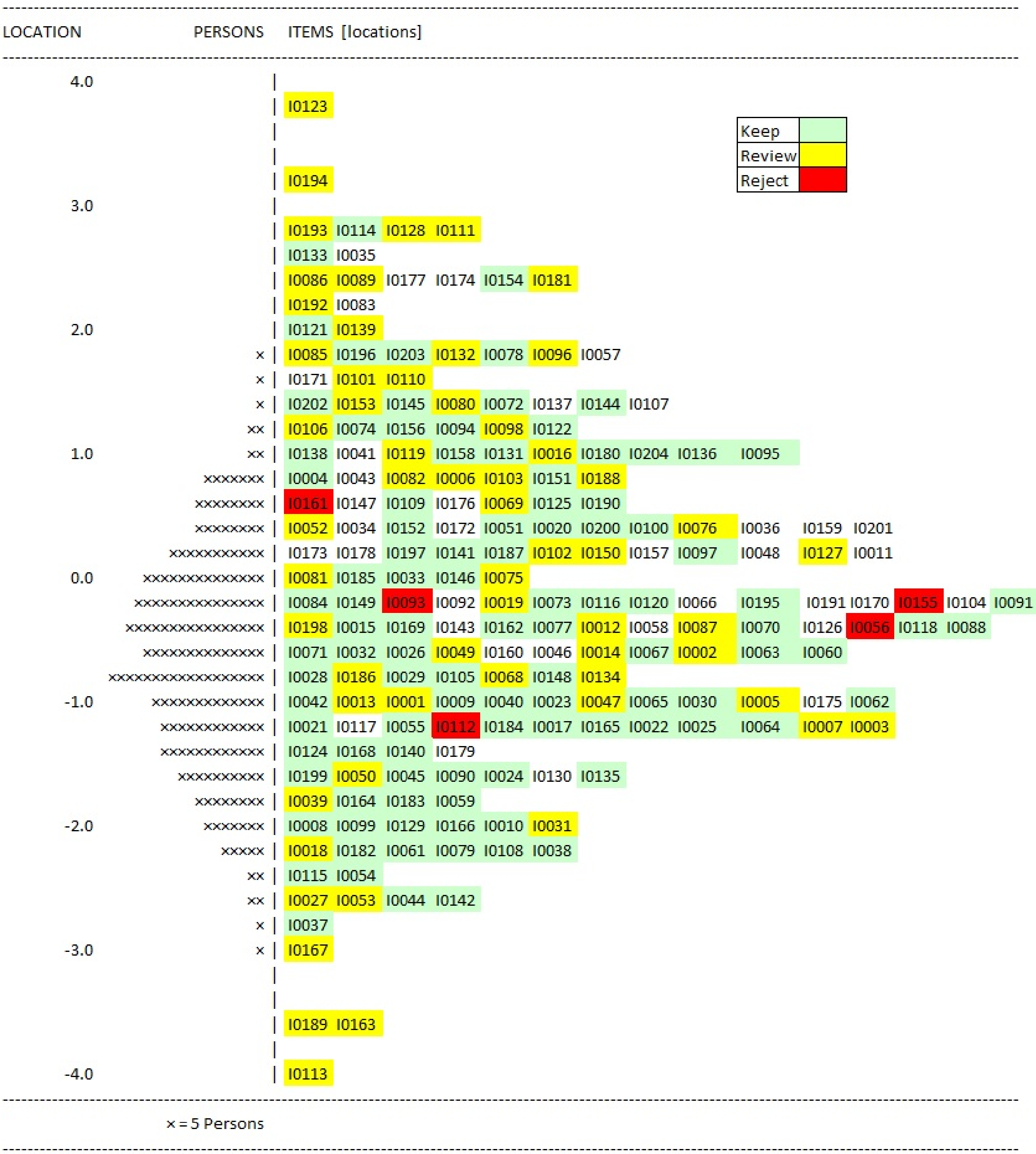
Based on this psychometric analysis as well as more general feedback from the trial, items were classified into four categories based on their overall quality. These categories were used to help inform the Science Literacy Working Group (SLWG) post-trial review.

Table 3.3 Categories used to classify trial items by quality and the number of items in each category

Category	Explanation	Number of items			
		Historical objective items	New objective items	Inquiry task items	Total
Keep	No obvious issues with the item	21	58	22	**101**
Review	Some issues that should be considered when including the item in a test	21	25	11	**57**
Low priority	The item could be viable but is not ideal for use	1	8	12	**21**
Reject	Not a suitable item for inclusion in a test	0	2	1	**3**
Total		**43**	**93**	**46**	**182**

The following item-person map for 2015 trial items shows the items as they were classified for the SLWG post-trial review.

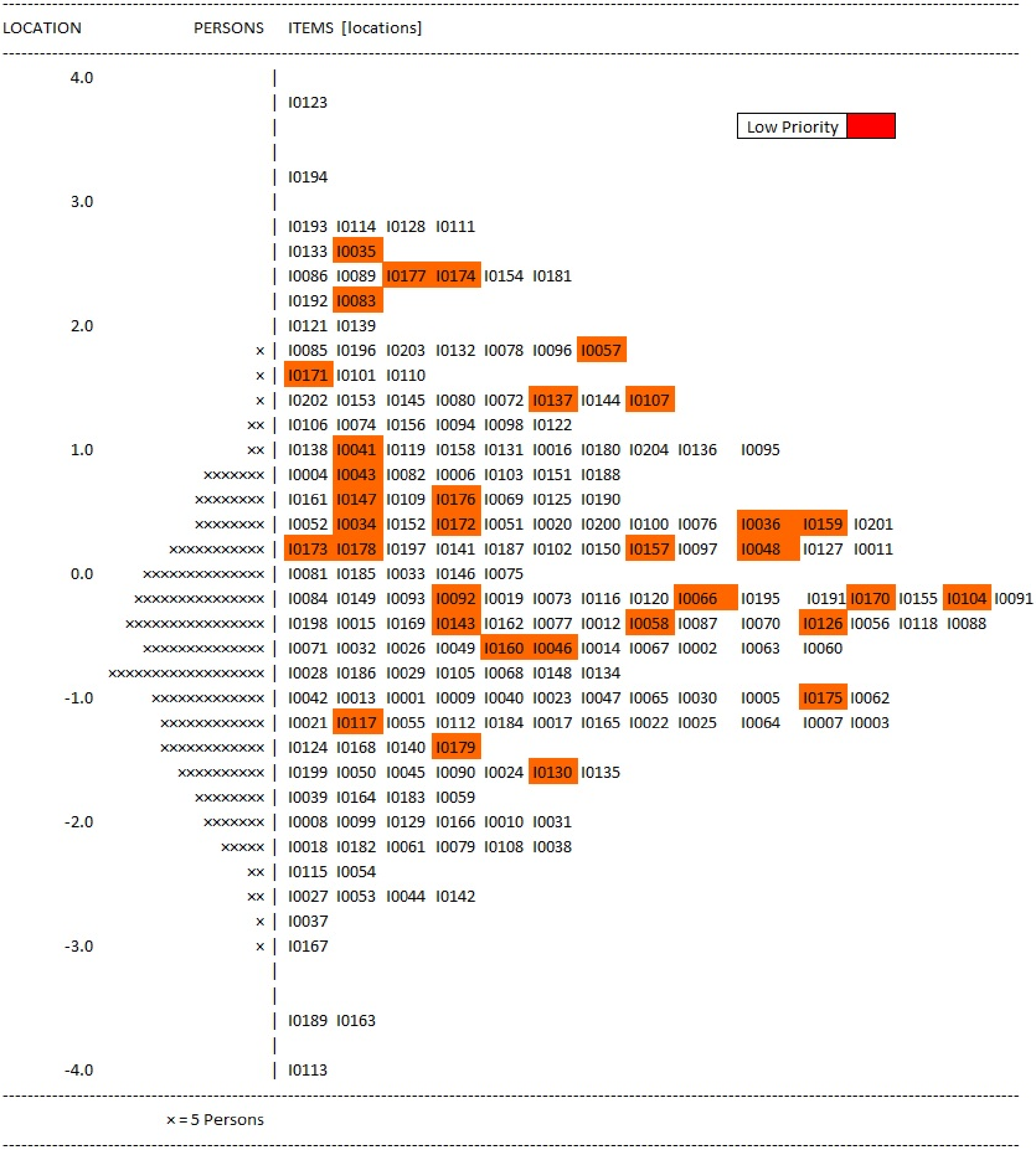
Figure 3.2 Item-person map of 2015 trial items for the SLWG review



As can be seen the removal of the rejected items does not diminish the difficulty range of the available item pool.

Items that were flagged during analysis on multiple criteria were flagged as ‘Low priority’. The highlighted items in Figure 3.3 show the distribution of the ‘Low priority’ items.

Figure 3.3 Item-person map of low priority 2015 trial items



As can be seen even with the ‘Low priority’ items removed there are sufficient items at all levels of difficulty in the remaining pool.

3.1.5. Reports to trial schools

Reports were developed and provided to schools that participated in the trial. The reports were provided to schools in June 2015. They contained a number of sheets; one for each of the eight trial test forms and four inquiry tasks used. Individual students’ results were given for the test form and inquiry task which they completed in the assessment. An information sheet providing advice on interpreting the reports was also included.

3.2. Item selection process for the main study assessment

The development of the main study assessment was done in a series of stages:

1. The pool of trialled objective items and inquiry tasks was reviewed and approved by the SLWG.
2. From the approved pool objective items were selected to create a balanced set of test forms. The item selection was reviewed against the test specifications.
3. Two inquiry tasks were selected.
4. The selection of items and tasks were sent to ACARA for initial approval.
5. The selected items were allocated into online test forms for review in the test delivery system.
6. A process of review and feedback on the selected items was then conducted until the online test forms were approved.

3.2.1. Post-trial review by the SLWG

All trialled items were provided to the SLWG to view in the authoring and review system. Members were invited to view the stimulus and items, as well as the associated metadata, before the upcoming panel discussion in Sydney. They were also provided with a spreadsheet with selected metadata and a summary recommendation as shown in this table.

Table 3.4 Types of data given to SLWG members during post-trial review

Hyperlink	Path to item
Item code	System code for the item
System item type	The IARS item type
Analysis item type	The item type used for analysis (constructed response=scores analysed, multiple choice=option choices analysed)
Historical/New	Historical items were adapted from paper version for online use. New items were written specifically for online delivery.
Year first used	The NAP—SL cycle the item was first developed for
Also in mode-effect study?	Historical items were also used to compare performance with paper-based versions of the items. "In study" item was used in the mode-effect study (historical items). "Not in study" item was not used in the mode-effect study (2015 items).
Unit title	Title for the set of items
Concept area	NAP—SL concept area
Strand code	NAP—SL strand
Item status	Provisional status of the item. Keep: no obvious issues with the item. Review: Some issues that should be considered when including the item in a test. Low priority: The item could be viable but is not ideal for use. Reject: Not a suitable item for inclusion in a test.
Gender DIF comment	Items were compared statistically for different levels of performance between boys and girls. "Significant gender DIF" indicates an item that showed strong evidence of differences in performance between boys and girls. Other statements indicate lesser evidence of differences in performance that may not be significant but which should be

	considered when constructing the final tests.
Comment	Comments on the item that give further elaboration on possible issues with the item or factors that may be of interest to reviewers.

This pool was discussed at a meeting with the SLWG in Sydney on 26 June 2015 and approved for use in the 2015 assessment.

During the meeting SLWG members were presented relevant psychometric data on the items that included:

- facility (per cent correct)
- weighted MNSQ
- discrimination (overall and for each distractor [where applicable])
- gender DIF
- ICC graphs.

3.2.2. Survey

Also at this meeting, the results from the trial NAP—SL student survey were presented. SLWG members were invited to comment on the survey items and provide a priority for inclusion in the final form. SLWG members discussed the results, recommended changes to a number of survey items and agreed on a final list of survey items which will appear as part of the main study.

3.2.3. Inquiry tasks

The four inquiry tasks were discussed in detail. For the main study, two inquiry tasks of approximately 10 items were required and the SLWG discussed the performance of all four tasks to help evaluate which two tasks should be selected. Specifically:

- Bouncing Balls proved very difficult and had one item which did not function well and impacted on the other items in the set. It was agreed that this task would not proceed to the main study.
- Mustard Seeds had a few items with minor issues.
- Pendulums only had one problematic item.
- Sunscreen worked really well and overall the SLWG liked this task.

Overall the Pendulums and Sunscreen tasks were most favoured for inclusion in the main study, with Mustard Seeds as a possible alternative choice instead of Pendulums.

3.2.4. Objective test items

The working group looked at all the objective test items that had been flagged as ‘Review’ and were also invited to raise issues with any other items.

Wording of items was discussed but to ensure that the psychometric properties of items did not change only minor changes in wording were agreed to. Additionally, some items flagged as 'Review' were moved to 'Low priority' due to issues highlighted during discussion. Specifically:

- NSL15E_C007.2 due to poor fit in its item characteristic curve
- NSL15E_F005.3 due to overall difficulty and poor performance of the whole item set
- NSL15E_V001.4 due to a factual problem with the item
- NSL15E_C004.1 due to the numeracy level of the content.

It was also agreed to make some general editorial changes to ensure consistency of style in multiple choices style items by putting references to number of options to be chosen in bold.

3.3. Main study forms

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

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

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

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

3.3.1. Cluster development

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

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

3.3.2. Inquiry task selection

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

The performance of each item in a task was considered against psychometric criteria. The overall difficulty of each task was also considered as well as the spread of difficulty across the items in each task. In addition to these criteria the tasks were also judged in terms of the variety of skills tested and the content areas they covered.

Of the four tasks, Bouncing Balls was deemed to be least effective psychometrically and Sunscreen was deemed to be the most effective. Of the remaining two tasks it was felt that Pendulums provided a better balance of skills in conjunction with Sunscreen than Mustard Seeds would provide.

Consequently, Pendulums and Sunscreen were selected for the main study.

3.4. Distribution of assessment item types in main study

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

The item types equivalent to paper-based were:

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

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

Online interaction types were classified as:

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

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

3.4.1. Item types used in the objective test

These tables show the item types used in the objective test forms.

Table 3.5 Objective test paper-based equivalent item types

Item type: paper-based equivalent	Total
Long constructed response	29
Short constructed response	5
Multiple choice	47
Other	7

Table 3.6 Objective test interaction types

Online Interaction Type	Total
Extended text	27
Multiple choice	40
Multiple choices	4
Hotspot	3
Interactive gap match	2
Interactive graphic gap match	4
Select point	1
Composite	7

Note that composite types combine two interactions but typically included extended text or text entry interactions.

3.4.2. Item types used in the inquiry tasks

Table 3.7 Inquiry task paper-based equivalent item types

Item type: paper-based equivalent	Total
Long constructed response	10
Short constructed response	1
Multiple choice	7
Other	3

Table 3.8 Inquiry task interaction types

QTI Interaction Type	Total
Extended text	9
Multiple choice	6
Multiple choices	1
Interactive graphic gap match	2
Select point	1
Composite	2

As with the objective test, composite types combine two interactions but typically included extended text or text entry interactions.

3.4.3. Item types across both the objective test and inquiry task

This table shows items across both sections classified against both schemes.

Table 3.9 Interaction types and paper-based equivalents

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

3.5. Coverage of scientific literacy in main study

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

3.5.1. Coverage of skills and content using historical framework

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

Table 3.10 Coverage of concept areas

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

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

Table 3.11 Coverage of A, B, C strands

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

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

Table 3.12 Coverage by concept area and A, B, C strands

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

3.5.2. Coverage of skills and content using the Australian Curriculum

Every item developed was classified against an Australian Curriculum code. These classifications were then reviewed by ACARA's science curriculum specialists and adjusted accordingly.

This table shows the items by the main curriculum strands.

Table 3.13 Coverage by curriculum strands

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

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

Table 3.14 Coverage by curriculum strands and curriculum year level

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

When an item was classified against a curriculum code at a year level below Year 5 it was done so as the best fit between the content of the curriculum statement and the content of the item. It does not indicate that the item was necessarily suitable for students at a younger year level. In some cases, an item may include content from an earlier year level but test a more sophisticated understanding of that content in a cognitively complex way.

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

Table 3.15 Coverage by curriculum sub-strand

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

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

3.5.3. Intersection with general capabilities

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

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

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

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

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

The critical and creative thinking (CCT) capability was overtly investigated to identify connections between the CCT framework and the NAP—SL assessment. Each item was compared against the CCT framework and, when there was a strong match between the cognitive skills employed in the item and the framework, the connection was noted. In many cases there was no exact match as a given item might include a variety of cognitive skills or address critical thinking skills in a way that was not easily summarised by a single reference. Therefore it should not be inferred from the items that were not directly matched, that either they required no critical and creative thinking skills or that the skills they did employ were not present in the framework.

Table 3.16 Coverage by critical and creative thinking organising strands

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

CHAPTER 4 SAMPLING PROCEDURES

4.1. Overview

The desired (target) population for the NAP sample assessments – science literacy (NAP—SL) consisted of all students enrolled in Year 6 in Australian schools in 2015.

The sample design for NAP—SL was a two-stage stratified cluster sample. Stratification involves ordering and grouping schools according to different school characteristics (for example, state, sector, size and school location). This helps ensure adequate coverage of all desired school types in the sample.

Stage 1 consisted of selecting schools that had Year 6 students. In this stage, schools were selected with probabilities proportional to their measure of size (MOS). The MOS is related to estimated enrolment size of Year 6 students at the school. This selection procedure is referred to as ‘probability proportional to size’ (PPS) sampling. Stage 2 involved the random selection of Year 6 students within each school selected in Stage 1.

The sample was designed with the aim of achieving a final sample size of approximately 12 000 students who were enrolled at approximately 600 schools across Australia.

Ideally the final sample would result in estimated mean scores for all jurisdictions that were of similar precision. However, it was recognised that reduced sample sizes would be needed for the smaller jurisdictions (that is, Australian Capital Territory, Northern Territory and Tasmania). This is because most schools in the smaller jurisdictions would need to participate to form a large enough sample. Additionally, as there are a number of national and international assessment projects implemented in Australia, many schools from the smaller jurisdictions would need to participate in multiple assessment projects. Consequently, there would be too much administrative burden on the schools if a larger sample size was used in the smaller jurisdictions.

4.2. Target population

The operational definition of the target population was a sampling frame which consisted of a list of all Australian schools and their Year 6 enrolment sizes as supplied by the Australian Curriculum, Assessment and Reporting Authority (ACARA).

Table 4.1 shows the 2015 estimate of the number of educational institutions and students in the sampling frame for each jurisdiction, as provided by ACARA.

Table 4.1 Estimated 2015 Year 6 enrolment figures as provided by ACARA

State/Territory	Institutions	Students	Percentage of students
NSW	2368	87087	32.0
Vic.	1816	65396	24.0
Qld	1364	57378	21.1
WA	858	29504	10.8
SA	600	18888	6.9

Tas.	214	6239	2.3
ACT	99	4290	1.6
NT	158	3289	1.2
Aust.	**7477**	**272071**	**100.0**

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

4.3. School and student non-participation

In large scale assessments of this kind it is important to document reasons for non-participation so that interpretations of the main findings from the study can be appropriately made within the contexts of the assessment.

As in the 2009 and 2012 cycles, the 2015 study made provisions to document the reasons for exemptions and refusals. School exemptions were handled on a case by case basis in liaison with state representatives and ACARA. Student exemptions and refusals were documented by Test Administrators using an online portal. A table showing the drop-down menu options and the definitions for each category is given in section 5.7.1 as Table 5.10.

4.4. Sampling size estimation

For consistency with previous cycles, a sample size of 14 250 students drawn from 633 sampled schools was proposed for 2015 (2012 NAP—SL Technical Report, Table 3.4, p. 21). This figure was based on the sampling approach discussed in 2012 NAP—SL Technical Report.

4.5. Stratification

The sampling frame was partitioned into 24 separate school lists with each list being a unique combination of state and territory (eight) and school type (three – government, Catholic and independent). This explicit stratification was performed to ensure that an adequate number of students were sampled from each school type in each jurisdiction.

Within each of the separate strata, schools were ordered (implicitly stratified) firstly according to their NAPLAN data, then by geographic location and lastly according to the school measure of Year 6 enrolment size.

For most schools, the MOS for a school was set to the 2013 Year 6 enrolment size (ENR) of the school. A school's MOS was adjusted if the school had a small or, alternatively, a very large number of Year 6 students. Whilst sampling methods for both these school sizes are described in more detail in the subsequent sections, in general small schools had their MOS adjusted so that their selection in the sample would not result in excessively large sampling weights. In addition, very large schools had their MOS reduced so that they were not selected more than once.

The number of Queensland students in Year 6 in the sample frame was lower than usual due to a change in school starting age for this cohort. To ensure the MOS was appropriate for Queensland, the Year 5 enrolment was used instead of Year 6 students if the

Queensland school had Year 5 students. If the school had no Year 5 students, Year 6 enrolments were multiplied by a constant so that the total MOS for Queensland Year 6 was approximately equal to the expected cohort size.

Within each sampled school the target number of students to sample was set at 26 which was one larger than the target cluster size (TCS) of 25 students used in previous NAP—SL cycles. The slight increase in the TCS was made to cover the expected increase in non-responses that could occur due to technology issues in schools (for example, computer breakdowns during testing, internet outages etc.).

Schools with an ENR less than the TCS had a MOS set to the average ENR of the same strata (small schools within each jurisdiction). This was performed to prevent excessively large sampling weights and was only applied after stratification had occurred. See Appendix 3 for more details.

4.5.1. Small schools

If a large number of schools that were sampled had an ENR less than the TCS, then the actual number of students sampled could be less than the overall target sample. Schools with enrolment sizes less than the TCS are classified as small schools in both OECD (2012) and IEA (2009). Both studies have different approaches for the treatment of small schools within the sampling frame. In the 2015 NAP—SL, OECD (2012) guidelines were utilised for classifying and stratifying small schools, whilst an adapted version of IEA’s (2009) treatment of small school MOS values was used.

OECD (2012) guidelines were used for classifying and stratifying small schools, which involved deliberately under-sampling small schools and slightly over-sampling large schools. This ensured that small schools were represented in the sample while still achieving an adequate overall student sample size without substantially increasing the total number of schools sampled (see OECD 2012, pp. 68–74).

The MOS for a small school was set to the average ENR of all schools within the same explicit stratum and school size category. This strategy was adapted from the IEA (2009) approach to ensure that selection of very small schools would not result in excessively large sampling weights (see IEA 2009, pp. 85–87, section 5.4.2).

4.5.2. Very large schools

Selecting schools with a probability proportional to size (PPS) can result in a school being sampled more than once if its ENR is sufficiently large. This can occur when the school enrolment size is larger than the explicit stratum sampling interval. To overcome this, very large schools had their MOS set equal to the size of the sampling interval of the explicit stratum that the school belonged to (an option that was utilised in IEA 2009, pp. 85–87, section 5.4.2).

4.6. Replacement schools

Replacement schools were included in the sample to help overcome problems in relation to school non-participation. For example, if the non-participation rate is high, then the target sample sizes will not be achieved. Further, if non-participating schools tend to be lower performing schools, then a bias in the estimated achievement levels will likely occur.

If a school elected not to participate for some reason, then a replacement school was selected for inclusion in the sample. Replacement schools were assigned as per OECD 2005 procedures (p. 60). That is, for a sampled school, the school immediately following it in the sampling frame was assigned as the first replacement school for it, and the school immediately preceding it was assigned as the second replacement school.

4.7. Student selection

Each selected school provided Educational Assessment Australia (EAA) a list containing the details of all eligible Year 6 students. Twenty-six students from the student list were then randomly selected for inclusion in NAP—SL for each school. For schools with 26 students or fewer, all students were included in the study. The random selection of students within schools was a new approach for the 2015 cycle. In previous cycles a Year 6 class was randomly selected where all Year 6 students in the selected class were included in the NAP—SL sample.

4.8. 2015 NAP—SL sample results

Table 4.2 provides a breakdown of the sample at the school level according to jurisdiction.

Table 4.2 School participation rates by jurisdiction

State/Territory	Number of schools sampled	Number of schools that participated	School participation (per cent)
NSW	93	91	97.8
Vic.	92	91	98.9
Qld	91	89	97.8
WA	92	88	95.7
SA	94	86	91.5
Tas.	62	60	96.8
ACT	55	55	100.0
NT	49	39	79.6
Aust.	**628**	**599**	**95.4**

Table 4.3 provides a breakdown of the sample according to jurisdiction. The target sample is the number of Year 6 students *enrolled at the time of testing* in the sampled schools. The achieved sample is the number of selected Year 6 students who participated (attempted the test).

Table 4.3 2015 NAP—SL target and achieved sample sizes by jurisdiction

State/ Territory	Number of selected students enrolled at the time of testing		Number of selected students who participated in the test	
	Students	Percentage of the sample	Students	Percentage of the sample
NSW	2185	15.1	1911	15.4
Vic.	2162	14.9	1930	15.6

Qld	2177	15.0	1833	14.8
WA	2126	14.7	1878	15.1
SA	2178	15.0	1790	14.4
Tas.	1366	9.4	1198	9.7
ACT	1366	9.4	1221	9.8
NT	920	6.4	649	5.2
Aust.	**14480**	**100.0**	**12410**	**100.0**

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

The numbers of non-participating students are provided in Table 4.4, broken down by jurisdiction and reason for non-participation.

Table 4.4 Student non-participation by jurisdiction

State/ Territory	Non-inclusion code					Total
	Absent	Functional disability	Intellectual disability	Limited language proficiency	Student or parent refusal	
NSW	186	0	12	1	10	209
Vic.	154	1	8	0	10	173
Qld	176	0	12	0	7	195
WA	162	1	11	1	5	180
SA	161	0	12	2	35	210
Tas.	106	4	18	2	4	134
ACT	95	0	11	0	17	123
NT	117	2	2	5	1	127
Aust.	**1157**	**8**	**86**	**11**	**89**	**1351**

Additional technical specifications and details related to sampling can be found in Appendices 3 and 4.

CHAPTER 5 TEST ADMINISTRATION PROCEDURES AND DATA PREPARATION

5.1. Registration of students

For most jurisdictions, School Contact Officers nominated by the sample schools were informed that they were to register their students using the templates provided by Educational Assessment Australia (EAA). In some jurisdictions the student registration task was completed centrally.

Table 5.1 Data collection by jurisdiction and sector

State/Territory	Government	Catholic	Independent
NSW	Central	School	School
Vic.	Central	School	School
Qld	Central	School	School
WA	Central	School	School
SA	Central	Central	School
Tas.	Central	Central	School
ACT	Central	Central	Central
NT	Central	School	School

5.2. Pre-assessment preparation

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

5.2.1. Contacting schools

After schools had been selected in the sampling process (see Chapter 4), they were contacted in consultation with the Australian Curriculum, Assessment and Reporting Authority (ACARA) and jurisdictions. Letters were sent to participating schools advising them of their participation in NAP—SL and of the processes involved.

A website was made available to schools as an online portal (see Appendix 8). This portal provided general information to schools including an overview of the assessment program and details of what the school would need to do to participate in the assessment.

The portal included online forms for schools to provide data on students. Schools were also asked about the number of computers (or equivalent devices) that they had available for the assessment. Student data was collected at various stages in conjunction with the progress of the sampling process.

Feedback was sought from schools on their capacity to take part in the assessment. When a school indicated that there may be difficulties with them participating, this was discussed with the jurisdictional representative and ways of resolving the difficulties were investigated. If the difficulties were insurmountable, the school was removed from the sample and an equivalent replacement school was sought. In total 43 schools in the sample were replaced because of such difficulties.

Schools were contacted by email on a regular basis as the project moved through the stages.

Table 5.2 Overview of school participation in the assessment program

Stage	Description	Date
Stage 1	Schools provided a school contact to be responsible for managing the assessment in their school. School contact provided contact details and student lists.	1–29 May 2015
Stage 2	Schools were notified of students selected to participate. School contact provided background details of each selected student, such as language spoken at home, parental education and occupation information (some jurisdictions/sectors may have had these lists uploaded from a central registry and the school contact checked and confirmed details were correct).	20 July – 28 August 2015
Stage 3	A technical readiness test (TRT) and online Pre-practice test were performed prior to the testing period. The school contact ensured that the TRT had been performed on each computer/device to be used for the assessment.	Available from 17 August 2015
	Online pre-practice test was available for students after TRT was completed.	17 August – 19 October 2015
Stage 4	The administration manual was read before the testing period. Login details for the selected students were provided via the NAP—SL school portal.	6 October 2015
Stage 5	Testing period. The test administrator was a class teacher with a reasonable knowledge of the school's computers/devices to be used for the test.	NSW/NT/Qld/Vic. 12–30 October 2015 ACT/SA/Tas./WA 19 October – 6 November 2015
Stage 6	Each school was emailed a school report with feedback on the performance of the student cohort that sat the test.	11 December 2015

5.2.2. Technical readiness test and pre-practice test

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

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

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

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

5.2.3. Helpdesk provision and online support

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

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

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

EAA provided a helpdesk with a toll-free telephone number and an email address to ensure all queries were dealt with promptly.

5.3. Assessment administration procedures

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

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

Teachers were able to review the administration manual before the assessment date and raise questions with the NAP—SL coordinator in their jurisdiction. Extracts from the administration manual are shown in Appendix 5.

5.3.1. School observations

A quality-monitoring program was established to gauge the extent to which class teachers followed the specified administration procedures. This involved trained observers monitoring the administration of the 2015 assessment in a random sample of classes in 33 of the 617 schools involved.

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

This table shows the number of schools visited per state/territory.

Table 5.3 School observations by jurisdiction

State/Territory	Number of schools visited
NSW	5
Vic.	5
Qld	6
WA	4

SA	5
Tas.	3
ACT	2
NT	3
Aust.	**33**

5.3.2. Test administrator feedback

Teachers acting as test administrators provided feedback on the administration of the test. This feedback was collected via the online portal and was then collated. The collated feedback was passed on to ACARA.

5.4. Online delivery of NAP in schools

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

5.4.1. Platforms used in schools

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

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

Table 5.4 Operating systems used by schools in the technical readiness test

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

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

This table shows what percentage of schools used a given browser during the TRT.

Table 5.5 Browsers used by schools in the technical readiness test

Browser	Total per cent
Chrome	34.2
Firefox	4.6

IE	46.7
Safari	14.5

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

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

Table 5.6 Operating systems used across states and territories

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

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

Also there was some regional variation in the browser used.

Table 5.7 Browsers used across states and territories

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

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

5.4.2. Common classroom arrangements used

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

- Some schools had a dedicated IT/computer room. This room was typically arranged either with computers in rows, with students facing a board, or arranged with computers around the edge of the room, sometimes with a centre island.
- Some schools had at least one class-set of devices that could be moved from one classroom to another. In this case the assessment was held either in a normal classroom or in a larger space (such as the school hall or library).
- Some schools had a mix of portable devices and desktop computers held in a library or learning centre. In this case students were seated in different areas of a larger space.
- Some schools operated a ‘bring your own device’ policy in which students brought their own device. Not all schools operating such a policy used student devices for the assessment but rather adopted one of the other methods.

- Some schools used a mix of methods due to there being an insufficient number of devices of any one kind. The students were split into two or more groups and supervised separately (for example, one group on desktop machines in a library and a separate group using notebooks in a classroom).
- In a small number of schools there were either an insufficient number of devices for all students to take the assessment simultaneously or the schools' bandwidth was insufficient for students to access the internet simultaneously. In this case the school ran split sessions.

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

Table 5.8 Room type used during NAP—SL

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

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

5.4.3. Challenges for schools

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

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

1. Logging onto the school network.
2. Logging onto an internet gateway to allow access to the web.
3. Opening a suitable web browser.
4. Navigating to the assessment platform website.
5. Logging onto the assessment platform.

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

5.4.4. Technical issues faced by schools

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

1. Students were locked out of the assessment because they had not logged off correctly during the break between the objective test and the inquiry task. A security

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

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

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

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

5.4.5. USB delivery

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

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

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

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

5.4.6. Student experience of the test

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

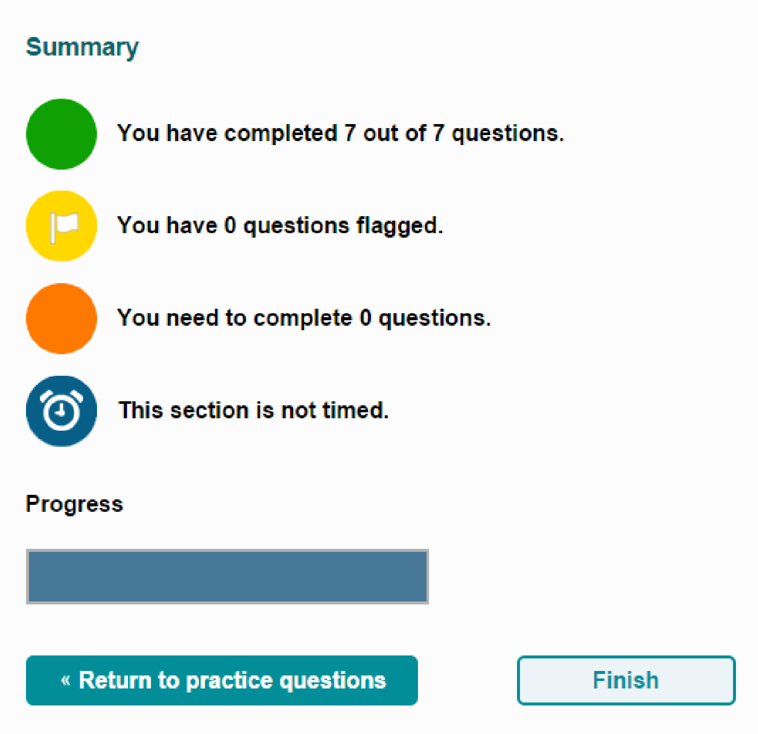
Figure 5.1 Icons shown to students

Students had to take each step in order:

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

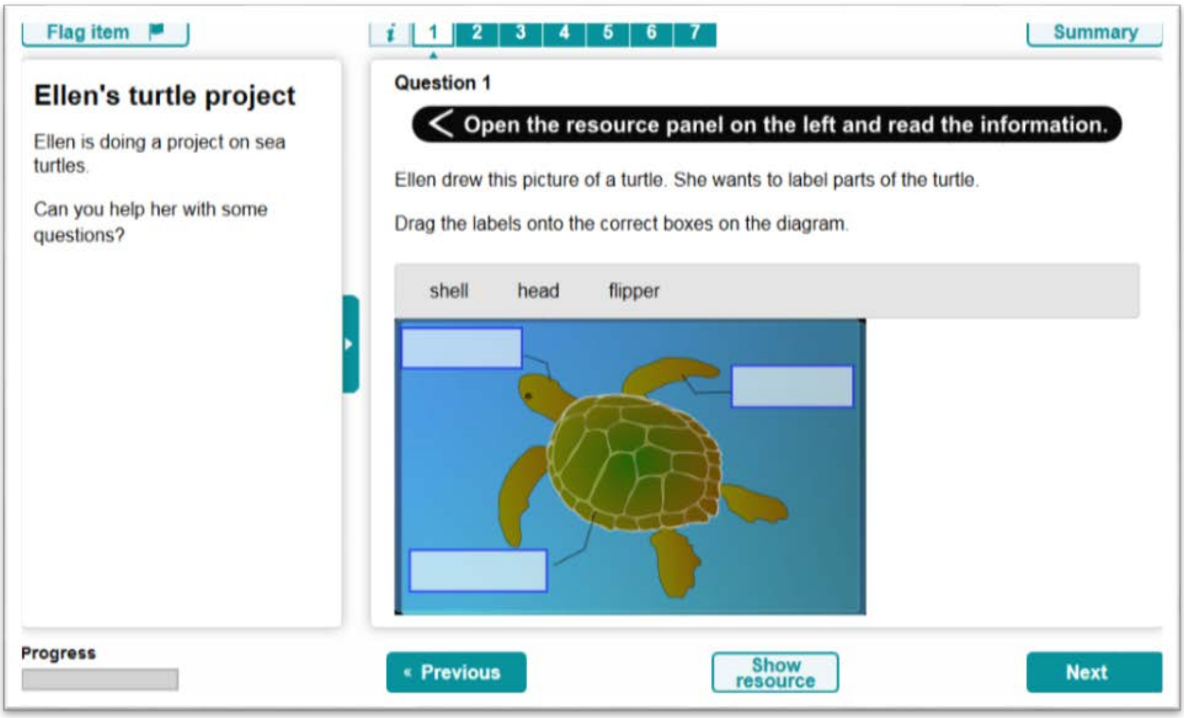
When students completed a step, the system would show them a summary of what they had done. This image shows the summary that appeared at the end of the practice items.

Figure 5.2 Summary screen shown to students



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

Figure 5.3 Sample screen layout of test delivery system



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

5.5. Marking of responses to open-ended items

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

5.5.1. Development of marking guides

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

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

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

5.5.2. Online marking

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

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

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

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

The team leaders underwent a half-day training day presented by the Marking Professional Leader who had a major role in developing the marking guides. During this training, the usual procedure was followed whereby the marking guide for each item was explicated and the Professional Leader responded to any questions from the team leaders. The session involved a brief formal presentation followed by hands-on practice with pre-marked sample student responses. Presentations included leading markers through an overview of sets of items or inquiry task and discussing the marking criteria and illustrative answers for correct and incorrect student responses exemplified in the marking guides.

At the end of the session, the team leaders marked the same set of student answer responses. The scores were compared to those agreed to by expert scorers (the Project Director, Test Development Manager and the Professional Leader). This process gave greater scope for discussion and also resulted in two team leaders being proficient in marking those items even though only one would actually be marking them. This provided extra support for team leaders during the marking operation.

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

Markers then practised marking with a pre-marked sample of items and discussed the scores assigned to each item to help clarify distinctions between score levels. At the end of the session, all markers were asked to mark the same set of student answer responses. The scores were compared to those agreed to by expert scorers (the Project Director, Test Development Manager, the Professional Leader and the team leaders). Team leaders discussed with markers agreements and disagreements between their scores and the scores given by expert scorers. Additional practice was provided to markers for items where consistency and/or accuracy were low.

The above steps were undertaken outside of the online marking system using print outs of selected student responses.

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

The online marking system provided its own "quality marking algorithm" which is described in the manual.

> *The quality rating algorithm for the markers is a rating between 0.0–1.0, with 1.0 being the best possible, and 0.0 being the worst possible.*
>
> *The algorithm essentially measures the extent to which a marker disagrees with the various yardsticks available. For example –*
>
> - *Double marking – each double marking pair is analysed for difference between this marker, and the other marker.*
> - *Senior Marker sampling – the extent of disagreement with any of this markers' scores that have been sampled by senior markers is analysed.*
> - *Monitoring scores – the extent of disagreement with monitoring scripts is analysed.*
>
> *The algorithm will only ever consider the 'most relevant' score. For example, consider the scenario where two double markers A and B have substantial conflict. Subsequently, that conflict was resolved by a senior member of the marking team who sided with Marker A.*
>
> *Marker A and B's scores will no longer be compared against each other, but with the 'correct' score instead (the one from the more senior marker). This means that Marker A's ranking will improve, but Marker B's ranking will remain low.*
>
> *The guideline on interpreting this score is that it's a ranking only. The score itself is meaningless without comparing other markers. (Janison CLS – Extended Marking V1.2 2014 p.40)*

The system also provided information on marker speed.

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

5.6. School summary reports

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

For the first time, Australian Curriculum: science reference codes were provided in the school reports to help schools identify the key curriculum aspects that had been assessed.

Item descriptors were written using the content of the Australian Curriculum: science as a starting point so as to more clearly identify the relevant aspect of the curriculum assessed. A sample report is shown in Appendix 2.

5.7. Data processing procedures

After completion of the marking centre, student response data and marks were exported from the test-delivery/test-marking system. Data were provided in ten separate files, organised by task and date as shown:

- Answer report - NAPSL Main Study - Inquiry Task by date (24Oct-10Nov).xlsx
- Answer report - NAPSL Main Study - Inquiry Task by date (8Oct-23Oct).xlsx
- Answer report - NAPSL Main Study - Objective Test by date (17Oct-23Oct).xlsx
- Answer report - NAPSL Main Study - Objective Test by date (24Oct-30Oct).xlsx
- Answer report - NAPSL Main Study - Objective Test by date (31Oct-10Nov).xlsx
- Answer report - NAPSL Main Study - Objective Test by date (8Oct-16Oct).xlsx
- Answer report - NAPSL Main Study - Practice Questions by date (08Oct-23Oct).xlsx
- Answer report - NAPSL Main Study - Practice Questions by date (24Oct-10Oct).xlsx
- Answer report - NAPSL Main Study - Student Survey by date (24Oct-10Nov).xlsx
- Answer report - NAPSL Main Study - Student Survey by date (8Oct-23Oct).xlsx

The objective test and inquiry task files were then analysed together. The practice question files were not analysed and the student survey files were treated separately.

The data in the objective test and inquiry task files was structured so that each row represented a single student's response to a single online interaction. For composite items (that is, single items that contain two or more interactions), two or more lines of data represented a single interaction.

Each column of the spreadsheet represented a different field of data.

Table 5.9 Data fields used in student response data

Column	Description	Example
A	Jurisdiction: the relevant state or territory	NSW
B	School: the name of the student's school	[Suburb Name] Public School
C	School identifier: a code identifying the school	31077
D	Full name: the full name of the student	Sue Student
E	Username: a code used by the student as a username which consisted of the school code, a dash and a two-digit number	31077-01
F	Student number: an internally generated student identifier used by the system	361421
G	Assessment event identifier: a code used by the system to identify which assessment event the data was taken from	ms-2-objective
H	Test identifier: a code indicating which test form the student attempted	NAP—SL -2015-objective-test-paper-1
I	Item name: The IARS item name created during item authoring. Typically an alphanumeric code starting with NSL	NSL12H551.51C (item) or NSL12H551.51C Interaction 01 (interaction in a composite item)
J	Item identifier: The internal code generated by IARS to identify the item uniquely	x00000581
K	Interaction identifier: a code for each interaction in a composite item	[blank] (item) x00002815-01 (interaction in a

		composite item)
L	Composite number: the position in the test form where the interaction appears, with composite interactions shown as a decimal (initial screen is 1, so this value is one greater than the displayed question number)	19 (item) 19.1 (interaction in a composite item)
M	Item type: the system item type	Composite, Extended text
N	Answer response: the student response as recorded by the system; for example, a multiple-choice item would be recorded as 1, 2, 3, or 4 depending on the option chosen	1 (multiple choice) 0001 (hotspot)
O	Display response: the student response as displayed to the student; for example, a multiple choice item this would be the text of the option they chose	stormy (multiple choice) D (hotspot)
P	Weight: the number of marks allocated to the item, typically 1 except for polytomous items	1, 2
Q	Is answer correct: A true/false field stating whether the answer is correct or not	True, False
R	Mark awarded: the value of the mark awarded	0,1,2
S,T,U,V, W,X	Answer duration, Date answered, Date last submitted, Date attempted, Date completed, Date marked: Various time stamped fields used by the system	15/10/2015 10:14:28 AM
Y	Status: used by the system to track marking status	Marked

EAA's data processing team reprocessed the data files to create files in a matrix format. In this format each row represented a single student's data and each column represented either data about the student or a single item. This format was preferred for use with the data analysis software.

In the raw data exported from the test delivery system, even unanswered items were shown as having a mark of 0. As the system had no distinct field to show an item had not been attempted, this needed to be inferred from other fields during the initial processing. Specifically, the student response fields and the date fields were used to identify when a student had not provided an answer to an item. For composite items, it was necessary to check these fields for all interactions to ascertain whether a student had partly attempted the item.

5.7.1. Student participation data

Student participation codes were recorded by the test administrator via an online portal. Participation was recorded distinctly for the objective test and the inquiry task. The available options were presented as a drop down menu for the test administrator to select.

Table 5.10 NAP—SL exemption and refusal codes

Category	Explanation
Absent	Student was absent for the test.
Functional disability	Student has a moderate to severe permanent physical disability such that he/she cannot perform in the testing situation. Functionally disabled students who can respond to the assessment should be included.

Intellectual disability	Student has a mental or emotional disability and is cognitively delayed such that he/she cannot perform in the testing situation. This includes students who are emotionally or mentally unable to follow even the general instructions of the assessment. Students should NOT be excluded from participating solely because of poor academic performance or disciplinary problems.
Limited test language proficiency	The student is unable to read or speak the language of the assessment (i.e. English) and would be unable to overcome the language barrier in the testing situation. Typically a student who has received less than one year of instruction in the language of the assessment may be excluded.
Student or parent refusal	Parent/caregiver requested that student not participate OR student refused to participate.

CHAPTER 6 COMPUTATION OF SAMPLING WEIGHTS

Sampling weights were calculated for the NAP sample assessments – science literacy (NAP—SL) as the two-stage stratified cluster sample design meant that students had unequal probabilities of selection. Non-response adjustments were also included to take into account schools and students who were sampled, but did not participate in the assessments. Finally, a post-stratification adjustment was included so that weights corresponded to known student population totals.

The following sections detail the steps required to calculate the sampling weights.

6.1. School weight

6.1.1. School base weight

School level base weight for school i

$$BW_{sc}^{i} = \frac{M}{n \cdot m_i} \tag{1}$$

where n was the total number of schools sampled within each explicit stratum and m_i was the measure of size (MOS) assigned to the i^{th} school, and

$$M = \sum_{i=1}^{N} m_i \tag{2}$$

where N was the total number of schools (that is, both sampled and not sampled) in the explicit stratum.

For small school strata, schools were assigned equal MOS values. Small school sampling weights, using the above equations, can be given by:

$$BW_{sc}^{i} = \frac{N \cdot m_i}{n \cdot m_i} \tag{3}$$

This can be simplified to:

$$BW_{sc}^{i} = \frac{N}{n} \tag{4}$$

6.1.2. School non-participation adjustment

School level base weights were calculated for all sampled and replacement schools that satisfied the condition that more than half of the eligible students actually participated in the study. In total, 628 schools were sampled of which there were 29 schools that did not participate in the testing (and could not be replaced).

A school-level non-response adjustment was calculated separately for each explicit stratum to account for schools that were sampled but did not participate. Specifically, the non-response adjustment was calculated as:

$$A_{sc} = \frac{n_s + n_{r1} + n_{r2} + n_{nr}}{n_s + n_{r1} + n_{r2}} \quad (5)$$

where:

n_s was the number of originally sampled schools that participated,

n_{r1} and n_{r2} was the number of first and second replacement schools, respectively, that participated, and

n_{nr} was the number of schools that did not participate.

6.1.3. Final school weight

The final school weight was then the product of the school base weight and non-participation adjustment:

$$FW_{sc}^{i} = BW_{sc}^{i} \cdot A_{sc} \quad (6)$$

6.2. Student weight

6.2.1. Student base weight

For schools with 26 Year 6 enrolments or fewer, the student base weight (BW_{st}) was equal to 1. For schools with more than 26 Year 6 enrolments the base weight was given by:

$$BW_{st}^{i} = \frac{Gr_{06Enr}}{26} \quad (7)$$

Where Gr_{06Enr} is the number of Year 6 students enrolled at school at the time of testing.

6.2.2. Student non-participation adjustment

A student non-participation adjustment was calculated for any school that had at least one student who was sampled and eligible to do the test but did not participate for some reason. These are the absent and refusal students not including exclusions, such as functionally disabled. This was given by:

$$A_{st}^{i} = \frac{s_{rs}^{i} + s_{nr}^{i}}{s_{rs}^{i}} \quad (8)$$

where s_{rs}^{i} was the number of eligible students that participated, and s_{nr}^{i} was the number of eligible students that did not participate, at the i^{th} school.

6.2.3. Final student weight

The final student weight is then equal to the product of the student base weight and non-participation adjustment:

$$FW_{st}^{i} = BW_{st}^{i} \cdot A_{st}^{i} \tag{9}$$

6.3. Pre weight

The pre weight is the product of the final school and student weights:

$$preW^{i} = FW_{sc}^{i} \cdot FW_{st}^{i} \tag{10}$$

6.4. Post stratification adjustment

A post stratification adjustment was included for each explicit stratum. The adjustment was equal to the student population according to Australian Bureau of Statistics enrolment information ($popEst$) divided by the total number of participating students when weighted using the pre weight ($preW$):

$$psAdj = \frac{popEst}{\sum preW} \tag{11}$$

6.5. Final weight

In summary, the final weight is the product of the pre weight and the post stratification adjustment:

$$W^{i} = preW^{i} \cdot psAdj^{i} \tag{12}$$

CHAPTER 7 ITEM ANALYSIS OF THE FINAL TEST

7.1. Item analyses

This chapter presents the item analyses of the 2015 NAP sample assessments – science literacy (NAP—SL) main assessment data.

7.1.1. Sample size

In all, 12 410 students participated in at least the objective component of the NAP—SL test. Table 7.1 shows the number of participating students by state and territory.

Table 7.1 Number of participating students by state and territory

State/Territory	Number of students
NSW	1911
Vic.	1930
Qld	1833
WA	1878
SA	1790
Tas.	1198
ACT	1221
NT	649
Aust.	**12410**

7.1.2. Number of students by test form

Seven test forms with link items were rotated in each class (see section 7.2 for test design). Each student completed only one test form. Table 7.2 shows the number of students that completed each test form. It can be seen that the test rotation scheme worked well, as the number of students per form is approximately equal across the seven forms. As each item appears in three test forms, the number of students taking each item is around 5320.

Table 7.2 Number of students by test form

Form	Number of students
1	1758
2	1781
3	1780
4	1787
5	1775
6	1770
7	1759
Total	**12410**

7.1.3. Initial item analysis

The first item analysis carried out was on all data records. No sampling weights were used. This analysis aimed to detect any items that did not function well. In this analysis, all trailing missing item responses were treated as not-administered, except for the first item following the last non-missing item. Embedded missing responses were treated as incorrect. A complete list of items and their codes can be found in Table 7.3.

7.1.3.1. Item-person map

The following figure shows an item-person map from this analysis.

Figure 7.1 Item-person map of 2015 main study items

```
    Persons          Items
----------------------------------------------------
             |
             |
             |44
             |
             |
             |14
  3          |53 94
             |
             |
             |106
             |
             |107
            X|90
             |47 97
  2         X|37
           XX|22
            X|12 24 105
           XX|11 19 58 64
           XX|8 38 48 80 95 104
         XXXX|27 103
        XXXXX|35 50 102
  1     XXXXX|65 67 68
      XXXXXXX|
       XXXXXX|16 18 79
       XXXXXX|39 43 57 76
      XXXXXXX|20 23 100
     XXXXXXXX|87
    XXXXXXXXX|55 62 86 91
   XXXXXXXXXX|34 61
  0  XXXXXXXX|41 42 60 96
    XXXXXXXXX|82
      XXXXXXX|2 5 49 70 81 85
     XXXXXXXX|4 9 31 33 56
     XXXXXXXX|36 84 101
     XXXXXXXX|6 7 26 45 59 92
      XXXXXXX|17
 -1    XXXXXX|1 10 13 21 30
         XXXX|52 99
         XXXX|3 29 63 93
        XXXXX|66 73 74
          XXX|15
           XX|25 46 78
          XXX|89
           XX|28 40 83
 -2         X|75
            X|32 54 69 71
            X|72
            X|77
             |98
            X|88
             |
             |
 -3          |
             |
             |
             |
             |51
             |
             |
 -4          |
====================================================
```

Each 'X' represents 73.5 students

The vertical scale in Figure 7.1 shows increasing proficiency, with student ability distribution shown in the left panel (indicated by an 'X'). The items are placed in the right panel (indicated by item numbers) in item difficulty order, where items at the top are most difficult.

Figure 7.1 shows that the items cover a wide range of difficulty levels. The average item difficulty is zero logit, while the average ability is –0.1 logit, showing that the match between item difficulties and person abilities is quite good overall.

7.1.3.2. Summary item statistics

Table 7.3 shows summary item statistics for each of the 107 items included in the final analysis and indicates the two items which were deleted.

Table 7.3 Summary item statistics in 2015

Cluster code	2015 item name	2012 item name	Sample size	Omit rate (%)	Discrimination index	Weighted MNSQ	Facility (%)
C1Q01	NSL06H149.49	IDOB149	5266	0.87	0.43	0.95	66.08
C1Q02	NSL06H149.50	IDOB150	5260	2.81	0.40	0.98	52.98
C1Q03	NSL09H308.08	IDOB308	5257	1.01	0.49	0.89	71.33
C1Q04	NSL09H308.09	IDOB309	5247	1.70	0.42	0.96	57.94
C1Q05	NSL09H308.10	IDOB310	5242	0.78	0.34	1.05	54.77
C1Q06	NSL06H084.84	IDOB084	5239	2.27	0.31	1.07	63.39
C1Q07	NSL06H084.85	IDOB085	5236	1.28	0.36	1.02	62.05
C1Q08	NSL06H084.86	IDOB086	5228	9.37	0.23	1.08	21.54
C1Q09	NSL06H084.87	IDOB087	5218	3.32	0.44	0.94	57.67
C1Q10	NSL06H084.88	IDOB088	5211	4.15	0.44	0.92	66.07
C1Q11	NSL09H405.05	IDOB405	5202	1.19	0.18	1.07	20.03
C1Q12	NSL09H405.06	IDOB406	5190	3.66	0.24	1.05	17.90
C1Q13	NSL12H503.03	IDOB503	5180	0.64	0.42	0.94	68.19
C1Q14	NSL12H503.06	IDOB506	5168	2.59	0.23	1.01	5.92
C2Q01	NSL09H360.60	IDOB360	5270	0.34	0.34	0.99	76.60
C2Q02	NSL09H360.62	IDOB363	5267	1.12	0.32	1.02	33.64
C2Q03	NSL12H559.59	IDOB559	5266	1.14	0.45	0.92	66.35
C2Q04	NSL12H559.61	IDOB561	5259	5.08	0.44	0.93	32.63
C2Q05	NSL12H559.62	IDOB562	5251	2.97	0.39	0.94	21.22
C2Q06	NSL12H564.64C	IDOB564	5245	3.26	0.38	1.09	41.46
C2Q07	NSL12H564.65	IDOB565	5233	0.94	0.40	0.97	67.28
C2Q08	NSL12H564.66C	IDOB566	5221	1.15	0.33	0.98	16.93
C2Q09	NSL12H564.69	IDOB569	5214	3.97	0.40	0.97	38.36
C2Q10	NSL12H564.70	IDOB570	5191	6.09	0.26	1.05	18.07
C2Q11	NSL06H041.41	IDOB041	5170	0.60	0.25	1.06	79.56
C2Q12	NSL06H041.44	IDOB044	5159	1.69	0.34	1.03	61.97
C3Q01	NSL15E_C006.2		5282	0.40	0.27	1.06	23.76
C3Q02	NSL15E_C006.1		5279	0.49	0.36	0.93	82.93
C3Q03	NSL15E_W002.1		5274	0.59	0.41	0.95	71.44
C3Q04	NSL15E_W002.2		5268	0.78	0.38	0.99	67.07
C3Q05	NSL15E_W002.3		5265	1.25	0.48	0.91	56.73
C3Q06	NSL15E_W002.4		5258	1.33	0.37	0.91	84.80
C3Q07	NSL15E_W002.5		5256	0.53	0.31	1.06	58.75
C3Q08	NSL15E_Z002.2C		5255	2.47	0.39	1.00	45.58

Cluster code	2015 item name	2012 item name	Sample size	Omit rate (%)	Discrimination index	Weighted MNSQ	Facility (%)
C3Q09	NSL15E_Z002.4		5247	1.66	0.13	1.18	27.60
C3Q10	NSL15E_Z007.2		5230	0.57	0.33	1.04	61.22
C3Q11	NSL15E_Z007.3		5218	5.48	0.42	0.92	16.10
C3Q12	NSL15E_Z007.4		5165	2.81	0.38	1.05	19.18
C4Q01	NSL12H517.17	IDOB517	5285	6.51	0.38	1.00	35.52
C4Q02	NSL12H517.18	IDOB518	5277	0.45	0.37	0.94	82.49
C4Q03	NSL12H517.19	IDOB519	5272	1.76	0.39	1.00	50.32
C4Q04	NSL12H517.21	IDOB521	5267	0.66	0.23	1.14	49.12
C4Q05	NSL12H517.22	IDOB522	5261	1.82	0.39	0.98	36.61
C4Q06	NSL12H551.51C	IDOB551	5250	2.50	0.13	1.13	4.36
C4Q07	NSL12H551.52	IDOB552	5240	0.71	0.36	1.02	63.53
C4Q08	NSL12H551.53	IDOB553	5229	1.36	0.41	0.92	78.14
C4Q09	NSL12H551.54	IDOB554	5220	8.54	0.15	1.10	13.62
C4Q10	NSL06H021.21	IDOB021	5196	1.35	0.26	1.04	22.59
C4Q11	NSL06H021.22	IDOB022	5178	1.80	0.36	1.03	53.40
C4Q12	NSL06H021.23	IDOB023	5160	2.09	0.21	1.11	25.56
C5Q01	NSL15E_F001.1		5302	0.40	0.17	0.94	95.87
C5Q02	NSL15E_F001.2		5298	1.06	0.40	0.98	68.93
C5Q03	NSL15E_F001.4C		5295	4.15	0.27	0.99	6.76
C5Q04	NSL15E_M003.1		5294	0.81	0.42	0.88	84.36
C5Q05	NSL15E_M003.2		5284	2.59	0.53	0.86	44.36
C5Q06	NSL15E_A006.1		5279	0.55	0.41	0.98	57.40
C5Q07	NSL15E_A006.3		5275	1.55	0.45	0.93	36.74
C5Q08	NSL15E_W006.1		5271	2.12	0.35	0.97	20.36
C5Q09	NSL15E_W006.2		5266	0.85	0.28	1.10	63.01
C5Q10	NSL15E_W006.3		5261	2.34	0.38	1.02	48.58
C5Q11	NSL15E_F004.2		5252	0.57	0.21	1.17	47.37
C5Q12	NSL15E_F004.3		5234	4.09	0.44	0.95	43.06
C6Q01	NSL15E_H009.1		5286	0.72	0.35	1.01	71.02
C6Q02	NSL15E_H009.3	Deleted from analysis due to poor fit					
C6Q03	NSL15E_E001.1		5278	4.91	0.34	1.08	23.80
C6Q04	NSL15E_E001.4		5275	10.58	0.34	1.00	28.08
C6Q05	NSL15E_H004.1		5266	0.85	0.43	0.93	74.31
C6Q06	NSL15E_H004.2		5258	4.96	0.42	0.93	27.33
C6Q07	NSL15E_H004.3		5251	5.01	0.29	1.04	26.78
C6Q08	NSL15E_H002.1		5243	0.90	0.21	1.09	84.89
C6Q09	NSL15E_H002.2		5240	3.15	0.37	1.02	52.94
C6Q10	NSL15E_H002.3		5233	0.86	0.19	1.09	84.48
C6Q11	NSL15E_A007.1		5228	0.63	0.27	1.04	85.69
C6Q12	NSL15E_A007.2		5217	1.69	0.32	1.04	74.60
C6Q13	NSL15E_A007.8		5210	1.19	0.36	1.00	72.55
C7Q01	NSL15E_H007.1C		5288	0.21	0.26	1.07	82.66
C7Q02	NSL15E_H007.3C		5288	0.42	0.30	1.06	36.02
C7Q03	NSL15E_V001.1		5283	0.66	0.36	0.95	87.24

Cluster code	2015 item name	2012 item name	Sample size	Omit rate (%)	Discrimination index	Weighted MNSQ	Facility (%)
C7Q04	NSL15E_V001.2		5275	1.59	0.44	0.92	77.52
C7Q05	NSL15E_H008.4		5270	7.93	0.42	0.95	32.37
C7Q06	NSL15E_Z027.1		5254	4.61	0.37	0.94	21.47
C7Q07	NSL15E_Z027.2		5242	1.47	0.29	1.10	54.64
C7Q08	NSL15E_M004.1		5233	0.80	0.41	0.99	52.55
C7Q09	NSL15E_M004.3		5229	0.69	0.37	0.97	80.80
C7Q10	NSL15E_M004.4		5220	1.17	0.29	1.11	60.42
C7Q11	NSL15E_M004.5		5215	0.92	0.34	1.05	54.55
C7Q12	NSL15E_F003.3		5204	1.19	0.41	0.97	44.27
C7Q13	NSL15E_F003.4		5194	1.17	0.44	0.93	39.64
CCQ01	NSL15Esi3-1.1		6036	2.70	0.27	1.02	89.79
CCQ02	NSL15Esi3-1.2		6036	2.52	0.31	1.04	78.99
CCQ03	NSL15Esi3-1.4		6036	6.44	0.25	0.99	11.91
CCQ04	NSL15Esi3-2.1		6036	2.80	0.32	1.05	42.74
CCQ05	NSL15Esi3-3.1		6036	3.15	0.41	0.99	62.38
CCQ06	NSL15Esi3-4.1		6036	3.36	0.35	1.02	72.08
CCQ07	NSL15Esi3-5.1		6036	7.42	0.21	0.95	6.13
CCQ08	NSL15Esi3-5.2		6036	6.28	0.38	0.94	21.19
CCQ09	NSL15Esi3-5.3		6036	5.65	0.42	0.97	48.51
CCQ10	NSL15Esi3-5.5		6036	10.52	0.39	0.89	13.19
CDQ01	NSL15Esi5-1.1		6057	2.36	0.30	0.99	89.07
CDQ02	NSL15Esi5-1.2		6057	3.35	0.31	1.06	70.17
CDQ03	NSL15Esi5-2.1		6057	4.09	0.44	0.93	38.43
CDQ04	NSL15Esi5-3.1		6057	3.30	0.40	0.99	58.59
CDQ05	NSL15Esi5-4.1		6057	3.85	0.33	0.99	25.08
CDQ06	NSL15Esi5-4.2	Deleted from analysis due to poor fit					
CDQ07	NSL15Esi5-4.3		6057	4.82	0.33	0.98	22.83
CDQ08	NSL15Esi5-4.4		6057	5.37	0.26	1.03	21.15
CDQ09	NSL15Esi5-5.1C		6057	4.21	0.29	0.99	17.91
CDQ10	NSL15Esi5-5.2C		6057	4.66	0.24	0.97	8.55
CDQ11	NSL15Esi5-5.3		6057	7.71	0.32	1.01	11.67

7.1.3.3. Test reliability

Person separation reliability for the 2015 NAP—SL tests is 0.88. In comparison, the reported reliability for PISA 2003 mathematics is 0.85, and 0.89 for TIMSS 2003 Grade 8 mathematics.

7.1.4. Test form effect

'Test form effect' refers to the differences in test form difficulties after equating of the forms has been carried out. That is, students may be advantaged or disadvantaged by taking a particular test form, even after forms have been equated. Table 7.4 shows the test form difficulty estimates. The estimation of test form adjustments was carried out through a ConQuest analysis with the model statement:

test form + item + item*step

Table 7.4 Test form difficulty parameters

Test form	Test form parameter (logit)	Error
1	0.012	0.006
2	–0.011	0.006
3	0.036	0.006
4	–0.020	0.006
5	0.015	0.006
6	–0.005	0.006
7	–0.026	0.015

The test form parameters shown in Table 7.4 are very close to zero, indicating that test form effect was not a serious issue for this assessment. It is noted that test form 7 seems to be somewhat easier and test form 3 appears to be more difficult than the other test forms. However, in estimating the student proficiency levels, the test form effect was taken into account. To do so, the test form effect was set as one of the model parameters in estimating the student parameters in ConQuest.

7.1.5. Item statistics by state and territory

While the items worked quite well in general for the overall sample, it is important to check if the items performed well within each state and territory, and whether the item difficulties are similar across states and territories

The following table shows the item location for each item by state and territory.

Table 7. 5 Item locations by state and territory

Cluster code	NSW	Vic.	Qld	WA	SA	Tas.	ACT	NT
C1Q01	-0.93	-0.87	-0.91	-0.82	-0.92	-0.70	-1.15	-0.93
C1Q02	-0.40	-0.36	-0.21	-0.25	-0.15	-0.09	-0.28	-0.30
C1Q03	-0.97	-1.12	-1.25	-1.21	-1.17	-1.17	-1.47	-1.33
C1Q04	-0.29	-0.53	-0.58	-0.48	-0.37	-0.34	-0.50	-0.29
C1Q05	-0.07	-0.38	-0.37	-0.44	-0.31	-0.18	-0.32	-0.52
C1Q06	-0.66	-0.64	-0.88	-0.63	-0.91	-0.44	-0.90	-0.88
C1Q07	-0.61	-0.52	-0.89	-0.74	-0.70	-0.59	-0.70	-0.93
C1Q08	1.44	1.63	1.13	1.42	1.44	1.35	1.49	1.92
C1Q09	-0.65	-0.53	-0.46	-0.47	-0.22	-0.65	-0.30	-0.32
C1Q10	-0.94	-0.84	-0.98	-0.47	-0.90	-0.94	-0.99	-0.66
C1Q11	1.46	1.46	1.75	1.70	1.59	1.77	1.36	1.26
C1Q12	1.66	1.83	1.86	1.63	1.72	1.72	1.44	2.09
C1Q13	-0.91	-0.88	-0.89	-0.91	-0.94	-0.88	-1.12	-0.98
C1Q14	3.27	2.86	3.03	3.06	3.50	2.79	3.01	3.90
C2Q01	-1.59	-1.36	-1.67	-1.60	-1.47	-1.39	-1.42	-1.45

Cluster code	NSW	Vic.	Qld	WA	SA	Tas.	ACT	NT
C2Q02	0.60	0.67	0.71	0.71	0.86	0.72	0.83	0.27
C2Q03	-1.06	-0.94	-0.87	-0.92	-0.86	-1.06	-0.81	-0.65
C2Q04	0.78	0.71	0.96	0.65	0.89	0.40	0.79	0.96
C2Q05	1.42	1.56	1.79	1.39	1.52	1.03	1.43	1.40
C2Q06	0.37	0.69	0.34	0.30	0.31	0.53	0.58	0.39
C2Q07	-0.79	-0.96	-1.05	-0.81	-1.04	-0.83	-0.97	-0.88
C2Q08	1.57	1.89	1.76	1.94	1.88	1.59	1.80	1.91
C2Q09	0.44	0.58	0.37	0.70	0.65	0.18	0.49	0.35
C2Q10	1.71	1.66	1.60	1.74	1.75	1.74	1.74	1.47
C2Q11	-1.67	-1.70	-1.86	-1.57	-1.56	-1.22	-1.83	-1.81
C2Q12	-0.64	-0.59	-0.61	-0.72	-0.60	-0.73	-0.54	-0.56
C3Q01	1.20	1.43	1.24	1.22	1.17	1.24	1.35	0.98
C3Q02	-2.15	-1.76	-2.12	-1.98	-2.20	-1.77	-2.00	-2.09
C3Q03	-1.37	-1.21	-1.22	-1.24	-1.05	-1.01	-1.21	-1.02
C3Q04	-1.06	-0.99	-0.97	-0.83	-0.91	-0.91	-1.00	-0.87
C3Q05	-0.31	-0.33	-0.41	-0.54	-0.52	-0.32	-0.52	-0.53
C3Q06	-2.23	-2.08	-1.98	-2.02	-2.14	-2.33	-1.98	-2.30
C3Q07	-0.61	-0.35	-0.33	-0.57	-0.70	-0.51	-0.22	-0.89
C3Q08	0.48	0.52	-0.45	-0.05	0.04	0.16	0.00	0.18
C3Q09	1.26	1.29	0.76	0.93	1.08	1.18	0.99	0.82
C3Q10	-0.67	-0.63	-0.63	-0.71	-0.74	-0.26	-0.40	-1.09
C3Q11	1.99	1.96	1.75	1.86	1.87	1.65	1.88	1.60
C3Q12	1.54	1.61	1.33	1.32	1.53	1.38	1.58	1.36
C4Q01	0.41	0.88	0.88	0.74	0.52	0.29	0.56	0.53
C4Q02	-1.92	-1.87	-2.10	-1.92	-1.77	-1.76	-1.94	-1.74
C4Q03	-0.13	0.08	-0.25	-0.07	-0.12	-0.15	-0.04	-0.30
C4Q04	-0.05	-0.10	0.03	0.07	-0.14	-0.03	-0.17	0.05
C4Q05	0.51	0.67	0.47	0.57	0.60	0.69	0.47	0.27
C4Q06	3.18	3.01	3.39	3.78	3.24	3.41	3.63	3.39
C4Q07	-0.71	-0.90	-0.77	-0.88	-0.83	-0.52	-0.82	-0.07
C4Q08	-1.51	-1.72	-1.49	-1.47	-1.71	-1.56	-1.54	-1.16
C4Q09	2.20	2.12	2.22	1.96	1.74	1.94	2.20	1.88
C4Q10	1.35	1.25	1.34	1.41	1.42	1.55	1.44	1.40
C4Q11	-0.26	-0.27	-0.27	-0.17	-0.07	-0.30	-0.02	-0.47
C4Q12	1.34	1.12	1.24	1.26	1.01	1.04	1.30	1.12
C5Q01	-3.73	-3.79	-3.71	-3.43	-3.94	-3.28	-3.15	-3.01
C5Q02	-0.81	-1.04	-0.94	-1.16	-0.99	-1.15	-1.41	-1.27
C5Q03	2.70	2.79	3.27	3.05	3.00	2.82	3.09	2.93
C5Q04	-2.06	-2.14	-2.04	-2.01	-2.09	-2.12	-2.03	-2.16

Cluster code	NSW	Vic.	Qld	WA	SA	Tas.	ACT	NT
C5Q05	0.32	0.30	0.08	0.19	0.32	0.06	0.04	0.25
C5Q06	-0.26	-0.47	-0.48	-0.34	-0.38	-0.33	-0.65	-0.68
C5Q07	0.58	0.35	0.62	0.74	0.72	0.61	0.46	0.63
C5Q08	1.33	1.38	1.60	1.59	1.89	1.47	1.62	1.99
C5Q09	-0.59	-0.71	-0.82	-0.88	-0.84	-0.56	-0.62	-0.73
C5Q10	-0.17	0.06	-0.07	-0.04	0.01	0.10	0.16	-0.01
C5Q11	0.10	0.18	-0.07	0.10	0.02	0.26	0.08	-0.45
C5Q12	0.32	0.36	0.30	0.29	0.19	0.15	0.06	0.47
C6Q01	-1.13	-1.15	-1.11	-1.32	-1.28	-1.04	-1.03	-1.23
C6Q02	Deleted from analysis due to poor fit							
C6Q03	1.62	1.68	1.43	1.49	1.43	1.42	1.99	1.60
C6Q04	1.19	1.31	0.80	0.92	0.92	0.98	0.61	1.61
C6Q05	-1.43	-1.34	-1.43	-1.20	-1.52	-1.11	-1.29	-1.23
C6Q06	1.14	1.12	1.12	0.91	1.06	1.02	1.22	1.33
C6Q07	1.26	1.16	1.17	1.03	1.01	0.86	1.29	1.21
C6Q08	-2.13	-2.29	-2.02	-1.88	-2.30	-1.67	-1.86	-2.73
C6Q09	-0.11	-0.08	-0.28	-0.09	-0.22	-0.31	-0.31	-0.56
C6Q10	-2.01	-2.02	-1.97	-1.84	-2.17	-2.00	-2.02	-1.83
C6Q11	-2.10	-2.15	-1.85	-2.06	-2.34	-2.00	-2.19	-2.36
C6Q12	-1.24	-1.35	-1.42	-1.19	-1.50	-1.30	-1.54	-0.99
C6Q13	-1.08	-1.24	-1.30	-1.20	-1.28	-1.08	-1.11	-1.36
C7Q01	-1.64	-1.97	-2.03	-2.03	-2.03	-1.88	-1.98	-1.81
C7Q02	0.60	0.34	0.49	0.69	0.56	0.73	0.64	1.09
C7Q03	-2.24	-2.59	-2.30	-2.25	-2.39	-2.32	-2.27	-1.89
C7Q04	-1.41	-1.74	-1.51	-1.50	-1.52	-1.45	-1.61	-1.63
C7Q05	0.84	0.84	0.89	0.84	0.69	0.83	0.83	0.67
C7Q06	1.59	1.75	1.11	1.37	1.61	1.46	1.56	1.23
C7Q07	-0.43	-0.43	-0.48	-0.18	-0.06	-0.23	-0.42	-0.20
C7Q08	-0.37	-0.23	-0.18	-0.16	-0.10	-0.06	0.06	-0.73
C7Q09	-1.58	-1.89	-1.90	-1.50	-1.83	-1.62	-1.67	-1.78
C7Q10	-0.68	-0.61	-0.42	-0.45	-0.71	-0.48	-0.52	-0.73
C7Q11	-0.24	-0.33	-0.28	-0.30	-0.23	-0.16	-0.36	-0.57
C7Q12	0.19	0.18	0.09	0.36	0.23	0.25	0.32	0.28
C7Q13	0.55	0.52	0.42	0.54	0.36	0.52	0.11	0.57
CCQ01	-3.02	-2.65	-2.78	-3.07	-2.78	-2.92	-2.65	-2.71
CCQ02	-1.91	-1.89	-1.40	-1.75	-1.64	-1.97	-1.78	-2.08
CCQ03	2.22	2.34	2.34	2.13	2.41	2.07	2.22	2.31
CCQ04	0.22	0.28	0.58	0.22	0.20	-0.05	0.16	0.05

Cluster code	NSW	Vic.	Qld	WA	SA	Tas.	ACT	NT
CCQ05	-0.71	-0.92	-0.39	-0.87	-0.76	-0.79	-0.96	-0.73
CCQ06	-1.35	-1.45	-1.06	-1.44	-1.34	-1.26	-1.61	-1.51
CCQ07	2.86	2.87	3.65	2.55	3.12	2.86	3.37	3.53
CCQ08	1.35	1.37	1.21	1.48	1.75	1.49	1.48	1.55
CCQ09	0.06	-0.16	-0.01	-0.35	-0.04	0.33	-0.20	-0.18
CCQ10	2.19	2.35	1.97	1.79	2.28	1.97	2.06	2.23
CDQ01	-2.88	-2.88	-2.17	-2.77	-2.57	-2.71	-3.05	-2.73
CDQ02	-1.14	-1.41	-0.97	-1.31	-1.19	-1.26	-1.11	-1.04
CDQ03	0.89	0.38	0.14	0.02	0.44	0.57	0.50	0.86
CDQ04	-0.63	-0.70	-0.50	-0.43	-0.59	-0.48	-0.68	-0.90
CDQ05	1.15	1.28	1.47	1.03	1.24	0.61	1.39	0.99
CDQ06	Deleted from analysis due to poor fit							
CDQ07	1.45	1.22	1.42	1.26	1.45	1.09	1.21	1.48
CDQ08	1.18	1.22	1.62	1.57	1.84	1.03	1.59	1.79
CDQ09	1.63	1.83	1.91	1.61	1.41	1.65	1.54	1.23
CDQ10	2.12	2.64	2.65	2.91	2.94	2.52	2.60	2.88
CDQ11	2.44	2.46	2.11	2.47	2.21	2.18	2.63	2.56

For twenty-three items, the discrimination index fell below 0.2 in some states or territories. In most cases there were items with either very high or very low percent-correct values nationally. Four items with national percentage correct values between 75 per cent and 25 per cent had a discrimination index of less than 0.2 for at least one state.

Table 7.6 Items with low discrimination index in specific states or territories

Cluster code	2015 item name	2012 item name	Minimum discrimination index	State with the minimum value
C3Q09	NSL15E_Z002.4		0.06	SA
C4Q04	NSL12H517.21	IDOB521	0.15	NT
C4Q12	NSL06H021.23	IDOB023	0.17	SA
C5Q11	NSL15E_F004.2		0.18	VIC

Although the content of each item was examined, it was unclear why these items would perform differently between states and territories.

The following tables show the discrimination indices for each state and territory as well as the minimum value for the discrimination index and the state or territory in which the item was least discriminating.

Table 7. 7 Discrimination index by state and territory

Cluster code	Discrimination index								Minimum discrimination index	State with the minimum value
	NSW	Vic.	Qld	WA	SA	Tas.	ACT	NT		
C1Q01	0.43	0.39	0.42	0.46	0.37	0.34	0.41	0.49	0.34	TAS
C1Q02	0.38	0.39	0.33	0.37	0.44	0.37	0.34	0.42	0.33	QLD
C1Q03	0.46	0.50	0.48	0.49	0.47	0.52	0.42	0.40	0.40	NT
C1Q04	0.37	0.40	0.36	0.47	0.36	0.36	0.35	0.44	0.35	ACT
C1Q05	0.34	0.30	0.30	0.37	0.32	0.27	0.34	0.28	0.27	TAS
C1Q06	0.22	0.26	0.30	0.32	0.27	0.35	0.35	0.38	0.22	NSW
C1Q07	0.33	0.37	0.38	0.41	0.31	0.33	0.41	0.33	0.31	SA
C1Q08	0.27	0.21	0.24	0.20	0.22	0.28	0.23	0.17	0.17	NT
C1Q09	0.43	0.38	0.42	0.48	0.35	0.50	0.42	0.44	0.35	SA
C1Q10	0.40	0.41	0.41	0.42	0.37	0.44	0.40	0.59	0.37	SA
C1Q11	0.26	0.18	0.13	0.10	0.11	0.17	0.25	0.16	0.10	WA
C1Q12	0.25	0.23	0.20	0.20	0.24	0.29	0.27	0.10	0.10	NT
C1Q13	0.40	0.34	0.36	0.47	0.46	0.39	0.33	0.55	0.33	ACT
C1Q14	0.22	0.23	0.19	0.23	0.14	0.23	0.24	0.11	0.11	NT
C2Q01	0.31	0.29	0.27	0.28	0.35	0.28	0.32	0.43	0.27	QLD
C2Q02	0.39	0.27	0.27	0.27	0.30	0.36	0.33	0.26	0.26	NT
C2Q03	0.40	0.43	0.42	0.46	0.40	0.37	0.47	0.43	0.37	TAS
C2Q04	0.41	0.35	0.38	0.39	0.37	0.49	0.42	0.45	0.35	VIC
C2Q05	0.38	0.33	0.32	0.34	0.30	0.47	0.35	0.37	0.30	SA
C2Q06	0.33	0.34	0.35	0.33	0.35	0.41	0.38	0.46	0.33	WA
C2Q07	0.39	0.36	0.34	0.35	0.37	0.43	0.41	0.44	0.34	QLD
C2Q08	0.31	0.24	0.31	0.24	0.31	0.35	0.32	0.27	0.24	VIC
C2Q09	0.39	0.33	0.37	0.34	0.32	0.47	0.39	0.44	0.32	SA
C2Q10	0.30	0.22	0.27	0.22	0.25	0.25	0.24	0.30	0.22	VIC
C2Q11	0.19	0.18	0.24	0.27	0.24	0.27	0.31	0.37	0.18	VIC
C2Q12	0.31	0.33	0.24	0.31	0.29	0.33	0.37	0.39	0.24	QLD
C3Q01	0.23	0.26	0.23	0.21	0.17	0.30	0.24	0.35	0.17	SA
C3Q02	0.29	0.35	0.30	0.35	0.33	0.32	0.35	0.43	0.29	NSW
C3Q03	0.34	0.31	0.35	0.37	0.39	0.31	0.43	0.42	0.31	TAS
C3Q04	0.38	0.35	0.34	0.38	0.34	0.35	0.39	0.44	0.34	QLD
C3Q05	0.46	0.44	0.45	0.43	0.46	0.41	0.42	0.43	0.41	TAS
C3Q06	0.38	0.34	0.40	0.40	0.40	0.33	0.42	0.45	0.33	TAS
C3Q07	0.33	0.35	0.30	0.31	0.33	0.38	0.33	0.32	0.30	QLD
C3Q08	0.33	0.28	0.36	0.34	0.39	0.34	0.32	0.34	0.28	VIC
C3Q09	0.09	0.11	0.13	0.12	0.06	0.16	0.15	0.18	0.06	SA
C3Q10	0.26	0.34	0.31	0.30	0.35	0.41	0.33	0.34	0.26	NSW
C3Q11	0.44	0.36	0.35	0.39	0.41	0.51	0.37	0.42	0.35	QLD
C3Q12	0.33	0.38	0.35	0.33	0.35	0.39	0.31	0.40	0.31	ACT

Cluster code	Discrimination index								Minimum discrimination index	State with the minimum value
	NSW	VIC	QLD	WA	SA	TAS	ACT	NT		
C4Q01	0.33	0.36	0.27	0.34	0.35	0.30	0.30	0.29	0.27	QLD
C4Q02	0.29	0.32	0.30	0.29	0.31	0.31	0.36	0.41	0.29	NSW
C4Q03	0.32	0.36	0.29	0.34	0.34	0.30	0.34	0.44	0.29	QLD
C4Q04	0.19	0.18	0.17	0.24	0.17	0.28	0.23	0.15	0.15	NT
C4Q05	0.34	0.35	0.30	0.35	0.31	0.30	0.37	0.37	0.30	QLD
C4Q06	0.13	0.10	0.09	0.13	0.12	0.06	0.10	0.12	0.06	TAS
C4Q07	0.25	0.28	0.26	0.36	0.31	0.33	0.39	0.44	0.25	NSW
C4Q08	0.36	0.37	0.34	0.34	0.28	0.35	0.38	0.49	0.28	SA
C4Q09	0.16	0.11	0.13	0.16	0.11	0.17	0.12	0.32	0.11	SA
C4Q10	0.21	0.21	0.28	0.17	0.21	0.21	0.27	0.28	0.17	WA
C4Q11	0.32	0.25	0.30	0.31	0.31	0.29	0.35	0.45	0.25	VIC
C4Q12	0.18	0.25	0.19	0.21	0.17	0.18	0.17	0.26	0.17	SA
C5Q01	0.18	0.18	0.13	0.22	0.17	0.30	0.15	0.38	0.13	QLD
C5Q02	0.31	0.38	0.38	0.37	0.39	0.35	0.32	0.30	0.30	NT
C5Q03	0.28	0.24	0.22	0.23	0.29	0.27	0.17	0.24	0.17	ACT
C5Q04	0.41	0.34	0.42	0.40	0.40	0.47	0.37	0.40	0.34	VIC
C5Q05	0.46	0.44	0.46	0.43	0.47	0.51	0.44	0.47	0.43	WA
C5Q06	0.38	0.38	0.38	0.38	0.41	0.44	0.38	0.29	0.29	NT
C5Q07	0.43	0.32	0.39	0.38	0.44	0.51	0.41	0.36	0.32	VIC
C5Q08	0.29	0.30	0.28	0.30	0.32	0.33	0.32	0.26	0.26	NT
C5Q09	0.28	0.20	0.26	0.26	0.25	0.25	0.31	0.36	0.20	VIC
C5Q10	0.31	0.37	0.39	0.39	0.32	0.41	0.32	0.40	0.31	NSW
C5Q11	0.22	0.18	0.23	0.25	0.20	0.21	0.24	0.21	0.18	VIC
C5Q12	0.36	0.38	0.38	0.38	0.32	0.47	0.37	0.39	0.32	SA
C6Q01	0.33	0.26	0.30	0.36	0.29	0.35	0.35	0.34	0.26	VIC
C6Q02	Deleted from analysis due to poor fit									
C6Q03	0.32	0.29	0.28	0.29	0.23	0.37	0.27	0.33	0.23	SA
C6Q04	0.30	0.26	0.26	0.36	0.29	0.31	0.23	0.23	0.23	ACT
C6Q05	0.36	0.37	0.40	0.44	0.36	0.42	0.39	0.38	0.36	SA
C6Q06	0.33	0.36	0.35	0.37	0.31	0.40	0.33	0.28	0.28	NT
C6Q07	0.29	0.21	0.24	0.24	0.25	0.22	0.22	0.27	0.21	VIC
C6Q08	0.18	0.26	0.25	0.30	0.19	0.25	0.29	0.26	0.18	NSW
C6Q09	0.31	0.31	0.30	0.30	0.30	0.34	0.36	0.40	0.30	SA
C6Q10	0.16	0.18	0.27	0.32	0.17	0.25	0.25	0.41	0.16	NSW
C6Q11	0.21	0.25	0.34	0.34	0.30	0.33	0.30	0.33	0.21	NSW
C6Q12	0.29	0.25	0.36	0.36	0.33	0.38	0.35	0.48	0.25	VIC
C6Q13	0.37	0.36	0.37	0.37	0.29	0.35	0.39	0.42	0.29	SA
C7Q01	0.24	0.24	0.22	0.25	0.21	0.21	0.26	0.31	0.21	SA

Cluster code	Discrimination index								Minimum discrimination index	State with the minimum value
	NSW	VIC	QLD	WA	SA	TAS	ACT	NT		
C7Q02	0.20	0.22	0.30	0.23	0.25	0.31	0.28	0.34	0.20	NSW
C7Q03	0.29	0.31	0.39	0.36	0.36	0.31	0.38	0.47	0.29	NSW
C7Q04	0.41	0.41	0.40	0.44	0.42	0.39	0.43	0.44	0.39	TAS
C7Q05	0.34	0.36	0.36	0.38	0.34	0.40	0.36	0.40	0.34	NSW
C7Q06	0.29	0.32	0.28	0.32	0.29	0.32	0.29	0.43	0.28	QLD
C7Q07	0.25	0.21	0.23	0.27	0.22	0.22	0.26	0.33	0.21	VIC
C7Q08	0.36	0.39	0.38	0.47	0.33	0.32	0.39	0.41	0.32	TAS
C7Q09	0.35	0.29	0.39	0.40	0.32	0.34	0.41	0.46	0.29	VIC
C7Q10	0.24	0.25	0.30	0.28	0.22	0.30	0.26	0.30	0.22	SA
C7Q11	0.30	0.35	0.28	0.34	0.31	0.32	0.30	0.40	0.28	QLD
C7Q12	0.41	0.36	0.40	0.37	0.41	0.44	0.34	0.48	0.34	ACT
C7Q13	0.42	0.41	0.41	0.38	0.44	0.38	0.40	0.49	0.38	WA
CCQ01	0.17	0.29	0.20	0.27	0.21	0.28	0.37	0.23	0.17	NSW
CCQ02	0.27	0.33	0.24	0.26	0.27	0.31	0.30	0.36	0.24	QLD
CCQ03	0.22	0.21	0.23	0.25	0.20	0.26	0.18	0.20	0.18	ACT
CCQ04	0.29	0.27	0.28	0.26	0.27	0.30	0.23	0.32	0.23	ACT
CCQ05	0.36	0.35	0.43	0.39	0.38	0.41	0.38	0.31	0.31	NT
CCQ06	0.30	0.38	0.36	0.31	0.36	0.37	0.35	0.37	0.30	NSW
CCQ07	0.24	0.18	0.08	0.24	0.14	0.24	0.24	0.15	0.08	QLD
CCQ08	0.31	0.31	0.29	0.31	0.28	0.30	0.24	0.27	0.24	ACT
CCQ09	0.33	0.41	0.33	0.32	0.40	0.41	0.32	0.36	0.32	WA
CCQ10	0.34	0.30	0.32	0.37	0.27	0.34	0.30	0.25	0.25	NT
CDQ01	0.25	0.25	0.26	0.24	0.26	0.27	0.31	0.29	0.24	WA
CDQ02	0.25	0.26	0.29	0.23	0.28	0.23	0.27	0.31	0.23	WA
CDQ03	0.35	0.30	0.37	0.36	0.42	0.30	0.41	0.41	0.30	VIC
CDQ04	0.31	0.26	0.29	0.31	0.31	0.30	0.34	0.32	0.26	VIC
CDQ05	0.37	0.29	0.29	0.31	0.28	0.21	0.27	0.29	0.21	TAS
CDQ06	Deleted from analysis due to poor fit									
CDQ07	0.35	0.32	0.29	0.31	0.33	0.32	0.31	0.30	0.29	QLD
CDQ08	0.24	0.24	0.29	0.29	0.22	0.29	0.26	0.30	0.22	SA
CDQ09	0.34	0.23	0.23	0.19	0.24	0.26	0.36	0.21	0.19	WA
CDQ10	0.31	0.27	0.27	0.20	0.23	0.18	0.29	0.24	0.18	TAS
CDQ11	0.33	0.28	0.31	0.23	0.29	0.37	0.29	0.27	0.23	WA

7.1.6. Item difficulty by gender

Table 7.8 shows item parameters calibrated separately for gender groups, arranged in order of the difference between the item difficulty parameters. The left side of the table shows items where boys performed better, and the right side of the table shows items where girls performed better. For most items, the difference in item difficulty parameters is small.

If one takes 0.5 logits as a cut-off value for identifying a relatively large gender difference, then only seven items fall in this category. Boys performed better on items C2Q04, C5Q08, CDQ05 and CDQ11. Girls performed better on items C1Q06, C6Q09 and CCQ01.

These seven items were retained in the analysis as the estimation model included gender as a regression term and was thus able to absorb the observed gender DIF for these seven items. Item C6Q01showed the smallest difference (0.00 logits – no observed difference) in item difficulty between boys and girls.

Table 7.8 Items by difference in performance by gender

Girls performed better			
Cluster Code	**Girls**	**Boys**	**Diff**
C1Q01	-0.65	-1.12	-0.47
C1Q02	-0.22	-0.24	-0.02
C1Q05	-0.26	-0.32	-0.06
C1Q06	-0.47	-1.08	-0.61
C1Q07	-0.49	-0.86	-0.38
C1Q11	1.7	1.51	-0.19
C1Q13	-0.83	-1.15	-0.32
C1Q14	3.19	3.06	-0.14
C2Q01	-1.46	-1.53	-0.07
C2Q02	0.88	0.66	-0.22
C2Q07	-0.92	-0.96	-0.04
C2Q11	-1.69	-1.74	-0.05
C2Q12	-0.57	-0.77	-0.2
C3Q01	1.4	1.32	-0.08
C3Q02	-1.85	-2.11	-0.26
C3Q03	-1.16	-1.21	-0.04
C3Q05	-0.25	-0.56	-0.31
C3Q06	-2.13	-2.29	-0.16
C3Q08	0.14	0.12	-0.03
C3Q10	-0.57	-0.65	-0.08
C4Q01	0.62	0.57	-0.06
C4Q02	-1.77	-2.1	-0.33
C4Q04	0.09	-0.1	-0.19
C4Q06	3.55	3.36	-0.19
C4Q08	-1.58	-1.69	-0.11
C4Q09	2.14	1.99	-0.15
C4Q10	1.5	1.33	-0.17
C4Q12	1.3	1.14	-0.17
C5Q02	-0.96	-1.12	-0.15
C5Q03	3.02	2.96	-0.06
C5Q05	0.4	0.03	-0.36

Boys performed better			
Cluster Code	**Girls**	**Boys**	**Diff**
C1Q03	-1.25	-1.13	0.12
C1Q04	-0.54	-0.41	0.13
C1Q08	1.35	1.44	0.1
C1Q09	-0.55	-0.42	0.13
C1Q10	-0.98	-0.95	0.03
C1Q12	1.59	1.9	0.3
C2Q03	-0.98	-0.83	0.15
C2Q04	0.53	1.09	0.56
C2Q05	1.48	1.56	0.08
C2Q06	0.38	0.49	0.11
C2Q08	1.68	2.04	0.35
C2Q09	0.49	0.51	0.02
C2Q10	1.65	1.81	0.15
C3Q04	-0.99	-0.89	0.1
C3Q07	-0.56	-0.42	0.14
C3Q09	1.06	1.16	0.1
C3Q11	1.84	1.92	0.08
C3Q12	1.44	1.61	0.16
C4Q03	-0.11	-0.06	0.05
C4Q05	0.57	0.62	0.05
C4Q07	-0.76	-0.72	0.05
C4Q11	-0.26	-0.24	0.02
C5Q01	-3.82	-3.79	0.03
C5Q04	-2.2	-2.06	0.15
C5Q08	1.36	1.88	0.52
C5Q09	-0.76	-0.65	0.11
C5Q10	-0.1	0.12	0.22
C5Q12	0.24	0.27	0.03
C6Q03	1.48	1.77	0.28
C6Q04	0.91	0.98	0.07
C6Q05	-1.38	-1.37	0.02

Girls performed better			
Cluster Code	Girls	Boys	Diff
C5Q06	-0.39	-0.42	-0.03
C5Q07	0.77	0.49	-0.28
C5Q11	0.34	-0.14	-0.48
C6Q08	-2	-2.36	-0.37
C6Q09	0.2	-0.71	-0.92
C6Q11	-2.21	-2.26	-0.05
C7Q01	-1.89	-1.95	-0.07
C7Q02	0.68	0.66	-0.02
C7Q03	-2.34	-2.4	-0.06
C7Q06	1.49	1.48	-0.01
C7Q07	-0.26	-0.3	-0.04
C7Q10	-0.47	-0.65	-0.18
C7Q11	-0.23	-0.29	-0.06
C7Q12	0.29	0.2	-0.1
CCQ01	-2.63	-3.25	-0.62
CCQ02	-1.66	-1.89	-0.24
CCQ04	0.31	0.24	-0.07
CCQ05	-0.63	-0.86	-0.23
CCQ06	-1.18	-1.48	-0.3
CCQ08	1.62	1.34	-0.28
CCQ09	-0.04	-0.1	-0.06

Boys performed better			
Cluster Code	Girls	Boys	Diff
C6Q06	0.96	1.21	0.24
C6Q07	0.98	1.25	0.27
C6Q10	-2.24	-2.06	0.18
C6Q12	-1.47	-1.4	0.07
C6Q13	-1.44	-1.13	0.3
C7Q04	-1.62	-1.59	0.03
C7Q05	0.66	0.93	0.27
C7Q08	-0.21	-0.11	0.09
C7Q09	-1.82	-1.78	0.04
C7Q13	0.48	0.49	0.01
CCQ03	2.04	2.53	0.49
CCQ07	2.98	3.14	0.17
CCQ10	1.95	2.25	0.29
CDQ01	-3	-2.67	0.32
CDQ02	-1.33	-1.12	0.21
CDQ03	0.34	0.6	0.26
CDQ04	-0.62	-0.5	0.12
CDQ05	0.95	1.53	0.59
CDQ07	1.13	1.61	0.48
CDQ08	1.29	1.66	0.37
CDQ09	1.7	1.73	0.03
CDQ10	2.53	2.81	0.28
CDQ11	2.2	2.78	0.58

7.2. Test design

Students were required to complete an objective test and one of the two inquiry tasks. The objective tests were made up of item sets grouped into clusters. Each cluster appeared in three of the seven test forms – once at the beginning of the form, once in the middle and once at the end of the form. The following table shows how each cluster was arranged within the objective test forms.

Table 7.9 Structure of test forms by cluster

Objective test form	Block 1	Block 2	Block 3
1	Cluster 1	Cluster 2	Cluster 4
2	Cluster 2	Cluster 3	Cluster 5
3	Cluster 3	Cluster 4	Cluster 6
4	Cluster 4	Cluster 5	Cluster 7
5	Cluster 5	Cluster 6	Cluster 1
6	Cluster 6	Cluster 7	Cluster 2
7	Cluster 7	Cluster 1	Cluster 3

A complete running order of items is provided in Appendix 11.

7.3. Item analysis files

Access to the data files and output from the analyses can be made available to researchers or future contractors who want to replicate procedures on application for approval to the Australian Curriculum, Assessment and Reporting Authority (ACARA). Relevant data files are listed throughout this report.

CHAPTER 8 SCALING OF TEST DATA

8.1. Overview

The process of scaling refers to the estimation of student achievement distributions using information from students' responses to the test items. In the NAP sample assessments – science literacy (NAP—SL), the scaling process involved two separate phases. Firstly, the item parameters were calibrated using a sample of the data. These item parameters were used as the basis for equating the 2015 results to the 2006 scale. Secondly, student proficiency levels were then calculated based on the full dataset.

8.1.1. Calibration of item parameters

The calibration of item parameters used a calibration sample in which equal numbers of respondents from each jurisdiction were included. See section 8.2 on the selection of the calibration sample and the methodology for the calibration of item parameters.

8.1.2. Estimating student proficiency levels and producing plausible values

Once item parameters had been determined, student proficiency levels were estimated. As the main purpose of the study is to obtain profiles of student achievement at the population level, rather than at the individual student level, a methodology using plausible values (Wu, 2005) was adopted. The following sections describe in detail the two phases of the scaling process.

8.2. Calibration sample

8.2.1. Overview

To estimate item difficulty parameters, a subset of the responses, called the calibration sample, was used to ensure that each jurisdiction had an equal representation in the sample so that the larger states did not unduly influence the item parameter values. Since the Northern Territory had the smallest number of responses, all 649 responses were included in the calibration sample. For each of the other jurisdictions, a random sample of 649 responses was selected. Consequently, the calibration sample consisted of 5192 (=649×8) responses.

8.2.2. Data files availability

Access to the data files and output from the analyses is available under specific circumstances on application to the Australian Curriculum, Assessment and Reporting Authority (ACARA) at datarequest@acara.edu.au.

8.2.2.1. CalibrationSample.sav

The file CalibrationSample.sav contains student background variables as well as item responses.

The variables Q1 to Q107 are students' raw item responses, recoded with A, B, 9 and M. The following rules apply to the recoding:

- For the objective test, the first 'not reached' item is coded as 'A' with the remaining 'not reached' items as 'B', and embedded missing responses remain as '9'. Students with no responses at all for the whole test have responses recoded to 'M'.

- For the inquiry task, students with no responses at all have responses recoded to ‘M’. Missing responses, whether not-reached or embedded, are recoded to ‘9’. That is, there are no ‘A’ and ‘B’ codes. As the two inquiry tasks have only 10 and 11 items respectively, there does not appear to be a large number of clearly ‘not reached’ items at the end.
- To calibrate the item parameters, response codes ‘A’ and ‘9’ are treated as incorrect, whereas response codes ‘B’ and ‘M’ are treated as non-administered (that is, as missing data).
- In contrast, to calibrate the student abilities in subsequent analyses, response code ‘M’ is treated as not-administered, but response codes ‘A’, ‘9’ and ‘B’ are treated as incorrect.

The recoding process follows the same approach as used in 2012 so as to ensure consistency in the way the data was analysed. However, in future cycles additional data from the online test delivery platform may provide clearer indications of which items were not encountered by the student, thus enabling a clearer distinction between embedded missing and ‘not reached’ items.

8.2.2.2. CalibrationItems.dat

This ASCII (or text) file is used as input to IRT software to calibrate the item parameters. The codebook for the relevant data fields in the text file is given below:

Table 8.1 Codebook for CalibrationItems.dat

Field	Column range	Description
Form	123	Objective test form id (1 to 7)
Item responses	1 to 107	Student responses to items

8.2.3. IRT analysis for calibrating item parameters

The software program used to carry out the calibration of item parameters is ConQuest. A facets model is used where the test form number is regarded as a facet. More specifically, the model statement used in ConQuest is:

bookid + item + item*step

The full syntax of ConQuest commands is in the control file *Calibration.cqc*.

The use of the term ‘bookid’ in ConQuest model statements is to ensure that the estimation of the item parameters takes into account the so-called ‘booklet effect’ (OECD 2012, p. 141). However, as there is only one domain in 2015 NAP—SL (unlike PISA where there are three domains: mathematics, science and reading) and all items are calibrated together, it is not expected that there will be a significant booklet effect, as is shown later in the results of the item analysis.

Three output files are produced from ConQuest:

Calibration.shw

This is a summary file, showing form (booklet) and item parameter values, population parameter estimated and item-person maps.

Calibration.itn

This file is known as the 'itanal', showing classical test statistics as well as IRT statistics for each item.

Calibration.anc

This file is produced through an Export statement in ConQuest. It contains the values of the parameters that can be used as anchor values later when student abilities are estimated.

Once the calibrated item parameters are obtained, the transformation equations used to equate the 2015 results to the 2006 scale are then derived. Details of the equating process can be found in Chapter 9 of this report.

8.3. Estimating student proficiency levels and producing plausible values

In this phase, student proficiency levels are estimated for the full data set (NAPSL 2015_PV_2016-06-02.sav. See Appendix 6 for descriptions of variables).

The scaling model used is a one-parameter item response model with conditioning variables in the population latent regression model. See PISA 2009 Technical Report for a description of the model (OECD 2005).

The conditioning variables included are:

- school mean proficiency (average of students' weighted likelihood estimates for each school)
- state or territory
- sector
- gender
- Indigenous status
- geographic location
- language background.

To prepare the data to be used as conditioning variables, two separate steps are taken:

Step A: Produce a weighted likelihood estimate (WLE) for each student in the full data set, and compute the average WLE for each school. ConQuest was used for the estimation of WLE estimates, with item parameters anchored at values from the item calibration phase.

Step B: Dummy variables are created for state or territory, sector, gender, Indigenous status, geographic location and language background.

8.3.1. Production of plausible values

The software program ConQuest is used for the scaling of student proficiency levels and the generation of plausible values. Note that Case Weight is used in this analysis. Both form (booklet) parameters and item parameters are anchored. Both embedded-missing (code ‘9’) and not-reached items (codes ‘A’ and ‘B’) are treated as incorrect. If a test has no valid responses from a student, the responses (code ‘M’) are treated as not-administered. Ten plausible values are generated (instead of the usual five).

The ConQuest control file used is *NAPSL _pv.cqc*, which is shown in Appendix 7.

8.4. Estimation of statistics of interest and their standard errors

Once the plausible values are produced for each student, statistics of interest can be computed together with their standard errors. For example, the mean achievement level in science literacy for Year 6 students in Australia can be estimated, as well as jurisdiction average achievement levels. The estimates will also have associated standard errors to indicate the confidence which we have about the results.

The plausible-values methodology has been used for large-scale studies such as TIMSS, PISA and NAEP. In 2015 NAP—SL, this methodology was also used for the estimation of statistics and standard errors. For a detailed description of the methodology, see Mislevy, Beaton, Kaplan and Sheehan (1992), and Beaton and Gonzalez (1995).

Briefly, the methodology is summarised below. The plausible values for each student show the indicative level of the student’s achievement. So the estimate for a population statistic is computed using the plausible values as if they represent each student’s level of achievement. For example, to compute the estimated mean of the population, take the first plausible value for each student and compute the average across students, weighted by the sampling weight (student final weight). Repeat the process with all ten plausible values, and then average the ten estimated means for the ten runs. Similarly, for the estimation of percentiles and percentages in levels, plausible values are used in the same way.

The standard errors associated with the estimated statistics are not straightforward to compute, as the sampling method is not simple random sampling but a complex two-stage sampling. Typically, for complex sampling such as the one used for 2015 NAP—SL, replication methods such as Balanced Repeated Replicate (BRR) or Jackknife are used to compute standard errors (Rust and Rao 1996). In 2015 NAP—SL, the Jackknife method was used. Jackknife replication weights are computed (variables RW1 to RW309 in the file NAPSL 2015_PV_2016-06-02.sav).

The statistic of interest is computed using each of the replicate weights in turn. The variations in the estimated statistic obtained from using different replicate weights contribute to the estimate of the sampling variance for the estimated statistic. Combining this sampling variance with the variance from using the ten plausible values (measurement error) provides an estimate of the standard error for the estimated statistic. SPSS macros and R scripts were written to carry out the procedures of the estimation of statistics and their standard errors.

8.5. Transform logits to a scale with mean 400 and standard deviation 100

To facilitate the interpretation of the results, it is a common practice to transform logit scores. It was decided that, for NAP—SL assessments, the proficiency scale should have a national mean of 400 and a standard deviation of 100. This scale was chosen to avoid having negative values on the scale representing student proficiency. Further, a standard deviation of 100 provides easy interpretation of proficiency levels in terms of how far away a score is from the mean.

As part of the equating process (refer to Chapter 9 for details), the 2015 logit scores are first translated to the 2006 scale, then transformed to the 400/100 scale. The transformation used in 2015 is given below:

Score on proficiency scale = (Logit–0.200543797)/0.954513216*100+400

Where 'Logit' refers to a logit score on the 2006 scale. The conversion of 2015 logit scores to the 2006 scale is detailed in section 9.2.

Note that the mean of 400 is the national mean, computed using student sampling weights to reflect the average achievement of all Year 6 students in Australia. It is not the average of jurisdiction means, as that average does not take into account the number of students in each jurisdiction. In summary, house weights are used to set the average score of 400, not senate weights.

CHAPTER 9 EQUATING 2015 RESULTS TO 2006 RESULTS

9.1. Setting 2006 results as the baseline

While the first cycle of the NAP sample assessments – science literacy (NAP—SL) was conducted in 2003 (then known as PSAP), and the 2006 assessment was the second round of NAP—SL, it was decided that the 2006 assessment be used to set the scale of a mean of 400 and a standard deviation of 100 instead of the 2003 assessment. The reasons for this decision are summarised below:

(1) The 2006 assessment test design was more robust than the 2003 test design. In 2006, a balanced incomplete block (BIB) test design consisting of seven test booklets was used. In contrast, in 2003 only two test booklets were used, resulting in item-position effect for most items.

(2) There were considerably more items in 2006 than in 2003, resulting in a better coverage of the science literacy content in 2006. In 2006, 110 items were included in the final test, while only 72 items were included in the 2003 test.

(3) The 2006 assessment produced a much higher population variance in achievement than 2003 did. In logits, the 2006 population standard deviation was 0.95, while the 2003 population standard deviation was 0.78. This could be an indication that:

- the 2006 items were generally more discriminating than the 2003 items; that is, the 2006 items were higher quality items
- the 2006 sampling was more comprehensive, as remote schools were also included in the sample, while the 2003 sampling focused only on areas where students were generally well-resourced.

9.2. Equating 2015 results to 2006 results

As a consequence of the decision to use 2006 results as the baseline, 2015 results were equated to 2006 results. To carry out the equating, link items between the 2015, 2012, 2009 and 2006 tests were used.

9.2.1. Link items

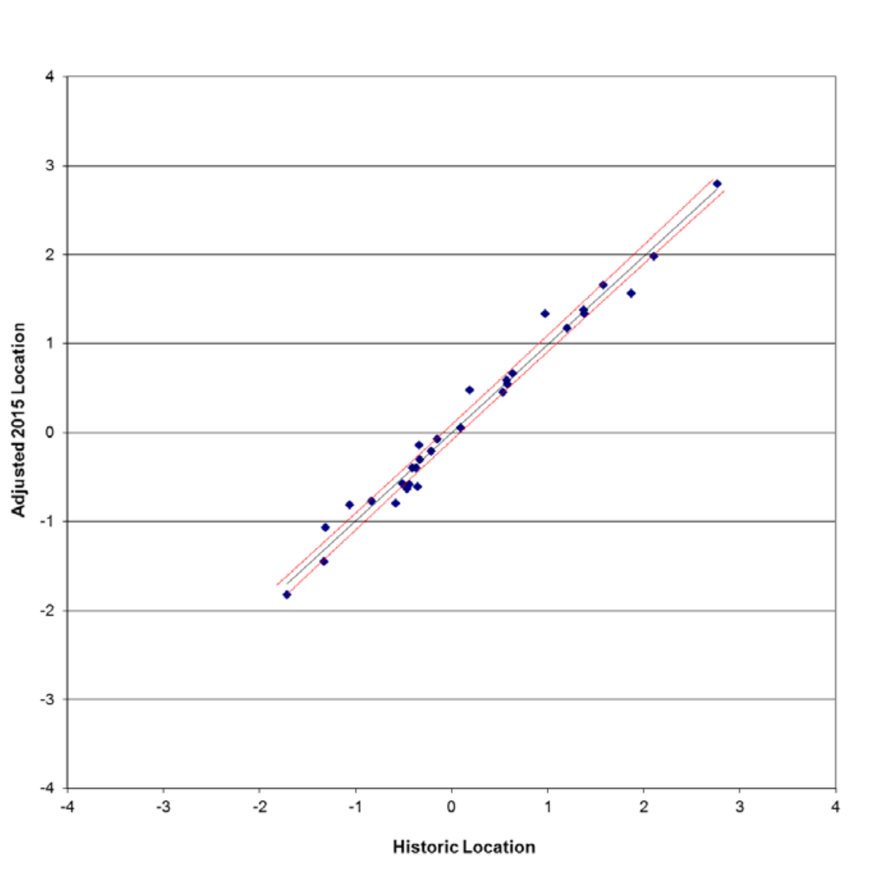
In order to equate the 2015 results to the science literacy scale, a total of 38 link items were included in the 2015 assessment. This included 12 link items from the 2006 assessment, 7 link items from the 2009 assessment and 19 items from the 2012 assessment. Care was taken to find items that performed well psychometrically and also covered the range of science literacy strands A, B and C and the concept areas.

9.2.2. Link item selection

The selection process for the final set of link items to conduct equating between 2015 and 2006 consisted of two parts. In the first part, the list of items was refined based on the comparisons of item locations in 2015 and the location of the items when located on the historic (2006) scale. In the second part, the final set of items was inspected by a content expert from the test construction team. The purpose of the latter was to ensure that the selected link items provided adequate coverage across the science literacy strands.

In the first part, the 2015 location of link items was independently estimated. In order to conduct comparisons of item locations between 2006 and 2015, the 2015 locations were adjusted to have the same mean and standard deviation as observed in 2006. Items that were clear outliers were removed with a final set of 30 link items retained. A plot of historic and the adjusted 2015 item difficulties for the final link items, including graphical representation of 95 per cent confidence interval for the statistical difference between item locations, is given in Figure 9.1.

Figure 9.1 Calibrated item difficulties in 2006 and 2015 for the final link item set



9.2.3. Equating procedures

The 2015 data were scaled and item parameters were obtained. The mean difference between the final set of link item parameters for 2015 compared to historic estimates was used to obtain the 'shift' required to place 2015 results onto the historic scale. This approach to equating is a simple, more common approach to equating than used for NAP—SL equating in the past. The methodology used for equating NAP—SL to historical cycles in previous cycles provides similar results to the 'shift' method and hence, the simplest equating model has been retained. In addition, the adjustment for students' interaction with online historic items was included in the transformation formula (see Chapter 3).

9.3. Equating transformation

The result of the equating process was the derivation of a transformation formula for the 2015 results to be placed on the 2006 scale:

2015 on 2006 scale = 2015logit + 0.025 + 0.334

For standard errors, no transformation is required, since the equating transformation does not involve a scaling factor:

2015 standard error on 2006 scale in logit = 2015 S.E. in logit

9.4. Link error

In establishing trends from 2006 to 2015, it is necessary to make judgments about the statistical significance of the difference in science literacy achievement between 2015 and 2006. An appropriate estimation of the magnitude of equating errors is important when trends are reported. An underestimate of the equating errors will often result in erroneous claims of change in achievement levels when there is no significant difference.

Equating errors come from at least two sources: the sampling of students and the sampling of items. Equating errors due to the sampling of students affect the accuracy with which the item parameters are estimated, and the magnitude of these errors diminishes when the

sample size increases. However, equating errors due to the sampling of items have not often been taken into account, and the magnitude of these errors does not diminish when the sample size increases. For the estimates of population parameters (for example, mean), the magnitude of equating errors due to the sampling of items tends to be much larger than the magnitude of equating errors due to the sampling of students. Consequently, it is important to estimate the equating error due to the sampling of items.

Equating error (called 'link error' in PISA) is computed following the approach used in PISA 2009 (OECD 2012). Firstly, calibrate the items using 2015, 2012, 2009 and 2006 data separately. If the link items behave exactly the same way in 2015, 2012, 2009 and 2006 (and they follow the Rasch model), there should only be a constant difference between 2015, 2012, 2009 and 2006 item parameters for matched items. However, in real life, items will vary from 2015 to 2006 and some items will vary more than others.

The link error for comparison between 2012 and 2015 is 0.048 logits; transformed to the science literacy scale it is equal to a scaled score of 5.03. Similarly, the link error for comparison between 2009 and 2015 is 0.048 logits; transformed to the science literacy scale it is equal to a scaled score of 5.01. Finally, the link error for comparison between 2006 and 2015 is 0.077 logits; transformed to the science literacy scale it is equal to a scaled score of 8.02.

Additional information about the computation of link errors can be found in data available to researchers or future contractors on application for approval to the Australian Curriculum, Assessment and Reporting Authority (ACARA).

The link error is used only when comparisons across 2015, 2012, 2009 and 2006 results are made. For example, to test whether the mean achievement in 2015 differs from the mean achievement in 2006, the link error is added to the standard error of the difference, as illustrated in Table 9.1.

Table 9.1 Example of link error application in calculating standard error of difference

	2015 mean on 2006 scale & S.E.	2006 mean & S.E.	2015 mean – 2006 mean	Standard error of difference	Standardised difference
Tas.	414 (5.97)	406 (6.17)	8	SQRT($5.97^2+6.17^2+8.02^2$)	0.68 = 8/11.75 (not significant)

CHAPTER 10 SCALE AND PROFICIENCY LEVELS

For reporting purposes, student results are often summarised through the definition of a number of proficiency levels. That is, the proficiency scale is divided into a number of levels, with descriptions of skills attached to each level, and percentages of students at various levels are reported.

10.1. Proficiency level cut-points

In 2006, cut-points along the proficiency scale were decided after consultations with science experts. In 2015, the same cut-points were used, as shown in Table 10.1.

Table 10.1 Cut-points for 2015 NAP—SL

Level	2006 cut-points	Transformed to 400/100 scale
2 and below	<–1.114	262.293
3.1	0.130	392.577
3.2	1.373	522.861
3.3	2.617	653.145
4	>2.617	>653.145

As for previous cycles, a response probability (RP) of 0.65 is used to place items in proficiency levels. The RP adjustment refers to '…the probability that a student in the middle of a level would correctly answer an item of average difficulty for that level.' (OECD 2000, p. 198).

10.2. Proficiency levels of items

Table 10.2 shows the 2015 NAP sample assessments – science literacy (NAP—SL) items and their corresponding levels on the proficiency scale.

Table 10.2 Proficiency levels of items

Item	Item ID	Historical item label	2015 difficulty	2015 difficulty on historic scale after adjusting for RP	Operational level	Design level	Scaled score
1	NSL06H149.49	IDOB149	–0.875	0.106	3.1	3	390
2	NSL06H149.50	IDOB150	–0.196	0.785	3.2	3	461
3	NSL09H308.08	IDOB308	–1.200	–0.219	3.1	3	356
4	NSL09H308.09	IDOB309	–0.475	0.506	3.2	3	432
5	NSL09H308.10	IDOB310	–0.368	0.613	3.2	3	443
6	NSL06H084.84	IDOB084	–0.728	0.253	3.2	3	405
7	NSL06H084.85	IDOB085	–0.704	0.277	3.2	3	408
8	NSL06H084.86	IDOB086	1.510	2.491	3.3	3	640
9	NSL06H084.87	IDOB087	–0.475	0.506	3.2	3	432
10	NSL06H084.88	IDOB088	–0.902	0.079	3.1	3	387
11	NSL09H405.05	IDOB405	1.465	2.446	3.3	3	635

Item	Item ID	Historical item label	2015 difficulty	2015 difficulty on historic scale after adjusting for RP	Operational level	Design level	Scaled score
12	NSL09H405.06	IDOB406	1.667	2.648	4 and above	4	656
13	NSL12H503.03	IDOB503	–0.971	0.010	3.1	3	380
14	NSL12H503.06	IDOB506	3.005	3.986	4 and above	4	797
15	NSL09H360.60	IDOB360	–1.535	–0.554	3.1	3	321
16	NSL09H360.62	IDOB363	0.687	1.668	3.3	3	554
17	NSL12H559.59	IDOB559	–0.925	0.056	3.1	3	385
18	NSL12H559.61	IDOB561	0.746	1.727	3.3	3	560
19	NSL12H559.62	IDOB562	1.415	2.396	3.3	3	630
20	NSL12H564.64C	IDOB564	0.453	1.434	3.3	3	529
21	NSL12H564.65	IDOB565	–0.909	0.072	3.1	3	387
22	NSL12H564.66C	IDOB566	1.769	2.75	4 and above	3	667
23	NSL12H564.69	IDOB569	0.485	1.466	3.3	3	533
24	NSL12H564.70	IDOB570	1.734	2.715	4 and above	4	663
25	NSL06H041.41	IDOB041	–1.700	–0.719	3.1	3	304
26	NSL06H041.44	IDOB044	–0.677	0.304	3.2	3	411
27	NSL15E_C006.2		1.246	2.227	3.3	3	612
28	NSL15E_C006.1		–2.028	–1.047	3.1	3	269
29	NSL15E_W002.1		–1.179	–0.198	3.1	3	358
30	NSL15E_W002.2		–0.983	–0.002	3.1	3	379
31	NSL15E_W002.3		–0.399	0.582	3.2	2	440
32	NSL15E_W002.4		–2.160	–1.179	2 and below	2	255
33	NSL15E_W002.5		–0.506	0.475	3.2	2	429
34	NSL15E_Z002.2C		0.148	1.129	3.2	3	497
35	NSL15E_Z002.4		1.046	2.027	3.3	3	591
36	NSL15E_Z007.2		–0.596	0.385	3.2	2	419
37	NSL15E_Z007.3		1.842	2.823	4 and above	4	675
38	NSL15E_Z007.4		1.478	2.459	3.3	3	637
39	NSL12H517.17	IDOB517	0.599	1.58	3.3	3	545
40	NSL12H517.18	IDOB518	–2.024	–1.043	3.1	3	270
41	NSL12H517.19	IDOB519	–0.123	0.858	3.2	3	469
42	NSL12H517.21	IDOB521	0.015	0.996	3.2	3	483
43	NSL12H517.22	IDOB522	0.553	1.534	3.3	3	540
44	NSL12H551.51C	IDOB551	3.514	4.495	4 and above	4	850
45	NSL12H551.52	IDOB552	–0.664	0.317	3.2	3	412
46	NSL12H551.53	IDOB553	–1.617	–0.636	3.1	3	312
47	NSL12H551.54	IDOB554	2.114	3.095	4 and above	4	703
48	NSL06H021.21	IDOB021	1.418	2.399	3.3	3	630
49	NSL06H021.22	IDOB022	–0.264	0.717	3.2	3	454
50	NSL06H021.23	IDOB023	1.238	2.219	3.3	3	611

Item	Item ID	Historical item label	2015 difficulty	2015 difficulty on historic scale after adjusting for RP	Operational level	Design level	Scaled score
51	NSL15E_F001.1		–3.604	–2.623	2 and below	2	104
52	NSL15E_F001.2		–1.142	–0.161	3.1	3	362
53	NSL15E_F001.4C		2.916	3.897	4 and above	3	787
54	NSL15E_M003.1		–2.099	–1.118	2 and below	2	262
55	NSL15E_M003.2		0.183	1.164	3.2	3	501
56	NSL15E_A006.1		–0.460	0.521	3.2	3	434
57	NSL15E_A006.3		0.557	1.538	3.3	3	540
58	NSL15E_W006.1		1.503	2.484	3.3	3	639
59	NSL15E_W006.2		–0.689	0.292	3.2	3	410
60	NSL15E_W006.3		0.014	0.995	3.2	3	483
61	NSL15E_F004.2		0.054	1.035	3.2	2	487
62	NSL15E_F004.3		0.350	1.331	3.2	3	518
63	NSL15E_H009.1		–1.191	–0.210	3.1	3	357
64	NSL15E_E001.1		1.524	2.505	3.3	3	641
65	NSL15E_E001.4		1.104	2.085	3.3	4	597
66	NSL15E_H004.1		–1.341	–0.360	3.1	2	341
67	NSL15E_H004.2		1.070	2.051	3.3	3	594
68	NSL15E_H004.3		1.069	2.050	3.3	3	594
69	NSL15E_H002.1		–2.125	–1.144	2 and below	3	259
70	NSL15E_H002.2		–0.318	0.663	3.2	3	448
71	NSL15E_H002.3		–1.993	–1.012	3.1	3	273
72	NSL15E_A007.1		–2.142	–1.161	2 and below	2	257
73	NSL15E_A007.2		–1.361	–0.380	3.1	3	339
74	NSL15E_A007.8		–1.237	–0.256	3.1	3	352
75	NSL15E_H007.1C		–1.883	–0.902	3.1	2	284
76	NSL15E_H007.3C		0.618	1.599	3.3	3	547
77	NSL15E_V001.1		–2.318	–1.337	2 and below	2	239
78	NSL15E_V001.2		–1.610	–0.629	3.1	2	313
79	NSL15E_H008.4		0.736	1.717	3.3	3	559
80	NSL15E_Z027.1		1.441	2.422	3.3	3	633
81	NSL15E_Z027.2		–0.225	0.756	3.2	3	458
82	NSL15E_M004.1		–0.233	0.748	3.2	3	457
83	NSL15E_M004.3		–1.799	–0.818	3.1	2	293
84	NSL15E_M004.4		–0.575	0.406	3.2	3	422
85	NSL15E_M004.5		–0.270	0.711	3.2	3	453
86	NSL15E_F003.3		0.203	1.184	3.2	3	503
87	NSL15E_F003.4		0.388	1.369	3.2	3	522
88	NSL15Esi3-1.1		–2.757	–1.776	2 and below	1	193
89	NSL15Esi3-1.2		–1.672	–0.691	3.1	1	307

Item	Item ID	Historical item label	2015 difficulty	2015 difficulty on historic scale after adjusting for RP	Operational level	Design level	Scaled score
90	NSL15Esi3-1.4		2.193	3.174	4 and above	2	712
91	NSL15Esi3-2.1		0.180	1.161	3.2	2	501
92	NSL15Esi3-3.1		–0.681	0.300	3.2	3	410
93	NSL15Esi3-4.1		–1.189	–0.208	3.1	3	357
94	NSL15Esi3-5.1		2.975	3.956	4 and above	4	793
95	NSL15Esi3-5.2		1.536	2.517	3.3	4	643
96	NSL15Esi3-5.3		0.004	0.985	3.2	4	482
97	NSL15Esi3-5.5		2.072	3.053	4 and above	3	699
98	NSL15Esi5-1.1		–2.560	–1.579	2 and below	2	214
99	NSL15Esi5-1.2		–1.071	–0.090	3.1	2	370
100	NSL15Esi5-2.1		0.505	1.486	3.3	3	535
101	NSL15Esi5-3.1		–0.558	0.423	3.2	3	423
102	NSL15Esi5-4.1		1.132	2.113	3.3	2	600
103	NSL15Esi5-4.3		1.346	2.327	3.3	3	623
104	NSL15Esi5-4.4		1.543	2.524	3.3	3	643
105	NSL15Esi5-5.1C		1.736	2.717	4 and above	3	664
106	NSL15Esi5-5.2C		2.787	3.768	4 and above	3	774
107	NSL15Esi5-5.3		2.402	3.383	4 and above	4	733

CHAPTER 11 SCALING THE STUDENT SURVEY

11.1. Background

The NAP sample assessments – science literacy (NAP—SL) survey items were written to investigate the aspects of student perception of and attitudes toward science as shown in Table 11.1.

Table 11.1 Aspects of student perception of and attitudes toward science investigated in the survey

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

Basic results from the survey are given in Appendix 10.

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

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

11.2. Scale construction

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

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

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

Figure 11.1 Item person map for 2015 survey analysis

```
------------------------------------------------------------------------------------------------------
Item Estimates (Category Deltas)
all on all (N = 11515 L = 43 Probability Level=0.50)
------------------------------------------------------------------------------------------------------
                                  |
                     Students     |      Item thresholds (deltas)
                                  |
  3.0                       X     |
                            X     |      16.3   20.3   22.3
                            X     |       5.3    6.3   12.3   21.3
                          XXX     |
  2.0                   XXXXX     |       2.3    7.3    8.3   18.3   24.3   38.3
                       XXXXXX     |      17.3   19.3   23.3   30.2
                  XXXXXXXXXXX     |       1.3    3.3   13.3   15.3   17.2   18.2   25.3   33     35.3
                  XXXXXXXXXXX     |       9.3   11.3   19.2   21.2   22.2   34     36.3
            XXXXXXXXXXXXXXXXX     |       4.2   16.2   39.2
  1.0    XXXXXXXXXXXXXXXXXXXX     |       8.2   14.2   20.2   25.2   42.3
         XXXXXXXXXXXXXXXXXXXX     |      24.2   37.2   43.3
              XXXXXXXXXXXXXXX     |      10.3   18.1   19.1   40.2
           XXXXXXXXXXXXXXXXXX     |      12.2
  0.0         XXXXXXXXXXXXXXX     |       5.2   17.1   38.2   41.2
                    XXXXXXXXX     |       2.2   21.1   23.2   28     29
                      XXXXXXX     |       6.2    7.2   27     31
                          XXX     |      13.2   15.2   20.1   26
                           XX     |       8.1    9.2   11.2   22.1   24.1
 -1.0                      XX     |       2.1   25.1   30.1   36.2   42.2   43.2
                            X     |       1.2    3.2   16.1   32     43.1
                                  |       3.1   10.2   35.2   38.1
                                  |       1.1   11.1   12.1   15.1   23.1
                                  |       5.1    6.1    7.1   10.1   35.1
 -2.0                             |       9.1   36.1   42.1
                                  |      13.1
                                  |
                                  |
 -3.0                             |       4.1   39.1
                                  |      37.1   40.1
                                  |      41.1
                                  |
                                  |
 -4.0                             |
                                  |      14.1
                                  |
                                  |
 -5.0                             |
------------------------------------------------------------------------------------------------------------
  Each X represents   68 students
============================================================================================================
```

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

Figure 11.2 shows frequencies of students on this scale compared with groups of items.

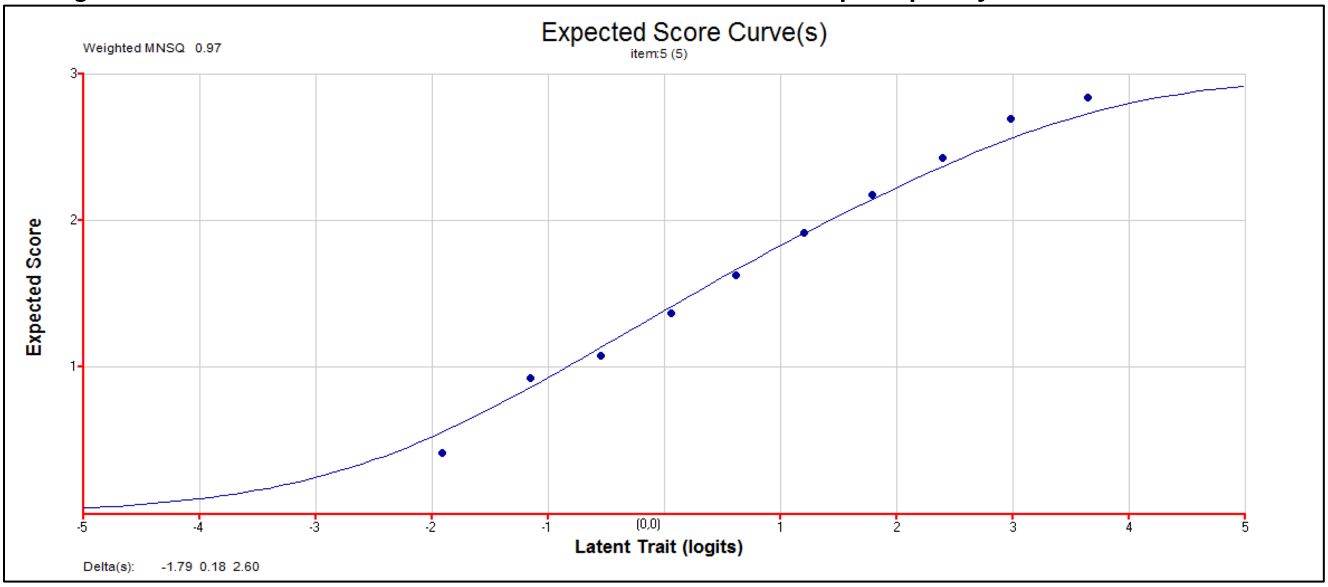
Figure 11.2 Item person map for 2015 survey analysis with dichotomised items

```
-------------------------------------------------------------------------------------------------------------
Item Estimates (Category Deltas)
all on all (N = 11515 L = 43 Probability Level=0.50)
-------------------------------------------------------------------------------------------------------------
  5.0                             |
                                  |
                                  |
                              X   |
                                  |
  4.0                             |
                                  |
                             XX   |
                                  |
                                  |
                            XXX   |
  3.0                             |
                           XXXX   |
                           XXXX   |
                                  |
                         XXXXXX   |  18
  2.0                  XXXXXXXX   |  33  17  19
                       XXXXXXXX   |  21
                      XXXXXXXXX   |  34  22
                     XXXXXXXXXX   |  16  20
                     XXXXXXXXXX   |
           XXXXXXXXXXXXXXXXXXXX   |   8  25
  1.0                XXXXXXXXXX   |  24
                     XXXXXXXXXX   |  12
                XXXXXXXXXXXXXXX   |
                        XXXXXXX   |
                         XXXXXX   |  38  2  5  28  29  30
  0.0                 XXXXXXXXX   |  27  23
                            XXX   |  31  6  7  26
                           XXXX   |
                             XX   |  13  15
                             XX   |
                              X   |  36  9  11  32
 -1.0                         X   |  42  43  3
                              X   |
                              X   |  35  39
                                  |   4  10
                                  |
 -2.0                             |  40  37
                                  |
                                  |  14
                                  |  41
                                  |
                                  |
 -3.0                             |
                                  |
                                  |
                                  |
                                  |
                                  |
 -4.0                             |
-------------------------------------------------------------------------------------------------------------
  Each X represents   72 students
=============================================================================================================
```

For the example of the two items mentioned above, it can be seen that student agreement is higher for the “I enjoy doing science” item (Item 3) than for the “I learn science topics quickly” item (Item 5).

The range of the student and item distributions on the scale is satisfactory (4 logits for the student distribution and 7 logits for the item threshold distribution). In addition, the distribution of items matches satisfactorily the distribution of students (at the 0.5 probability level) in the sense that the proportion of students who agree on each item is not too high or too low. The reliability of the survey was 0.91 and each item fitted the measurement model satisfactorily. Figure 11.3 shows the item characteristic curve (ICC) for the “I learn science topics quickly” item (Item 5).

Figure 11.3 Item characteristic curve for the “I learn science topics quickly” item



This ICC shows the mean agreement score (0 to 3) for 10 groups of students starting from low levels of agreement on the scale and ending with high levels of agreement on the scale. For the scores on this item to fit the measurement model, it is expected that the mean agreement score increases as the total survey scale score of a group increases according to the model probability curve. The dots show the observed mean scores on this item. The fit of this item to the measurement model is more than satisfactory. The ICCs for most survey items are similar to this ICC, supporting the conclusion of an underlying construct to the survey items.

It was found that an underlying construct has been measured with the 2015 NAP—SL survey items by fitting the data to the Rasch model. The data collected from the students in the NAP—SL sample showed that a measureable hierarchy of science perception items does exist. A similar analysis was performed on the survey data collected in 2012 confirming the existence of a measurable construct with the NAP—SL survey items.

The following table shows item parameters and fit statistics for the items in the survey for 2015 and 2012.

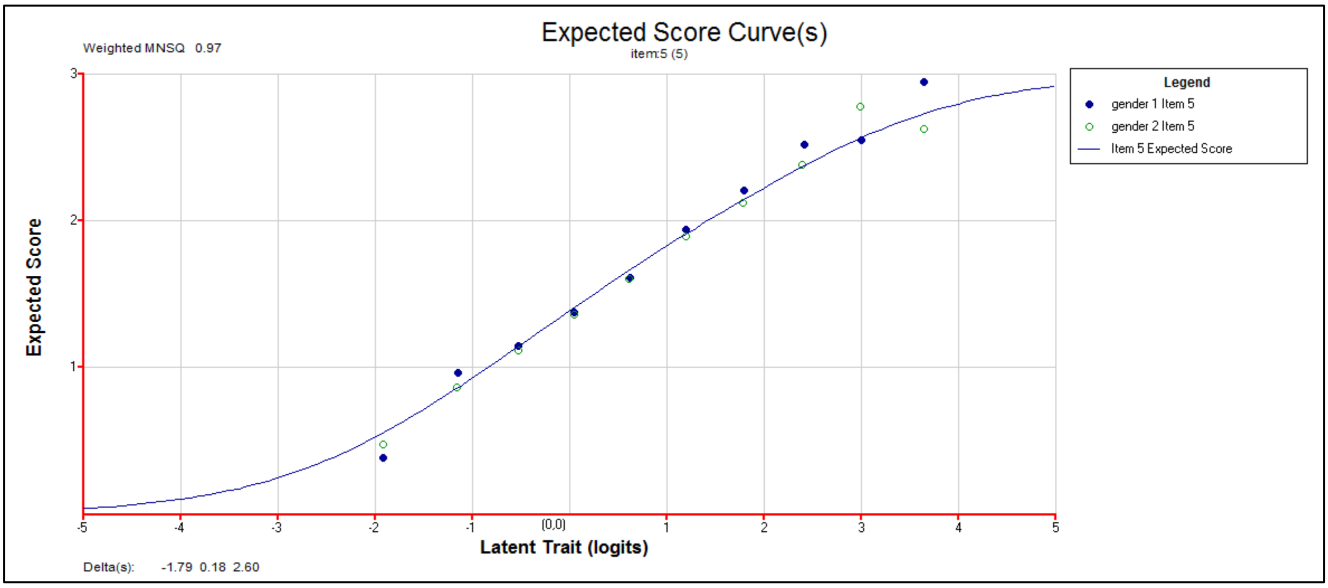
Table 11.2 Survey item statistics for 2015 and 2012

Group	Item	2015					2012				
		01	12	23	Infit MNSQ	Discrim	01	12	23	Infit MNSQ	Discrim
G01	1	-1.61	-1.14	1.66	0.89	0.54	-1.7	-1.05	1.51	0.88	0.56
	2	-1.01	-0.17	2.02	0.94	0.54	-0.86	-0.11	1.88	0.93	0.55
	3	-1.4	-1.16	1.51	0.87	0.57	-1.57	-1.35	1.26	0.84	0.6
	4	-1.23	-1.72	0.98	0.86	0.57	-1.73	-1.65	0.94	0.81	0.62
G02	5	-1.85	-0.01	2.4	0.94	0.51	-1.82	0	2.23	0.93	0.53
	6	-1.84	-0.37	2.51	0.93	0.51	-1.89	-0.35	2.38	0.92	0.52
	7	-1.96	-0.34	2.05	0.9	0.54	-1.95	-0.32	1.93	0.88	0.57

G03	8	-0.88	0.92	1.9	0.95	0.55	-0.75	1	1.57	0.88	0.6
	9	-2.01	-0.88	1.32	0.94	0.51	-2.11	-1	1.21	0.94	0.52
	10	-1.76	-1.42	0.57	0.93	0.52	-1.8	-1.34	0.58	0.95	0.51
	11	-1.75	-0.9	1.27	0.95	0.51	-1.95	-0.89	1.17	0.95	0.52
G04	12	-1.68	0.37	2.49	1.16	0.33	-1.81	0.07	2.13	1.12	0.37
	13	-2.23	-0.49	1.5	1.24	0.26	-2.43	-1.08	0.96	1.26	0.24
	14	-2.19	-2.16	0.9	0.94	0.47	-2.18	-2.39	0.7	0.97	0.45
	15	-1.64	-0.68	1.58	1	0.47	-1.96	-0.68	1.41	1.02	0.46
	16	-1.31	1.03	2.68	1.15	0.34	-1.42	0.52	2.2	1.11	0.39
G05	17	0.09	1.56	1.84	1	0.52	-0.17	1.33	1.61	0.95	0.55
	18	0.52	1.57	2.08	0.97	0.54	0.34	1.43	1.9	0.9	0.57
	19	0.53	1.35	1.82	0.93	0.58	0.48	1.39	1.75	0.87	0.59
G06	20	-0.65	1	2.55	1.09	0.43	-0.79	1.15	2.43	1.06	0.43
	21	-0.1	1.32	2.53	1.08	0.45	-0.53	1.22	2.16	1.04	0.47
G07	22	-0.91	1.25	2.55	1.02	0.47	-0.71	1.27	2.46	1	0.48
	23	-1.72	-0.15	1.82	1.05	0.43	-1.81	-0.16	1.7	1.07	0.43
	24	-0.88	0.69	1.92	1.16	0.4	-0.81	0.7	1.72	1.14	0.42
	25	-1.04	0.83	1.64	1.04	0.49	-1.1	0.92	1.53	1.03	0.49
G08	26	-0.54			0.99	0.31	-0.85			1.02	0.27
	27	-0.28			1.02	0.29	-0.84			1.01	0.29
	28	-0.17			1.02	0.29	-0.63			1.04	0.26
	29	-0.16			1.03	0.27	-0.3			1.04	0.27
G09	30	0.05	-0.42	1.67	1.41	0.26	-0.09	-0.54	1.32	1.32	0.34
G10	31	-0.39			1.08	0.2	-0.69			1.07	0.22
	32	-1.19			1	0.27	-1.4			0.98	0.3
	33	1.57			1.09	0.2	1.23			1.05	0.25
	34	1.29			1.1	0.2	0.93			1.08	0.22
G11	35	-1.79	-1.55	1.52	0.87	0.55					
	36	-2.18	-0.93	1.58	0.88	0.56					
	37	-1.45	-1.86	0.66	0.92	0.51					
	38	-1.46	-0.03	2.04	0.96	0.51					
	39	-1.26	-1.56	0.95	0.89	0.55	New items for 2015				
G12	40	-1.45	-1.9	0.47	0.91	0.52					
	41	-1.54	-2.16	-0.08	0.93	0.48					
	42	-1.97	-1.01	0.87	0.98	0.47					
	43	-1.29	-0.97	0.75	1.02	0.46					

Differential Item Functioning (DIF) analysis for gender, LBOTE (Language Background Other than English), ATSI (Aboriginal and Torres Straight Islanders) and state/territory (each state/territory versus all other states/territories) was performed on each item and it was found that only a small amount of DIF exists in a few items. When the sample is divided into score groups along the scale as it was shown in Figure 11.3, the mean item score for the two subsets of students (for example, male students and female students) was very close. Most items show no DIF indicating that important subgroups of the NAP—SL sample responded consistently to the items in the survey. Figure 11.4 shows the gender DIF ICC for the “I learn science topics quickly” item.

Figure 11.4 Gender DIF Item characteristic curve for the “I learn science topics quickly” item



The Pearson correlation between survey scale scores and scale scores on the NAP—SL cognitive scale was found to be positive but only 0.2. Student achievement in science is positively correlated with the level of science perception although the correlation is not strong.

A multilevel modelling of the data that takes into account the structure of the data (students within schools) confirmed the overall positive contribution of the survey measure of each student to the explanation of variance in NAP—SL scale scores. The higher the survey measure the higher the NAP—SL scale score tends to be. In the same modelling of the data it was found that the overall performance of female students was higher than for male students, non-ATSI students higher than ATSI students, non-LBOTE students higher than LBOTE students and Northern Territory students lower than the other states/territory.

CHAPTER 12 MULTILEVEL MODELLING OF THE 2015 NAP—SL DATA

12.1. Introduction

The measurement scales constructed for 2015 NAP sample assessments – science literacy (NAP—SL) by fitting data to the Rasch model provide interval measures for reporting student achievement in science (Nwle) based on plausible values (see section 8.3.1) and student levels of perception of science (AbilSurvey) based on the survey data (see Chapter 11).

In addition to an overall measure of responses on the survey, scores are available for each student for each of the 12 groupings (G01 to G12) of the survey (See Table 12.1).

Table 12.1 Aspects of student perception of and attitudes toward science investigated in the survey

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

Note: G09 has not been included in the multilevel modelling because it consists of one item only "How often do you have science lessons at school?".

One of the aims of the study is to find out whether differences in achievement and in measures of science perception of various groups of students (for example, male students and female students) are statistically significant.

Multilevel modelling provides more reliable tests of statistical significance than traditional *t*-tests because it models the hierarchical structure of the data due to students attending different schools and takes into account the joint contribution of variables to the explanation of variance (Rasbash et al. 2003). Traditionally, a *t*-test is performed separately for each comparison ignoring the interaction between explanatory variables and failing to account for the clustering of students within schools. Such tests assume that students participate in the study as individuals and do not take into account similarities of students within schools. It is well known that students within a school are more likely to share similar characteristics than students enrolled in different schools. Ignoring clustering of students is the source of technical problems. For example, generally standard errors of regression coefficients would be underestimated (Rasbash et al. 2003). Detailed discussion of multilevel modelling advantages over traditional *t*-tests and statistical derivations may be found in the books by Bryk and Raudenbush (1992), Goldstein (1995) and Longford (1993).

The aim of the multilevel modelling of 2015 NAP—SL data is to determine whether observed differences are statistically significant for the purpose of making inferences about the population from which the sample was drawn. In addition to carrying out tests of statistical significance, fitting multilevel models to data makes it possible to estimate the proportion of the residual variance that is explained at school and individual student levels, after the estimation of the proportion of variance that is accounted for by explanatory variables.

12.2. Variables used

The response variables used in the 20 multilevel models fitted to the 2015 NAP—SL data were the student achievement in science (Nwle) in Models 1 to 10 and the measure of perception of science (AbilSurvey) in Models 11 to 20.

The demographic information included in the models as explanatory variables were gender, ATSI and LOBTE. The state/territory of each student was also used as an explanatory variable as well as Nwle, AbilSurvey, and the 11 survey groups as shown above.

Nwle: Continuous variable

The science achievement of each student is based on plausible values. This variable was normalised in the multilevel modelling with a mean of 0 and standard deviation of 1.

AbilSurvey: Continuous variable

The measure of the perception of science of each student was based on responses to items in the survey.

G01 to G12 (excluding G09): Continuous variable

These variables were treated as continuous even if they are obtained by adding scores on items of the survey (see 2015 NAP—SL survey). A positive regression coefficient indicates a positive correlation between a variable and the responses variable.

Gender: Categorical variable with female students 2 and male students 1.

The reference category was taken to be male students. A positive regression coefficient corresponds to a difference in achievement in favour of female students while a negative value in favour of male students.

ATSI: Categorical variable with ATSI students 1 and nonATSI students 0.

The reference category was taken to be nonATSI students. A positive regression coefficient corresponds to a difference in favour of ATSI students while a negative value in favour of nonATSI students.

LBOTE: Categorical variable with LBOTE students 1 and nonLBOTE students 0.

The reference category was taken to be nonLBOTE students. A positive regression coefficient corresponds to a difference in favour of LBOTE students while a negative value in favour of nonLBOTE students.

State/territory: Dummy variable with WA taken as the reference state in Models 1 and 11.

A comparison of students from the eight states/territories (S1 ACT, S2 NSW, S3 NT, S4 Qld, S5 SA, S6 Tas., S7 Vic. and S8 WA) required the selection of a state/territory with which each of the other states/territories could be compared. A positive regression coefficient for a state/territory indicates a difference in favour of that state/territory while a negative value in favour of WA.

12.3. The multilevel models

Twenty multilevel models were fitted to the 2015 NAP—SL data as follows:

12.3.1. Multilevel Models 1 to 10

The response variable in the **first set of ten models** is Nwle.

The explanatory variables in **Model 1** are gender, Aboriginal and Torres Strait Islanders (ATSI), Language Background Other than English (LBOTE), state/territory and AbilSurvey. In this model differences in science achievement of the whole sample for gender, ATSI, LBOTE and state/territory are investigated as well as the contribution of AbilSurvey to the explanation of variance in Nwle measures.

The explanatory variables in **Model 2** are gender, ATSI, LBOTE, and 11 survey groups (G09 was not included because it consists of one item only).

The explanatory variables in **Models 3 to 10**, a model for the data of each of the eight states/territories, are as for Model 2.

12.3.2. Multilevel Models 11 to 20

The response variable in the **second set of ten models** is AbilSurvey.

The explanatory variables in **Model 11** are gender, ATSI, LBOTE, state/territory and Nwle. In this model differences in the level of perception of science of the whole sample for gender, ATSI, LBOTE and state/territory are investigated as well as the contribution of Nwle to the explanation of variance in AbilSurvey measures.

The explanatory variables in **Model 12** are gender, ATSI, LBOTE, and Nwle.

The explanatory variables in **Models 13 to 20**, a model for the data of each of the eight states/territories, are as for Model 12.

Each of the models fitted to the data consists of a two-level hierarchical structure: students (level 1) clustered within schools (level 2).

MLwiN software was used for this modelling (Rasbash et al. 2015). Output from the MLwiN program for the multilevel modelling of the data is shown for four of the models (Models 1, 2, 11 and 12). The table following the MLwiN output of each of these four models lists the regression coefficients for each explanatory variable, the corresponding standard error and the value of the associated *t*-statistic (Tables 12.2, 12.3, 12.5 and 12.6). For Models 3 to 10 and 13 to 20, a summary of the results of the multilevel modelling of the data of each state/territory is shown in Tables 12.4 and 12.7.

A positive regression coefficient for an explanatory variable indicates that the two variables (response and explanatory) are correlated positively. The ratio of the regression coefficient to its standard error is distributed according to the *t*-statistic. A value of *t* greater than 1.96 indicates a statistically significant contribution to the variance beyond the 0.05 probability level, that is, the 95 per cent confidence level.

The residual variance, after the estimation of the percentage of variance accounted for by explanatory variables, is partly at the school level and partly at the student level. The percentage of the residual variance at the school level is shown for each model.

The contribution of a two-category variable to the explanation of variance depends only on the difference in the achievement of the two groups. The level of the statistical significance of the difference in achievement indicated by a regression coefficient depends on sample size since error decreases with sample size. Therefore different levels of statistical significance may be associated with regression coefficients of similar values.

12.4. Twenty fitted models

12.4.1. Model 1 response variable: Nwle

This model investigates the contribution of variables to the explanation of science achievement variance with all students in the sample.

Explanatory variables: gender, ATSI, LBOTE, AbilSurvey, state/territory (DUMMY variable with WA taken as the reference state/territory).

The MLwiN output is shown in Figures 12.1 and 12.2.

Figure 12.1 Baseline model 1

$$\text{Nwle}_{STUD,\ SchoolID} \sim \text{N}(XB,\ \Omega)$$

$$\text{Nwle}_{STUD,\ SchoolID} = \beta_{0STUD,\ SchoolID}\text{Const}$$

$$\beta_{0STUD,\ SchoolID} = -0.025(0.021) + u_{0SchoolID} + e_{0STUD,\ SchoolID}$$

$$\left[u_{0SchoolID}\right] \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[0.221(0.015)\right]$$

$$\left[e_{0STUD,\ SchoolID}\right] \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[0.789(0.010)\right]$$

$$-2*loglikelihood(IGLS\ Deviance) = 33404.468(12410\text{ of }12410\text{ cases in use})$$

Figure 12.2 Model 1 with gender, ATSI, LBOTE, AbilSurvey and state/territory

$$\mathrm{Nwle}_{STUD,\ SchoolID} \sim \mathrm{N}(XB,\ \Omega)$$

$$\begin{aligned}
\mathrm{Nwle}_{STUD,\ SchoolID} = \ & \beta_{0STUD,\ SchoolID}\mathrm{Const} + \\
& 0.132(0.016)\mathrm{Gender}_{STUD,\ SchoolID} + \\
& \text{-}0.029(0.009)\mathrm{ATSI}_{STUD,\ SchoolID} + \\
& \text{-}0.075(0.016)\mathrm{LBOTE}_{STUD,\ SchoolID} + \\
& 0.229(0.010)\mathrm{AbilSurvey}_{STUD,\ SchoolID} + \\
& 0.078(0.078)\mathrm{S1}_{SchoolID} + \\
& 0.020(0.069)\mathrm{S2}_{SchoolID} + \\
& \text{-}0.539(0.092)\mathrm{S3}_{SchoolID} + \\
& \text{-}0.072(0.070)\mathrm{S4}_{SchoolID} + \\
& \text{-}0.102(0.071)\mathrm{S5}_{SchoolID} + \\
& 0.026(0.077)\mathrm{S6}_{SchoolID} + \\
& \text{-}0.080(0.069)\mathrm{S7}_{SchoolID}
\end{aligned}$$

$$\beta_{0STUD,\ SchoolID} = \text{-}0.280(0.056) + u_{0SchoolID} + e_{0STUD,\ SchoolID}$$

$$\left[u_{0SchoolID}\right] \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[0.165(0.012)\right]$$

$$\left[e_{0STUD,\ SchoolID}\right] \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[0.748(0.010)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 30111.557(11445 of 12410 cases in use)

Table 12.2 shows the summary of results for Model 1. It shows the regression coefficients of the explanatory variables with their standard errors, the corresponding t-values and the conclusion about statistical significance beyond the 0.05 probability level (t > 1.96).

Table 12.2 Summary of mm results for Model 1

Variable	Coefficient	Error	t-statistic	Comment
Gender	0.132	0.016	8.3	Girls sign higher than boys
ATSI	–0.029	0.009	3.2	nonATSI significantly higher than ATSI
LBOTE	–0.075	0.016	4.7	nonLBOTE significantly higher than LBOTE
AbilSurvey	0.229	0.010	22.9	Significant positive association with Nwle
NSW	0.020	0.069	0.3	NSW not significantly higher than WA
Vic.	–0.080	0.069	1.2	Vic. not significantly lower than WA
Qld	–0.072	0.070	1.0	Qld. not significantly lower than WA
SA	–0.102	0.071	1.4	SA not significantly lower than WA
Tas.	0.026	0.077	0.3	Tas. not significantly higher than WA
ACT	0.078	0.078	1.0	ACT not significantly higher than WA
NT	–0.539	0.092	5.9	NT significantly lower than WA

Variance	Baseline model	Model with variables
Between schools	0.221	0.165
Between student	0.789	0.748

		Residual variance	Explained variance
Total variance	1.01	0.913	0.097

		% total variance	% residual variance
Explained variance	0.097	9.6	
School	0.165	16.3	18.1
Student	0.748	74.1	81.9
Total	1.01	100	100

These explanatory variables account for 9.6 per cent of the total variance. After accounting for the effect of explanatory variables, 18.1 per cent of the residual variance is at the school level and 81.9 per cent at the individual student level.

12.4.2. Model 2 response variable: Nwle

This model investigates the contribution of survey groups, in addition to gender, ATSI and LBOTE, to the explanation of science achievement variance with all students in the sample.

Explanatory variables: gender, ATSI, LBOTE, G01 to G12 except G09.

The MLwiN output is shown in Figures 12.3 and 12.4.

Figure 12.3 Baseline Model 2

$$\text{Nwle}_{STUD, SchoolID} \sim \text{N}(XB, \Omega)$$

$$\text{Nwle}_{STUD, SchoolID} = \beta_{0STUD, SchoolID}\text{Const}$$

$$\beta_{0STUD, SchoolID} = -0.025(0.021) + u_{0SchoolID} + e_{0STUD, SchoolID}$$

$$\left[u_{0SchoolID}\right] \sim \text{N}(0, \Omega_u) : \Omega_u = \left[0.221(0.015)\right]$$

$$\left[e_{0STUD, SchoolID}\right] \sim \text{N}(0, \Omega_e) : \Omega_e = \left[0.789(0.010)\right]$$

$-2*loglikelihood(IGLS\ Deviance)$ = 33404.468(12410 of 12410 cases in use)

Figure 12.4 Model 2 with gender, ATSI, LBOTE and G01 to G12 (except G09)

$$\text{Nwle}_{STUD, SchoolID} \sim \text{N}(XB, \Omega)$$

$$\begin{aligned}\text{Nwle}_{STUD, SchoolID} = \beta_{0STUD, SchoolID}&\text{Const} + \\ &0.093(0.019)\text{Gender}_{STUD, SchoolID} + \\ &-0.014(0.010)\text{ATSI}_{STUD, SchoolID} + \\ &-0.073(0.018)\text{LBOTE}_{STUD, SchoolID} + \\ &-0.007(0.006)\text{Q1}_{STUD, SchoolID} + \\ &0.115(0.007)\text{Q2}_{STUD, SchoolID} + \\ &0.038(0.006)\text{Q3}_{STUD, SchoolID} + \\ &-0.113(0.005)\text{Q4}_{STUD, SchoolID} + \\ &0.001(0.005)\text{Q5}_{STUD, SchoolID} + \\ &-0.049(0.008)\text{Q6}_{STUD, SchoolID} + \\ &-0.008(0.005)\text{Q7}_{STUD, SchoolID} + \\ &0.075(0.012)\text{Q8}_{STUD, SchoolID} + \\ &-0.057(0.011)\text{Q10}_{STUD, SchoolID} + \\ &0.063(0.006)\text{Q11}_{STUD, SchoolID} + \\ &0.060(0.006)\text{Q12}_{STUD, SchoolID}\end{aligned}$$

$$\beta_{0STUD, SchoolID} = -0.858(0.076) + u_{0SchoolID} + e_{0STUD, SchoolID}$$

$$\left[u_{0SchoolID}\right] \sim \text{N}(0, \Omega_u) : \Omega_u = \left[0.125(0.011)\right]$$

$$\left[e_{0STUD, SchoolID}\right] \sim \text{N}(0, \Omega_e) : \Omega_e = \left[0.605(0.011)\right]$$

$-2*loglikelihood(IGLS\ Deviance)$ = 16519.568(6779 of 12410 cases in use)

Table 12.3 shows the summary of results for Model 2. It shows the regression coefficients of the explanatory variables with their standard errors, the corresponding t-values and the conclusion about statistical significance beyond the 0.05 probability level (t > 1.96).

Table 12.3 Summary of mm results for Model 2

Variables	Coefficient	Error	t-statistic	Comment
Gender	0.093	0.019	4.9	Girls significantly higher than boys
ATSI	–0.014	0.010	1.4	NonATSI not significantly higher than ATSI
LBOTE	–0.073	0.018	4.1	NonLBOTE significantly higher than LBOTE
G01	–0.007	0.006	1.2	negative not significant
G02	0.115	0.007	16.4	positive significant
G03	0.038	0.006	6.3	positive significant
G04	–0.113	0.005	22.6	negative significant
G05	0.001	0.005	0.2	positive not significant
G06	–0.049	0.008	6.1	negative significant
G07	–0.008	0.005	1.6	negative not significant
G08	0.075	0.012	6.3	positive significant
Q10	–0.057	0.011	5.2	negative significant
Q11	0.063	0.006	10.5	positive significant
Q12	0.060	0.006	10.0	positive significant

Variance	Baseline model	Model with variables
Between schools	0.221	0.125
Between student	0.789	0.605

		Residual variance	Explained variance
Total variance	1.01	0.730	0.280

		% total variance	% residual variance
Explained variance	0.280	27.7	
School	0.125	12.4	17.1
Student	0.605	59.9	82.9
Total	1.01	100	100

These explanatory variables account for 27.7 per cent of the total variance. After accounting for the effect of explanatory variables, 17.1 per cent of the residual variance is at the school level and 82.9 per cent at the individual student level.

12.4.3. Models 3 to 10 response variable: Nwle

These eight models, each with the data of a state/territory, investigate the contribution of survey groups, in addition to gender, ATSI and LBOTE, to the explanation of science achievement variance.

Explanatory variables: gender, ATSI, LBOTE, G01 to G12 except G09.

Table 12.4 shows the summary of the results for Models 3 to 10.

Table 12.4 Summary of results for Models 3 to 10 (one for the data of each state/territory)

	all cases	NSW	Vic.	Qld	WA	SA	Tas.	ACT	NT
Gender	p y	p y	p	0	p	p y	p	p y	p
ATSI	n	n	p	n y	0	n	p	n	n y
LBOTE	n y	p y	n y	p	n y	n y	n	n	n y
G01	n	n	p	p	n	n	p	n	n
G02	p y	p y	p y	p y	p y	p y	p y	p y	p y
G03	p y	p	p y	p y	p y	p	p y	p	p y
G04	n y	n y	n y	n y	n y	n y	n y	n y	n y
G05	p	p	n	n	p	p	p	p	n
G06	n y	n y	n y	n	n y	n y	n	n	n
G07	n	p	p	n	n	n	n	n	n
G08	p y	p	p	p y	p y	p y	p	p	p
Q10	n y	n y	n	n y	n	n	n	n	n y
Q11	p y	p y	p y	p y	p y	p y	p y	p y	p y
Q12	p y	p y	p y	p y	p y	p y	p y	p y	p y

p means positive regression coefficient
n means negative regression coefficient
y means statistically significant at the 0.05 level (t>1.96)

12.4.4. Model 11 response variable: AbilSurvey

This model investigates the contribution of explanatory variables in the measure of science perception with all students in the sample.

Explanatory variables: gender, ATSI, LBOTE, Nwle, state/territory (DUMMY variable with WA taken as the reference state).

The MLwiN output is shown in Figures 12.5 and 12.6.

Figure 12.5 Baseline model 11

$$\text{AbilSurvey}_{STUD,\ SchoolID} \sim \mathrm{N}(XB,\ \Omega)$$

$$\text{AbilSurvey}_{STUD,\ SchoolID} = \beta_{0STUD,\ SchoolID}\text{Const}$$

$$\beta_{0STUD,\ SchoolID} = 0.771(0.013) + u_{0SchoolID} + e_{0STUD,\ SchoolID}$$

$$\left[u_{0SchoolID}\right] \sim \mathrm{N}(0,\ \Omega_u) \ :\ \Omega_u = \left[0.056(0.005)\right]$$

$$\left[e_{0STUD,\ SchoolID}\right] \sim \mathrm{N}(0,\ \Omega_e) \ :\ \Omega_e = \left[0.624(0.008)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 27656.475(11445 of 12410 cases in use)

Figure 12.6 Model 11 with gender, ATSI, LBOTE, Nwle and state/territory

$$\text{AbilSurvey}_{STUD,\ SchoolID} \sim \text{N}(XB,\ \Omega)$$

$$\begin{aligned}\text{AbilSurvey}_{STUD,\ SchoolID} = {} & \beta_{0STUD,\ SchoolID}\text{Const} + \\ & \text{-0.138(0.014)Gender}_{STUD,\ SchoolID} + \\ & \text{0.010(0.007)ATSI}_{STUD,\ SchoolID} + \\ & \text{0.060(0.012)LBOTE}_{STUD,\ SchoolID} + \\ & \text{0.035(0.048)S1}_{SchoolID} + \\ & \text{0.004(0.042)S2}_{SchoolID} + \\ & \text{-0.031(0.058)S3}_{SchoolID} + \\ & \text{0.027(0.043)S4}_{SchoolID} + \\ & \text{-0.133(0.044)S5}_{SchoolID} + \\ & \text{-0.090(0.048)S6}_{SchoolID} + \\ & \text{-0.030(0.042)S7}_{SchoolID} + \\ & \text{0.178(0.008)Nwle}_{STUD,\ SchoolID}\end{aligned}$$

$$\beta_{0STUD,\ SchoolID} = 0.979(0.037) + u_{0SchoolID} + e_{0STUD,\ SchoolID}$$

$$\left[u_{0SchoolID}\right] \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[0.047(0.005)\right]$$

$$\left[e_{0STUD,\ SchoolID}\right] \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[0.595(0.008)\right]$$

$$\textit{-2*loglikelihood(IGLS Deviance)} = 27060.156(11445 \text{ of } 12410 \text{ cases in use})$$

Table 12.5 shows the summary of results for Model 11. It shows the regression coefficients of the explanatory variables with their standard errors, the corresponding t-values and the conclusion about statistical significance beyond the 0.05 probability level (t > 1.96).

Table 12.5 Summary of results for Model 11

Variable	Coefficient	Error	t-statistic	Comment
Gender	–0.138	0.014	9.9	Boys significantly higher than girls
ATSI	0.010	0.007	1.4	ATSI higher than nonATSI not significant
LBOTE	0.060	0.012	5.0	LBOTE significantly higher than NonLBOTE
Nwle	0.178	0.008	22.3	Significantly positive association with survey
NSW	0.004	0.042	0.1	NSW higher than WA not significant
Vic.	–0.030	0.042	0.7	Vic. lower than WA not significant
Qld	0.027	0.043	0.6	Qld higher than WA not significant
SA	–0.133	0.044	3.0	SA significantly lower than WA
Tas.	–0.090	0.048	1.9	Tas. lower than WA not significant
ACT	0.035	0.048	0.7	ACT higher than WA not significant
NT	–0.031	0.058	0.5	NT lower than WA not significant
Variance	**Baseline model**	**Model with variables**		
Between schools	0.056	0.047		
Between student	0.624	0.595		

		Residual variance	Explained variance
Total variance	0.68	0.642	0.038

		% total variance	% residual variance
Explained variance	0.038	5.6	
School	0.047	6.9	7.3
Student	0.595	87.5	92.7
Total	0.68	100	100

These explanatory variables account for 5.6 per cent of the total variance. After accounting for the effect of explanatory variables, 7.3 per cent of the residual variance is at the school level and 92.7 per cent at the individual student level.

12.4.5. Model 12 response variable: AbilSurvey

This model investigates the contribution of explanatory variables in the measure of science perception with all students in the sample.

Explanatory variables: gender, ATSI, LBOTE, Nwle.

The MLwiN output is shown in Figures 12.7 and 12.8.

Figure 12.7 Baseline model 12

$$\text{AbilSurvey}_{STUD,\ SchoolID} \sim \text{N}(XB,\ \Omega)$$
$$\text{AbilSurvey}_{STUD,\ SchoolID} = \beta_{0STUD,\ SchoolID}\text{Const}$$
$$\beta_{0STUD,\ SchoolID} = 0.771(0.013) + u_{0SchoolID} + e_{0STUD,\ SchoolID}$$
$$\left[u_{0SchoolID}\right] \sim \text{N}(0,\ \Omega_u) \ :\ \Omega_u = \left[0.056(0.005)\right]$$
$$\left[e_{0STUD,\ SchoolID}\right] \sim \text{N}(0,\ \Omega_e) \ :\ \Omega_e = \left[0.624(0.008)\right]$$
$$\textit{-2*loglikelihood(IGLS Deviance)} = 27656.475(11445 \text{ of } 12410 \text{ cases in use})$$

Figure 12.8 Model 12 with gender, ATSI, LBOTE and Nwle

$$\text{AbilSurvey}_{STUD,\ SchoolID} \sim \text{N}(XB,\ \Omega)$$
$$\begin{aligned}\text{AbilSurvey}_{STUD,\ SchoolID} = {} & \beta_{0STUD,\ SchoolID}\text{Const} + \\ & -0.139(0.014)\text{Gender}_{STUD,\ SchoolID} + \\ & 0.010(0.007)\text{ATSI}_{STUD,\ SchoolID} + \\ & 0.062(0.012)\text{LBOTE}_{STUD,\ SchoolID} + \\ & 0.179(0.008)\text{Nwle}_{STUD,\ SchoolID}\end{aligned}$$
$$\beta_{0STUD,\ SchoolID} = 0.956(0.024) + u_{0SchoolID} + e_{0STUD,\ SchoolID}$$
$$\left[u_{0SchoolID}\right] \sim \text{N}(0,\ \Omega_u) \ :\ \Omega_u = \left[0.050(0.005)\right]$$
$$\left[e_{0STUD,\ SchoolID}\right] \sim \text{N}(0,\ \Omega_e) \ :\ \Omega_e = \left[0.595(0.008)\right]$$
$$\textit{-2*loglikelihood(IGLS Deviance)} = 27081.350(11445 \text{ of } 12410 \text{ cases in use})$$

Table 12.6 shows the summary of results for Model 12. It shows the regression coefficients of the explanatory variables with their standard errors, the corresponding t-values and the conclusion about statistical significance beyond the 0.05 probability level ($t > 1.96$).

Table 12.6 Summary of mm results for Model 12

Variable	Coefficient	Error	t-statistic	Comment
Gender	–0.139	0.014	9.9	Boys significantly higher than girls
ATSI	0.010	0.007	1.4	ATSI higher than nonATSI not significant
LBOTE	0.062	0.012	5.2	LBOTE significantly higher than nonLBOTE
Nwle	0.179	0.008	22.4	Significantly positive association with survey

Variance	Baseline model	Model with variables
Between schools	0.056	0.050
Between student	0.624	0.595

		Residual variance	Explained variance
Total variance	0.680	0.645	0.035

		% total variance	% residual variance
Explained variance	0.035	5.1	
School	0.050	7.4	7.8
Student	0.595	87.5	92.2
Total	0.680	100	100

These explanatory variables account for 5.1 per cent of the total variance. After accounting for the effect of explanatory variables, 7.8 per cent of the residual variance is at the school level and 92.2 per cent at the individual student level.

12.4.6. Models 13 to 20 response variable: AbilSurvey

These eight models, each with the data of a state/territory, investigate the contribution of gender ATSI, LBOTE and Nwle to the explanation of measures of science perception variance.

Explanatory variables: gender, ATSI, LBOTE, Nwle.

Table 12.7 shows the summary of results for Models 13 to 20.

Table 12.7 Summary of results for Models 13 to 20 (one for the data of each state/territory)

Variable	all cases	NSW	Vic.	Qld	WA	SA	Tas.	ACT	NT
Gender	n y	n y	n y	n y	n y	n y	n y	n y	n
ATSI	p	p	n	p	p	n	n	p	n
LBOTE	p y	p y	p y	p	p	p y	n	p	n
Nwle	p y	p y	p y	p y	p y	p y	p y	p y	p y

p means positive regression coefficient

n means negative regression coefficient
y means statistically significant at the 0.05 level (t>1.96)

12.5. Summary of results

Results are reported for each explanatory variable separately for response variables Nwle and AbilSurvey.

12.5.1. Response variable Nwle

Gender: In Models 1 and 2 girls performed significantly better than boys. This result was confirmed in ACT, NSW and SA only. In the NT, Tas., Vic. and WA the better performance of girls was not statistically significant. No difference was observed for Qld students.

ATSI: In Model 1 NonATSI students performed significantly better than ATSI students but in Model 2 the better performance of NonATSI students was not statistically significant. In the NT and Qld the NonATSI students performed significantly better but not in ACT, NSW and SA. In Tas. and Vic. ATSI students' better performance was not significant. No difference was observed for the WA students.

LBOTE: In Models 1 and 2 NonLBOTE students performed significantly better than LBOTE students. This result was confirmed in the NT, SA, Vic. and WA, but it was not significant in the ACT and Tas. The better performance of LBOTE students in NSW was significant but it was not in Qld.

State/territory: The state/territory variable was not significant for all states/territories except NT. WA students performed significantly better than NT students.

G01 Interest in science: Overall, this variable does not contribute significantly to an explanation of science achievement variance and this result is confirmed in the data of each state/territory.

G02 Self-concept of science ability: Overall, this variable is positively related in a significant way to achievement in science. This result is confirmed in each state/territory.

G03 Value of science: Overall, this variable is positively related in a significant way to achievement in science. It is positively related in each state/territory but not significantly in ACT, NSW and SA.

G04 The nature of science 1: Overall, this variable is negatively related in a significant way to achievement in science overall and in each state/territory. Further empirical study is needed to fully explain this result.

G05 Science-related activities outside school: Overall, this variable is weakly related to science achievement. This result is confirmed in each state/territory.

G06 Science-related activities at school: Overall, this variable is negatively related to science achievement. This result is confirmed in each state/territory but it is not significant in ACT, NT, Qld and Tas. Further empirical study is needed to fully explain this result.

G07 Science teaching: Overall, this variable is weakly related to science achievement. This result is confirmed in each state/territory.

G08 Science topics studied: Overall, this variable is positively related to science achievement. The same result has been observed in each state/territory but it is significant only in Qld, SA and WA.

G10 Science teaching 2: Overall, this variable is negatively related to science achievement. There is a negative relation in each state/territory but it is significant only in NSW, NT and Qld. Further empirical study is needed to fully explain this result.

G11 The nature of science 2: Overall, this variable is positively related to science achievement. This result is significant in each state/territory.

G12 Who is involved in science: Overall, this variable is positively related to science achievement. This result is significant in each state/territory.

12.5.2. Response variable AbilSurvey

Gender: In Models 11 and 12 boys performed significantly higher than girls. This result is confirmed in each state/territory except NT for which the better performance of boys is not significant.

ATSI: In Models 11 and 12 the difference between NonATSI and ATSI students is not significant. This result is confirmed in each state/territory.

LBOTE: In Models 11 and 12 LBOTE students performed significantly higher than NonLBOTE students. This result is confirmed in NSW, SA and Vic. The higher performance of LBOTE students in ACT, Qld and WA is not significant as well as the higher performance of NonLBOTE students in NT and Tas.

State/territory: The state/territory variable is not significant for all states/territories except SA: WA students performed significantly higher than SA students.

Nwle: In Models 11 and 12, science achievement has a significant positive association with the measure of perception of science. This result is confirmed in each state/territory.

REFERENCES

Ball, S., Rae, I., & Tognolini, J. (2000). Options for the assessment and reporting of primary students in the key learning area of science to be used for the reporting of nationally comparable outcomes of schooling within the context of the National Goals for Schooling in the Twenty-First Century: National Education Performance Monitoring Taskforce.

Beaton, A.E., & Gonzalez, E. (1995). NAEP primer. Chestnut Hill, MA: Boston College.

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

Bryk, A.S. & Raudenbush, S.W. (1992). *Hierarchical linear models*. Newbury Park, California: Sage.

Goldstein, H. (1995). *Multilevel statistical models*. London: Arnold.

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

IEA (2009). TIMSS 2007 Technical Report. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

IMS Question and Test Interoperability v2.1 Final Specification 31 August 2012.

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

Longford, N.T. (1993). *Random coefficient models*. Oxford: Clarendon Press.

Mislevy, R.J., Beaton, A.E., Kaplan, B., & Sheehan, K.M. (1992). Estimating population characteristics from sparse matrix samples of item responses.

Journal of Educational Measurement, 29, 131–154.

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

OECD (2000). PISA 2000 Technical Report. OECD: Paris.

OECD (2012). PISA 2009 Technical Report. OECD: Paris.

Rasbash, J., Steele, F. & Browne, W. (2003). *A User's Guide to MLwiN*. Centre for Multilevel Modelling, Institute of Education, University of London.

Rasbash, J., Browne, W., Healy, M., Cameron, B. & Charlton, C. (2015). *MLwiN Version 2.35*. Centre for Multilevel Modelling, University of Bristol.

Rust, K.F., & Rao, J.N.K. (1996). Variance estimation for complex surveys using replication techniques. Statistical Methods in Medical Research, 5, 283–310.

Wu, M.L. (2005). The role of plausible values in large-scale surveys. Studies in Educational Evaluation, 31, 114–128.

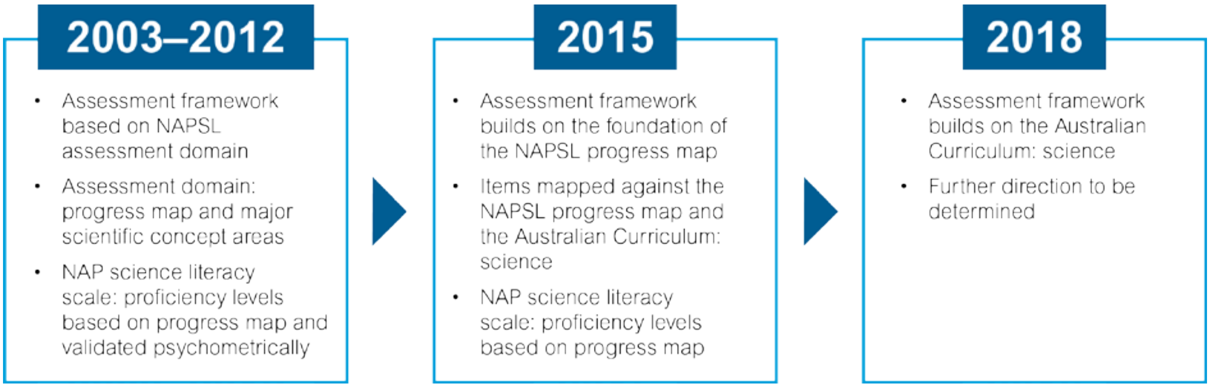
APPENDIX 1 NATIONAL YEAR 6 PRIMARY SCIENCE ASSESSMENT DOMAIN

The complete NAP sample assessments - science literacy (NAP—SL) assessment framework document can be found at the assessment frameworks page of the NAP website: **www.nap.edu.au/nap-sample-assessments/assessment-frameworks**

Introduction and background

This appendix provides a description and rationale for the framework that formed the basis of the 2015 NAP—SL. This framework is intended to serve as a bridge between past and future NAP—SL cycles (see Figure A1.1).

Figure A1.1 Development of the NAP—SL assessment framework and science literacy scale



One of the main objectives of NAP—SL is to monitor trends in science literacy performance over time. To enable effective historical comparison, it was important that the underlying construct of the NAP—SL assessment was maintained. At the same time, the new Australian Curriculum: science provided an opportunity to bring in aspects of science literacy that had not been assessed in previous cycles.

As a consequence, this framework was intended to describe the existing NAP—SL construct in terms of the Australian Curriculum: science and augment it in ways that reflect developments in the Australian Curriculum. It is expected that future NAP—SL assessments will be based mainly on the Australian Curriculum: science.

Additionally, this cycle marked a major advance in the way the NAP—SL construct was assessed. For the first time, NAP—SL was assessed online, with items developed to be compatible with the IMS Question & Test Interoperability™ Specification (**www.imsglobal.org/question**).

Science literacy assessment domain: A historical perspective

NAP—SL assesses science literacy in the context of a student's ability to apply broad conceptual understandings of science in order to make sense of the world; to understand natural phenomena; and to interpret media reports about scientific issues. It also includes the ability to ask investigable questions; conduct investigations; collect and interpret data; and make informed decisions. This construct was developed from an earlier definition of science literacy used by the Organisation for Economic Co-operation and Development (OECD) – Programme for International Student Assessment (PISA).

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

Science literacy progress map

For previous cycles of NAP—SL, a science literacy progress map was developed based on the agreed definition of science literacy and an analysis of the existing state and territory curriculum and assessment frameworks. The progress map describes the development of science literacy across three strands:

> **Strand A**: formulating or identifying investigable questions and hypotheses; planning investigations; and collecting evidence.
>
> **Strand B**: interpreting evidence and drawing conclusions from students' own or others' data; critiquing the trustworthiness of evidence and claims made by others; and communicating findings.
>
> **Strand C**: using science understandings for describing and explaining natural phenomena; and for interpreting reports about phenomena.

Major scientific concept areas

In previous cycles, the science literacy domain also comprised four major scientific concept areas: Earth and space; Energy and force; Living things; and Matter. Each concept area included a set of concepts – broad statements of scientific understandings that Year 6 students would be expected to demonstrate – found most widely in the various state and territory documents.

An illustrative list of examples for each of the concept areas provided elaboration of these broad conceptual statements and, in conjunction with the science literacy progress map, was used in previous cycles as a guide in the development of assessment items.

The Australian Curriculum: science

In 2010, the federal, state and territory education ministers of Australia endorsed the release of the Australian Curriculum: science. Since that time, the states and territories have been working on the implementation of the new curriculum in schools.

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

Accordingly, the Australian Curriculum: science has three interrelated strands – science as a human endeavour, science inquiry skills and science understanding – which are designed to be taught in an integrated way. Together, these three strands provide students with understanding, knowledge and skills through which they can develop a scientific view of the world. Students are challenged to explore science, its concepts, nature and uses through clearly described inquiry processes. Table A1.1 lists the strands of the curriculum and the sub-strands within each strand.

Table A1.1 Strands and sub-strands in the Australian Curriculum: science

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

Previous cycles of NAP—SL were developed with no common science curriculum across the states and territories. With the implementation of the Australian Curriculum: science in all states and territories in 2014, it was important that the NAP—SL construct was mapped onto and described in terms of the new Australian Curriculum: science.

Table A1.2 shows how the NAP—SL progress map strands map onto the strands/sub-strands of the Australian Curriculum: science.

Table A1.2 The NAP—SL progress map links to the strands/sub-strands of the Australian Curriculum: science

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

There is a high degree of alignment between NAP—SL progress map Strand A and Strand B and the science inquiry skills strand of the Australian Curriculum: science.

The NAP—SL progress map Strand C provides an abstract representation of progression in students' use of science concepts for describing and explaining natural phenomena and interpreting reports about phenomena that makes no reference to particular science concepts.

In previous NAP—SL cycles, Strand C has provided guidance for the development of items that reflect levels of increasing complexity and abstraction in students' understanding of science concepts. In the absence of a common science curriculum across states and territories, the major scientific concept areas provided the contexts and specific concepts used to assess science understanding.

Table A1.3 shows how the NAP—SL major scientific concept areas map onto the sub-strand of the Australian Curriculum: science – science understanding. The science understanding strand of the Australian Curriculum: science provides guidance about the specific concepts to be assessed in the NAP—SL tests. Table A1.7 shows the mapping between the major scientific concept areas and the science understanding strand at a finer level. There is no explicit equivalent of the abstracted progression articulated in Strand C in the Australian Curriculum: science.

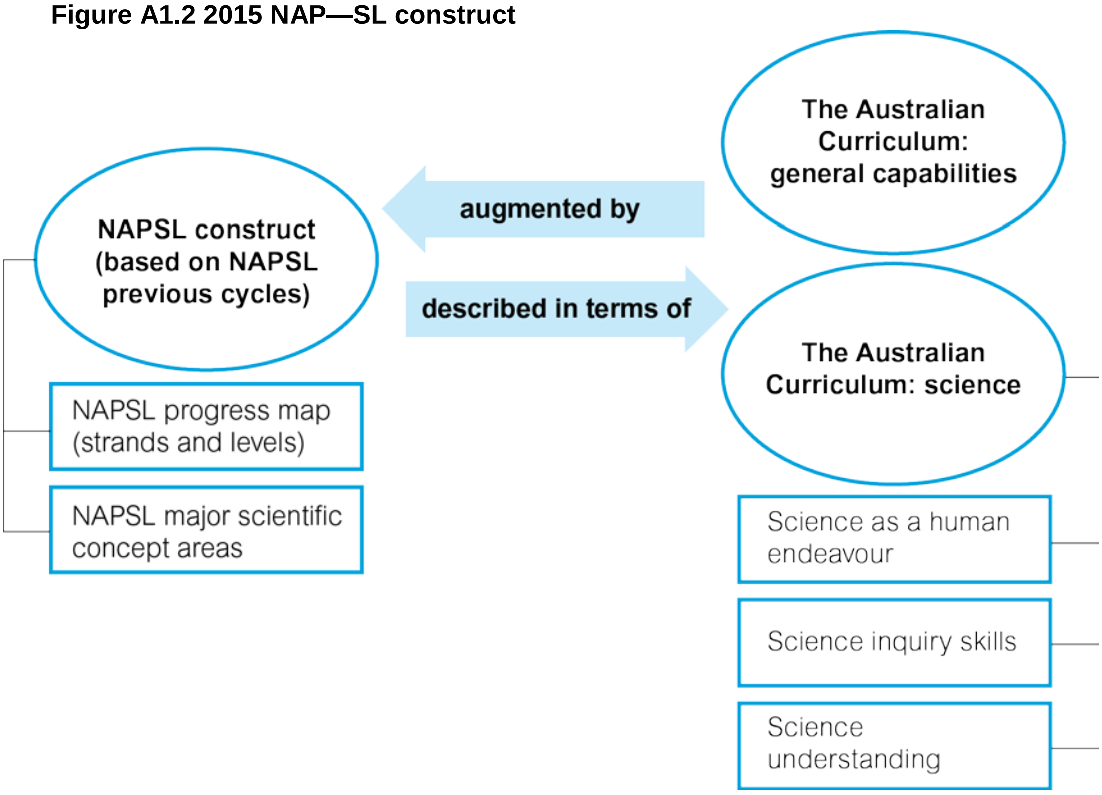
Table A1.3 The NAP—SL major scientific concept areas link to the sub-strand of the Australian Curriculum: science – science understanding

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

The 2015 NAP—SL Assessment Domain

One of the main objectives of NAP—SL is to monitor trends in science literacy performance over time. To enable effective historical comparison, it was important that the underlying construct of the NAP—SL assessment was maintained. At the same time, the Australian Curriculum: science provided an opportunity to bring in aspects of science literacy that had not been assessed in past cycles (see Figure A1.2).

Figure A1.2 2015 NAP—SL construct



As in previous cycles, the conceptual framework for the assessment of science literacy comprised the NAP—SL progress map, which describes growth in science inquiry skills and conceptual understandings (see Table A1.6). The A and B strands of the progress map were mapped against and augmented by the content descriptions articulated in the science inquiry skills strand of the Australian Curriculum: science.

Progress map Strand C describes a progression of how students use science concepts for describing and explaining natural phenomena and remained relevant for the 2015 cycle as it articulates performance levels that link directly to proficiency levels on the science literacy scale.

In previous cycles, the major scientific concept areas provided an indication of the pool of concepts from which item content related to science understanding was drawn. The science understanding strand of the Australian Curriculum: science provided more specific content descriptions related to science understanding and were used in the 2015 NAP—SL cycle as the basis for item development related to understanding of science concepts. As noted in the previous section, there is a high degree of alignment between the major scientific concept areas and the science understanding strand content descriptions.

Items in previous NAP—SL cycles did not assess understandings related to the scientific endeavour (nature and development of science; use and influence of science). In the 2015 NAP—SL cycle, the science as a human endeavour strand of the curriculum informed both stimulus context and attitudinal aspects of the student survey. Where appropriate, items were also classified against content descriptions related to this strand. In future cycles, the assessment domain will be further expanded to explicitly include the assessment of student understanding of the nature and development of science.

Given that the NAP—SL test instruments were constructed within constraints of test length, the content covered in the test was intended to be a sampling of available concepts rather than an exhaustive assessment of all the concepts listed. The focus of the assessment was on concepts and skills from the Australian Curriculum: science Years 4–6. However, as the Australian Curriculum represents a continuum, it was appropriate that concepts and skills from Foundation through to Year 6 be considered. This is consistent with the approach taken in previous NAP—SL cycles, in which the progress map articulates a progression in development of understanding and skills. Topic areas that are covered in Years 7–10 of the Australian Curriculum: science were not assessed.

The Australian Curriculum: general capabilities

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

> **Literacy:** aspects of the literacy capability are found within the reading comprehension demands of both the stimuli and the items of NAP—SL.
>
> **Numeracy:** aspects of the numeracy capability are found within NAP—SL, including the reading and construction of graphs and tables, calculations and measurement, as well as some elements of spatial reasoning.
>
> **Information and communication technology (ICT):** aspects of the ICT capability arose from online delivery.
>
> **Critical and creative thinking:** aspects of the critical and creative thinking capability arose from important cognitive skills inherent in scientific inquiry.

Items and stimulus could draw on aspects of the personal and social capability, the ethical understanding capability and the intercultural understanding capability when appropriate. The following sections describe in more detail how the relevant capabilities planned to be reflected in the NAP—SL assessment. It should be noted though that the focus of NAP—SL was the assessment of science literacy and not of general capabilities.

Literacy

In the NAP—SL tests, students were required to read and comprehend written and graphical stimulus. For some items they were also expected to write coherent explanations of ideas.

While literacy plays an important role in science learning and assessment, it was important that the difficulty of items and stimuli did not derive mainly from the amount and the complexity of the stimulus material and instructions. The NAP—SL stimuli and items were written to a level appropriate for the students assessed. The literacy demand of items was monitored by expert review to ensure that it was at an appropriate level for the assessment.

Numeracy

Many elements of numeracy are evident in the Australian Curriculum: science, particularly in science inquiry skills. These include practical measurement and the collection, representation and interpretation of data from investigations. Comparison between the Australian Curriculum: science and the Australian Curriculum: mathematics also shows a number of overlaps.

In the NAP—SL assessments, students were expected to show dispositions and capacities to use appropriate mathematical knowledge and skills as outlined in the Australian Curriculum: science (**www.australiancurriculum.edu.au/Science/General-capabilities**).

Information and communication technology

ICT plays a role when students take an online assessment. However, while students were interacting with technology to complete the assessment, it was not intended that the technology be a source of difficulty within NAP—SL. Consequently, it was important that access to, and navigation within, the test required relatively basic ICT skills.

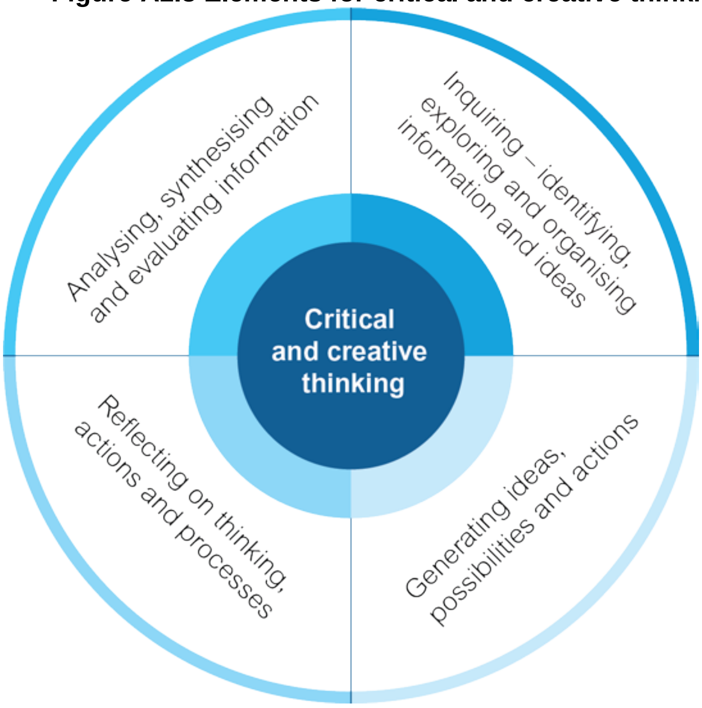
Critical and creative thinking

The critical and creative thinking continuum is described in the Australian Curriculum as consisting of two interrelated aspects:

> *Critical thinking is at the core of most intellectual activity that involves students in learning to recognise or develop an argument, use evidence in support of that argument, draw reasoned conclusions, and use information to solve problems. Examples of thinking skills are interpreting, analysing, evaluating, explaining, sequencing, reasoning, comparing, questioning, inferring, hypothesising, appraising, testing and generalising.*
>
> *Creative thinking involves students in learning to generate and apply new ideas in specific contexts, seeing existing situations in a new way, identifying alternative explanations, and seeing or making new links that generate a positive outcome.*
>
> *This includes combining parts to form something original, sifting and refining ideas to discover possibilities, constructing theories and objects, and acting on intuition.*

The critical and creative thinking continuum organises these concepts into four elements.

Figure A1.3 Elements for critical and creative thinking



These elements are further sub-divided to create 12 sub-elements (see Table A1.4).

Table A1.4 Organising elements and sub-elements of critical and creative thinking

Organising element	Sub-element
Inquiring – identifying, exploring and organising information and ideas	Pose questions
	Identify and clarify information and ideas
	Organise and process information
Generating ideas, possibilities and actions	Imagine possibilities and connect ideas
	Consider alternatives
	Seek solutions and put ideas into action
Reflecting on thinking, actions and processes	Think about thinking (metacognition)
	Reflect on processes
	Transfer knowledge into new contexts
Analysing, synthesising and evaluating information	Apply logic and reasoning
	Draw conclusions and design a course of action
	Evaluate procedures and outcomes

All elements/sub-elements of critical and creative thinking have a place within science education. However, only some of them can be overtly assessed within the parameters of the NAP—SL construct.

Different aspects of critical and creative thinking play different roles within an assessment. For example, many items within previous NAP—SL tests have included strong elements of logical reasoning. Other items targeted the evaluation of procedures and outcomes. More generally, the use of complex and rich stimulus within the NAP—SL assessments results in nearly all item sets addressing, at varying degrees of sophistication, a student's ability to transfer knowledge into new contexts.

When appropriate, the NAP—SL items were tagged against relevant statements (capability descriptions) from this continuum. This was done when there was a reasonable inference that students who answer the item correctly will have engaged in the relevant cognitive skill. Language from the continuum also helped inform item descriptors.

Items were not written to specifically address critical and creative thinking. Instead, the tagging of items against this capability will enable further discussion and analysis of how NAP—SL addresses this capability.

Online testing

Previous cycles of NAP—SL have been delivered using a printed document (otherwise known as paper-based testing). In 2015, NAP—SL was delivered as a computer-based assessment.

The item types available for 2015 NAP—SL included, but were not limited to, those in the following table (Table A1.5).

Table A1.5 Item types available for 2015 NAP—SL

Main type	Description	Paper-based equivalent
Multiple choice/choices	Select one option from radio buttons	Multiple choice
	Select one or more options from check boxes	Multiple select
Extended text/Text entry	Type text (manually scored) into box as separate paragraph	Short constructed response Long constructed response
Interactive gap match	Drag from a set of source options (text or images) into blank spaces within a passage of text or into a table	Not applicable
Match	Select which source objects match which destination categories by clicking a grid of radio buttons or checking checkboxes	Used for survey Likert style responses in the survey items
Hotspot	Select one or more predefined areas (circle, rectangle and polygon) on an image	Typically equivalent to multiple choice or multiple select with graphical options but with great freedom of layout and formatting
Interactive graphic gap match/ Position object	Drag objects (images/text) to hotspots on an image	Not applicable – but in some cases may be equivalent to multiple choice
Select point	Select a point on a background image	Used to simulate plotting points on a grid or graph
Composite	A combined item type that allows students to select an option and then write an explanation of their choice	This item type simulates a style of item used in past NAP—SL cycles, where students could choose from two alternatives and then explain their choice

Aside from the extended text item type, these item types can be more generally classified as forms of selected response. However, unlike more traditional multiple-choice items, students can be presented with a much greater range of choices. Practice items were included in the assessment to make students familiar with the various item types used.

To help maintain a consistent construct with past NAP—SL cycles, a similar proportion of long constructed response items (using the extended text item type) were included. These items required students to give longer, more open-ended responses that were then marked by expert markers in a similar way to the marking of long constructed response items in previous cycles.

Science inquiry and modifications to practical assessment tasks

Previous cycles of NAP—SL have included a practical component. The purpose of this component was to provide students with an opportunity to experience practical aspects of science within a formal assessment and test the conventions of science literacy in more depth than was possible in the objective component.

The practical component was not intended to be an assessment of a distinctly separate construct of 'practical science' or to provide a sub-scale measure of practical skills which might provide complementary information to a more general science scale. Instead, the test items in the practical section of NAP—SL were intended to be part of the same science literacy scale. This meant that at the item development, the piloting and the post-trial levels, practical items were judged against their performance with the objective items.

The approach for previous NAP—SL practical tasks was to use a two-stage structure. In the first stage, students participated in a science practical group work task that was classroom-based. Students then individually answered a range of items on the practical task they had just completed. These items included assessments of skills related to completing an experiment, including data representation tasks, writing conclusions and evaluating aspects of an experiment.

The 2015 NAP—SL was delivered online. Consequently, the previous approach for the NAP—SL practical tasks (hands-on investigations carried out by a group of students followed by individual responses to items) was no longer viable. At the same time, it was desirable that within online delivery, the 2015 NAP—SL content be as comparable as possible with previous NAP—SL cycles to allow comparisons of science literacy performance over time. Therefore, the 2015 NAP—SL contained a component that achieved similar objectives as the previous NAP—SL practical, but by different means.

A new approach for the NAP—SL online 'practical' task was developed that:

- highlighted the value of practical work within the science curriculum and its role in science literacy
- delivered valid and reliable data in the same way as the other component of the NAP—SL test
- addressed the same skill areas as the practical items in past NAP—SL cycles
- assessed/aligned with the relevant skills in the historical NAP—SL progress map
- assessed/aligned with the relevant skills in the new Australian Curriculum: science
- used items that could be written within the subset of item types available in the prescribed item authoring and review system
- was compatible with the test delivery system and the minimum technology requirements for participating schools (for example, screen size, appropriate file types, bandwidth limitations, etc.).

Linear staged task

An approach was developed after evaluation of different approaches to online tasks, taking into account the capabilities of the 2015 NAP—SL item-authoring and test-delivery systems and the available technology in schools.

The tasks presented were essentially closed, although the approach included some constructed-response items. Students were placed in a role more like observers of a practical task rather than active participants. This meant that students were not directly engaged in a practical activity but were tested on a range of relevant science inquiry skills. The approach was enhanced with video stimuli. To be sufficiently engaging, tasks with strong visual components were expected to be most effective.

To prevent dependencies, navigation needed to be more restricted than it was in the objective test. For example, if students were asked to make a prediction early in the task they needed to be prevented from changing that prediction once they had seen either the final results or the 'prediction' of an on-screen character.

It is hoped that this approach would provide good examples for primary schools of online assessment tasks in the science inquiry skills strand.

Stimuli and items embedded in the online tasks (and more generally in the NAP—SL assessments) use the term 'investigation' to refer to the process of answering a question, exploring an idea or solving a problem that requires activities such as planning a course of action, collecting data, interpreting data, reaching a conclusion and communicating these activities. Students were not expected to know that an experiment is an investigation which tests a hypothesis, and the term 'experiment' was not used in item stimulus or items. This is consistent with the treatment of the concepts 'investigation' and 'experiment' in the Australian Curriculum, where the term 'experiment' first appears in Year 7.

The NAP—SL progress map

During the previous cycles of NAP—SL, a science literacy progress map was developed based on the construct of science literacy and an analysis of state and territory curriculum and assessment frameworks. The progress map describes the development of science literacy across three strands:

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

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

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

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

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

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

Science literacy has been described in three strands to facilitate the interpretation of student responses to assessment tasks. However, authentic tasks should require students to apply concepts and processes together to address problems set in real-world contexts. The NAP—SL progress map (see Table A1.6) describes progression in four levels from 1 to 4 in terms of three aspects:

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

Strand C has been abstracted and makes no reference to particular science concepts or contexts.

NAP—SL focuses on Levels 2, 3 and 4 of the progress map; the levels of science literacy attained by students in Year 6. The agreed proficiency levels serve to further elaborate the progress map.

Table A1.6 NAP—SL progress map Levels 1 to 4

Level	Strands of science literacy		
	Strand A Formulating or identifying investigable questions and hypotheses; planning investigations; and collecting evidence. Process strand: experimental design and data gathering.	**Strand B** Interpreting evidence and drawing conclusions from students' own or others' data; critiquing the trustworthiness of evidence and claims made by others; and communicating findings. Process strand: interpreting experimental data.	**Strand C** Using science understandings for describing and explaining natural phenomena; and for interpreting reports about phenomena. Conceptual strand: applies conceptual understanding.
4	Formulates scientific questions, identifies the variable to be changed, the variable to be measured and in addition identifies at least one variable to be controlled. Uses repeated trials or replicates. Collects and records data involving two or more variables.	Calculates averages from repeat trials or replicates, plots line graphs where appropriate. Interprets data from line graph or bar graph. Conclusions summarise and explain the patterns in the science data. Makes general suggestions for improving an investigation (e.g. make more measurements).	Explains interactions, processes or effects that have been experienced or reported, in terms of a non-observable property or abstract science concept.
3 [includes 3.1, 3.2, and 3.3]	Formulates simple scientific questions for testing and makes predictions. Demonstrates awareness of the need for fair testing and appreciates scientific meaning of 'fair testing'. Identifies variable to be changed and/or measured but does not indicate variables to be controlled. Makes simple standard measurements. Records data as tables, diagrams or descriptions.	Displays data as tables or constructs bar graphs when given the variables for each axis. Identifies and summarises patterns in science data in the form of a rule. Recognises the need for improvement to the method. Applies the rule by extrapolating and predicting.	Describes the relationships between individual events (including cause and effect relationships) that have been experienced or reported. Generalises and applies the rule by predicting future events.

Level	Strands of science literacy		
	Strand A Formulating or identifying investigable questions and hypotheses; planning investigations; and collecting evidence. Process strand: experimental design and data gathering.	**Strand B** Interpreting evidence and drawing conclusions from students' own or others' data; critiquing the trustworthiness of evidence and claims made by others; and communicating findings. Process strand: interpreting experimental data.	**Strand C** Using science understandings for describing and explaining natural phenomena; and for interpreting reports about phenomena. Conceptual strand: applies conceptual understanding.
2	Given a question in a familiar context, identifies that one variable/factor is to be changed (but does not necessarily use the term 'variable' to describe the changed variable). Demonstrates intuitive level of awareness of fair testing. Observes and describes or makes non-standard measurements and limited records of data.	Makes comparisons between objects or events observed. Compares aspects of data in a simple supplied table of results. Completes simple tables and bar graphs given table column headings or prepared graph axes.	Describes changes to, differences between or properties of objects or events that have been experienced or reported.
1	Responds to the teacher's questions and suggestions, manipulates materials and observes what happens.	Shares observations; tells, acts out or draws what happened. Focuses on one aspect of the data.	Describes (or recognises) one aspect or property of an individual object or event that has been experienced or reported.

The NAP—SL major scientific concept areas

Table A1.7 NAP—SL major scientific concept areas

Major scientific concept areas	Examples	Australian Curriculum: science links
Earth and space		
Earth, sky and people: Our lives depend on air, water and materials from the ground; the ways we live depend on landscape, weather and climate.	Features of weather, soil and sky, and effects on me. People use resources from the Earth; need to use them wisely. Sustainability**	Daily and seasonal changes in our environment, including the weather, affect everyday life (F, ACSSU004). Earth's resources, including water, are used in a variety of ways (Year 2, ACSSU032). Earth's surface changes over time as a result of natural processes and human activity (Year 4, ACSSU075).
The changing Earth: The Earth is composed of materials that are altered by forces within and on its surface.	Changes in weather, weather data, seasons, soil landscape and sky (e.g. moon phases, weathering and erosion, movement of the sun and shadows, bush fires, land clearing). Climate change.*	Observable changes occur in the sky and landscape (Year 1, ACSSU019). Earth's surface changes over time as a result of natural processes and human activity (Year 4, ACSSU075). Light from a source forms shadows and can be absorbed, reflected and refracted (Year 5, ACSSU080). Sudden geological changes or extreme weather conditions can affect Earth's surface (Year 6, ACSSU096). Global systems, including the carbon cycle, rely on interactions involving the biosphere, lithosphere, hydrosphere and atmosphere (Year 10, ACSSU189).*
Our place in space: The Earth and life on Earth are part of an immense system called the universe.	Rotation of the Earth and night/day, spatial relationships between sun, Earth and moon. Planets of our solar system and their characteristics. Space exploration and new developments**	Observable changes occur in the sky and landscape (Year 1, ACSSU019). Earth's rotation on its axis causes regular changes, including night and day (Year 3, ACSSU048). The Earth is part of a system of planets orbiting around a star (the Sun) (Year 5, ACSSU078).

***Examples that cannot be explicitly found in the Australian Curriculum: science.*

**Examples that can be found in the Australian Curriculum: science in Years 7–10.*

Major scientific concept areas	**Examples**	**Australian Curriculum: science links**
Energy and force		
Energy and us: Energy is vital to our existence and our quality of life as individuals and as a society.	Uses of energy, patterns of energy use and variations with time of day and season. Energy sources, renewable and non-renewable.*	Energy from a variety of sources can be used to generate electricity (Year 6, ACSSU219). Some of Earth's resources are renewable, but others are non-renewable (Year 7, ACSSU116).*
Transferring energy: Interaction and change involve energy transfers; control of energy transfer enables particular changes to be achieved.	Sources, transfers, carriers and receivers of energy, energy and change. Types of energy, energy of motion – toys and other simple machines – light, sound.	Light and sound are produced by a range of sources and can be sensed (Year 1, ACSSU020). Heat can be produced in many ways and can move from one object to another (Year 3, ACSSU049). Light from a source forms shadows and can be absorbed, reflected and refracted (Year 5, ACSSU080). Electrical circuits provide a means of transferring and transforming electricity (Year 6, ACSSU097).
Energy sources and receivers: Observed change in an object or system is indicated by the form and amount of energy transferred to or from it.	Forces as pushes and pulls, magnetic attraction and repulsion.	A push or a pull affects how an object moves or changes shape (Year 2, ACSSU033). Heat can be produced in many ways and can move from one object to another (Year 3, ACSSU049). Forces can be exerted by one object on another through direct contact or from a distance (Year 4, ACSSU076).

**Examples that can be found in the Australian Curriculum: science in Years 7–10.*

Major scientific concept areas	Examples	Australian Curriculum: science links
Living things		
Living together: Organisms in a particular environment are interdependent.	Living vs. non-living. Plant vs. animal and major groups. Dependence on the environment: Survival needs – food, space and shelter. Interactions between organisms and interdependence (e.g. simple food chains).	Living things have basic needs, including food and water (F, ACSSU002). Living things live in different places where their needs are met (Year 1, ACSSU211). Living things grow, change and have offspring similar to themselves (Year 2, ACSSU030). Living things can be grouped on the basis of observable features and can be distinguished from non-living things (Year 3, ACSSU044). Living things, including plants and animals, depend on each other and the environment to survive (Year 4, ACSSU073).
Structure and function: Living things can be understood in terms of functional units and systems.	Major structures and systems and their functions. Healthy lifestyle, diet and exercise.**	Living things have a variety of external features (Year 1, ACSSU017). Living things have structural features and adaptations that help them to survive in their environment (Year 5, ACSSU043).
Biodiversity, change and continuity: Life on Earth has a history of change and disruption, yet continues generation to generation.	Change over lifetime, reproduction and lifecycles. Adaptation to physical environment.	Living things have life cycles (Year 4, ACSSU072). Living things have structural features and adaptations that help them to survive in their environment (Year 5, ACSSU043). The growth and survival of living things are affected by the physical conditions of their environment (Year 6, ACSSU094).

***Examples that cannot be explicitly found in the Australian Curriculum: science.*

Major scientific concept areas	**Examples**	**Australian Curriculum: science links**
Matter		
Materials and their uses: The properties of materials determine their uses; properties can be modified.	Materials have different properties and uses. Processing materials to make useful things produces waste. Use of alternative materials to better care for the environment.** Waste reduction – recycling. Nanotechnology.**	Objects are made of materials that have observable properties (Year F, ACSSU003). Everyday materials can be physically changed in a variety of ways (Year 1, ACSSU018). Different materials can be combined, including by mixing, for a particular purpose (Year 2, ACSSU031). Natural and processed materials have a range of physical properties; these properties can influence their use (Year 4, ACSSU074).
Structure and properties: The substructure of materials determines their behaviour and properties.	The properties of materials can be explained in terms of their visible substructure, such as fibres.*	The properties of the different states of matter can be explained in terms of the motion and arrangement of particles (Year 8, ACSSU151).* Differences between elements, compounds and mixtures can be described at a particle level (Year 8, ACSSU152).*
Reactions and change: Patterns of interaction of materials enable us to understand and control those interactions.	Materials can change their state and properties. Solids, liquids and gases.	A change of state between solid and liquid can be caused by adding or removing heat (Year 3, ACSSU046). Solids, liquids and gases have different observable properties and behave in different ways (Year 5, ACSSU077). Changes to materials can be reversible, such as melting, freezing, evaporating; or irreversible, such as burning and rusting (Year 6, ACSSU095).

***Examples that cannot be explicitly found in the Australian Curriculum: science.*

**Examples that can be found in the Australian Curriculum: science in Years 7–10.*

APPENDIX 2 SAMPLE SCHOOL REPORTS

Figure A2.1 Student report - introduction text

acara AUSTRALIAN CURRICULUM, ASSESSMENT AND REPORTING AUTHORITY

NAP NATIONAL ASSESSMENT PROGRAM

2015 National Assessment Program – Science Literacy

Information Sheet: Interpreting Student Reports

Each participating Year 6 student completed one of the seven objective tests and one of the two inquiry tasks. The report provides information about each student's achievement on the particular objective test and inquiry task that s/he completed. Each item (question) appeared in three of the seven objective tests in a different position. So, although each objective test was different, there were commonalities between the tests. Each test comprised a different combination of three clusters of questions and only one third of the questions in any test appeared in two other tests. Therefore, the total score achieved by any one student can only be compared with that achieved by other students who completed the same test.

For each test, the report provides the following information:

1. the relevant Australian Curriculum code addressed by the question (please refer to the link at the bottom of each page for more information);
2. a description of the aspect of the curriculum tested by the question;
3. the maximum possible score for the question and the percentage of students in your school (across multiple tests) who achieved that score;
4. the percentage of students in the sample who achieved the maximum score on the question;
5. the name of each student who completed the corresponding test, his/her achievement on the question and overall score on the test.

These reports can be used to:

- compare your students' achievement on each question against the sample (by comparing the two columns showing the percentage of students who achieved the maximum score)
- compare student achievement within the seven objective tests and two inquiry tasks by looking at the maximum possible score and the total score achieved by each student for each test
- identify areas in the curriculum that may need to be covered in more detail by examining the performance of students in each Australian Curriculum code.

Please note: the NAP–SL Assessment Framework is not yet fully aligned to the Australian Curriculum but questions have been tagged in the report with the Australian Curriculum codes.

Below is part of a sample report form with some key information explained.

Figure A2.2 Student report - explanation diagram

90% of students in the sample achieved the maximum score for this question.

This student achieved the maximum score (2) for this question.

These students completed Test 2.

2015 National Assessment Program – Science Literacy

SCHOOL NAME

Year 6 Objective Test 2

	(1)	(2)	(3)		(4)	(5)		
Q No.	Australian Curriculum code*	Item descriptor	Maximum question score	% Students with maximum score (your school)	% Students with maximum score (sample)	STUDENT 1	STUDENT 2	STUDENT 3
01	ACSSU094	Interprets data on the growth and survival of living things	1	100	95	1	1	1
02	ACSSU073	Shows awareness that plants and animals depend on each other to survive	1	65	75	1	1	1
03	ACSIS232	Identifies the research question of an investigation based on a description of its method	1	95	90	1	1	0
04	ACSSU077	Draws a conclusion about the observed properties of liquids in an investigation	1	75	85	1	1	1
05	ACSIS104	Explains why variables should be changed and measured in fair tests	1	65	70	-	1	1
06	ACSIS104	Evaluates which variables should be changed and measured in a fair test	2	65	80	1	2	0
		Maximum score possible	7	Total score		5	7	4

75% of students at your school achieved the maximum score for this question.

This student did not attempt this question.

This student attempted this question and achieved a score of 0.

Figure A2.3 Student report - inquiry task sample

2015 National Assessment Program – Science Literacy

Year 6 Inquiry Task 1

Q No.	Australian Curriculum code*	Item descriptor	Maximum question score	% Students with maximum score (your school)	% Students with maximum score (sample)												
01	ACSIS071	Represents ideas with a diagram	1	100	90	1	1	1	1	1	1	1	1	1	1	1	1
02	ACSIS090	Compares and describes observations of moving objects	1	100	79	1	1	1	1	1	1	1	1	1	1	1	1
03	ACSIS086	Suggests ways to conduct an investigation to find answers to a question	1	0	12	0	0	0	0	0	0	0	-	0	0	0	0
04	ACSIS087	Identifies the variables used in an investigation	1	75	43	1	1	1	1	1	0	1	0	1	1	1	0
05	ACSIS090	Infers the time taken for a process based on given data	1	92	62	1	1	1	1	1	1	1	1	1	1	0	1
06	ACSIS087	Measures and records data, using digital technologies as appropriate	1	100	72	1	1	1	1	1	1	1	1	1	1	1	1
07	ACSIS087	Explains why multiple trials are necessary in fair tests	1	8	6	0	0	1	0	0	0	0	-	0	0	0	0
08	ACSIS107	Describes patterns or relationships in data	2	0	0	0	0	1	0	0	0	0	0	0	1	0	1
09	ACSIS107	Constructs graphs to represent relationships in data using digital technologies	1	83	48	1	1	1	1	1	1	1	0	1	1	0	1
10	ACSIS103	Plans appropriate investigation methods to answer questions	2	8	3	1	0	0	0	1	0	2	-	0	1	0	0
		Maximum score possible	**12**	**Total score**		**7**	**6**	**8**	**6**	**7**	**5**	**8**	**4**	**6**	**8**	**4**	**6**

APPENDIX 3 SAMPLING

Technical notes

Stratification details

For each jurisdiction, schools were separated into three separate strata according to their size: very small; moderately small; and large. The target proportion of students and number of schools selected within each of the strata were determined using the OECD (2005) treatment of small schools (pp. 53–56). Essentially, the aim was to balance selecting an adequate sample without substantially increasing the number of sampled schools.

Large schools within each jurisdiction were further separated according to their school sector. The target numbers of large schools were proportionally allocated amongst the school sectors for each jurisdiction. Very small and moderately small strata were sorted according to school sector, then by the remaining implicit stratification variables – NAPLAN quintile, geographic location and measure of size (MOS). This strategy meant that the sampling frame was divided into 40 explicit strata overall. That is, there were 24 strata containing large schools (8 jurisdictions × 3 sectors); eight moderately small school strata (1 per jurisdiction); and eight very small school strata (1 per jurisdiction).

The stratification for small schools was slightly more complex than for large schools. Small schools were ordered by sector, NAPLAN quintile, geographic location and then MOS. The sort order was alternated so that ‘like schools’ were always nearby.

The stratum was sorted first by sector. Within each sector, schools were further sorted by NAPLAN quintile. This sort order was alternated between ascending to descending between sectors (that is, sector 1 had NAPLAN quintile sorted ascending, sector 2 had NAPLAN quintile sorted descending, sector 3 had NAPLAN quintile sorted ascending). Similarly, within each NAPLAN quintile category, schools were further sorted by geographic location. This sort order was alternated between ascending to descending between fields (that is, NAPLAN quintile 1 had geographic location sorted ascending, NAPLAN quintile 2 had geographic location sorted descending, NAPLAN quintile 3 had geographic location sorted ascending, etc.). The sort order for MOS was then alternated from low to high, then high to low, each time a new sector / NAPLAN quintile / geographic location classification was encountered. Table A3.1 illustrates the sort-order procedures that were employed for small Catholic schools.

Table A3.1 The sort ordering procedures employed for small Catholic schools

Sector	NAPLAN quintile	Geographic location	MOS
1	1	1	A
1	1	2	D
1	1	3	A
1	2	3	D
1	2	2	A
1	2	1	D
1	3	1	A
1	3	2	D
1	3	3	A
1	4	3	D
1	4	2	A
1	4	1	D
1	5	1	A
1	5	2	D
1	5	3	A

After small schools were stratified, the MOS for each school in the stratum was set equal to the average enrolment size (ENR) of all schools within that particular stratum. This was equivalent to selecting a simple random sample of small schools. Such a strategy meant that very small schools would not be assigned excessively large sampling weights.

Random start and sampling interval values

The sampling interval ([stratum enrolment size]/[planned number of schools]) is rounded to the nearest integer. Table A3.2 shows the starting values used to draw the sample for each explicit stratum.

Table A3.2 Stratum variables for sample selection

Stratum	Stratum enrolment size	Planned number of schools	Sampling interval	Random start
NSW_Large_C	14592	14	1042	98
NSW_Large_G	51672	51	1013	104
NSW_Large_I	10211	10	1021	366
NSW_Modsmall	6780	9	753	252
NSW_VerySmall	3417	8	427	56
VIC__Large_C	11976	16	749	486
VIC__Large_G	37441	49	764	616
VIC__Large_I	7361	10	736	291
VIC__ModSmall	6068	10	607	107
VIC__VerySmall	2604	7	372	62
QLD_Large_C	9575	14	684	579
QLD_Large_G	35677	53	673	210
QLD_Large_I	6432	10	643	134

Stratum	Stratum enrolment size	Planned number of schools	Sampling interval	Random start
QLD_Modsmall	3324	7	475	86
QLD_VerySmall	2204	7	315	130
WA__Large_C	4247	12	354	320
WA__Large_G	17766	52	342	325
WA__Large_I	3807	11	346	327
WA__ModSmall	2429	10	243	59
WA__VerySmall	1235	8	154	107
SA_Large_C	3024	14	216	172
SA_Large_G	9405	43	219	158
SA_Large_I	3060	14	219	107
SA_ModSmall	2445	15	163	124
SA_VerySmall	1020	9	113	81
TAS__Large_C	950	9	106	69
TAS__Large_G	3544	33	107	61
TAS__Large_I	557	5	111	65
TAS__ModSmall	676	8	85	77
TAS__VerySmall	493	7	70	48
ACT_Large_C	910	11	83	58
ACT_Large_G	2433	30	81	59
ACT_Large_I	842	11	77	50
ACT_Modsmall	269	4	67	46
ACT_verysmall	53	1	53	2
NT_Large_C	359	4	90	83
NT_Large_G	1735	21	83	50
NT_Large_I	325	4	81	48
NT_Modsmall	426	7	61	23
NT_VerySmall	456	12	38	15

Characteristics of the 2015 sample

The procedures used to draw the 2015 sample of schools were nearly identical to those used in the 2009 and 2012 assessments. Table A3.3 shows the number of sampled students. The table shows that the percentage of the students sampled from ACT, NT and Tas. is smaller compared with other jurisdictions. While equal samples are in principle desired from each jurisdiction, the sample sizes were reduced for the ACT, NT and Tas. given their relatively smaller populations and smaller number of schools.

Table A3.3 2015 NAP—SL target and achieved sample sizes by jurisdiction

State/Territory	Number of selected students enrolled at the time of testing		Number of selected students who participated in the test	
	Students	Percentage of the sample	Students	Percentage of the sample
NSW	2185	15.1	1911	15.4
Vic.	2162	14.9	1930	15.6

Qld	2177	15.0	1833	14.8
WA	2126	14.7	1878	15.1
SA	2178	15.0	1790	14.4
Tas.	1366	9.4	1198	9.7
ACT	1366	9.4	1221	9.8
NT	920	6.4	649	5.2
Aust.	**14480**	**100.0**	**12410**	**100.0**

Smaller samples from the ACT, NT and Tas. were necessary to ensure that schools in smaller states and territories are not over-burdened with survey assessments.

Table A3.4 shows the proportion of students in each sector by jurisdiction for both the selected sample and the population according to the sample frame. The table shows that the difference between the selected sample and the population is generally less than 2 per cent. This indicates that the proportion of students in the selected sample closely matches the population when comparing sector by sector within a jurisdiction.

Table A3.4 Comparison of selected sample and population sector proportions across jurisdictions

State or Territory	Sector	Population			Selected sample			Difference (population minus sample) of percentages
		Schools	Students	Percentage of students	Schools	Students	Percentage of students	
NSW	C	431	16777	19	18	365	19	0.2
	G	1627	58633	67	61	1283	67	0.2
	I	310	11677	13	12	263	14	-0.4
	Total	**2368**	**87087**	**100**	**91**	**1911**	**100**	**0.0**
Vic.	C	394	14182	22	20	444	23	-1.3
	G	1206	42932	66	60	1228	64	2.0
	I	216	8282	13	11	257	13	-0.7
	Total	**1816**	**65396**	**100**	**91**	**1929**	**100**	**0.0**
Qld	C	223	10343	18	17	364	20	-1.8
	G	973	39871	69	61	1218	66	3.0
	I	168	7164	12	11	251	14	-1.2
	Total	**1364**	**57378**	**100**	**89**	**1833**	**100**	**0.0**
WA	C	130	4829	16	15	344	18	-2.0
	G	595	20310	69	59	1256	67	1.9
	I	133	4365	15	14	276	15	0.1
	Total	**858**	**29504**	**100**	**88**	**1876**	**100**	**0.0**
SA	C	83	3407	18	16	357	20	-1.9
	G	428	11988	63	55	1093	61	2.4
	I	89	3493	18	15	340	19	-0.5
	Total	**600**	**18888**	**100**	**86**	**1790**	**100**	**0.0**
Tas.	C	32	1107	18	10	210	18	0.2
	G	148	4336	69	42	844	70	-1.0

	I	34	796	13	8	144	12	0.7
	Total	**214**	**6239**	**100**	**60**	**1198**	**100**	**0.0**
ACT	C	23	1004	23	12	291	24	-0.4
	G	61	2549	59	32	709	58	1.4
	I	15	737	17	11	221	18	-0.9
	Total	**99**	**4290**	**100**	**55**	**1221**	**100**	**0.0**
NT	C	12	403	12	5	91	14	-1.8
	G	130	2503	76	29	469	72	3.8
	I	16	383	12	5	89	14	-2.1
	Total	**158**	**3289**	**100**	**39**	**649**	**100**	**0.0**
Aust.	C	1328	52052	19	113	2466	20	-0.7
	G	5168	183122	67	399	8100	65	2.0
	I	981	36897	14	87	1841	15	-1.3
	Total	**7477**	**272071**	**100**	**599**	**12407**	**100**	**0.0**

Schools were also classified according to their enrolment size. Small schools (i.e. *moderately small* and *very small* schools) were under-sampled and large schools were slightly over-sampled. This approach was adopted to ensure that an adequate number of students would be assessed, while still ensuring very small schools would be represented without vastly increasing the overall number of schools sampled. *Very small* schools were under-sampled to a larger degree than *moderately small* schools. Table A3.5 shows the number of schools according to school size for the population and the selected sample. Table A3.5 also shows the percentage of students in the population compared to the selected sample according to school size. When considered in terms of the number of students, the under-sampling of small schools is not as noticeable. For example, 4.3 per cent of the population attend a *very small* school which is very similar to the 3.1 per cent of students from *very small* schools included in the selected sample.

Table A3.5 Comparison of population and selected sample proportions according to school size

State or Territory	School size	Population				Selected sample				Difference (population minus sample) of percentages
		Schools	Percentage of schools	Students	Percentage of students	Schools	Percentage of schools	Students	Percentage of students	
NSW	Large	1388	59	76475	88	74	81	1693	89	-1
	Mod Small	364	15	6916	8	9	10	147	8	0
	Very Small	616	26	3696	4	8	9	71	4	1
	Total	2368	100	87087	100	91	100	1911	100	0
Vic.	Large	1073	59	56778	87	75	82	1733	90	-3
	Mod Small	320	18	6080	9	10	11	175	9	0

	Very Small	423	23	2538	4	6	7	22	1	3
	Total	1816	100	65396	100	91	100	1930	100	0
Qld	Large	789	58	51684	90	75	84	1676	91	-1
	Mod Small	187	14	3366	6	7	8	106	6	0
	Very Small	388	28	2328	4	7	8	51	3	1
	Total	1364	100	57378	100	89	100	1833	100	0
WA	Large	512	60	25820	88	74	84	1700	91	-3
	Mod Small	134	16	2412	8	8	9	116	6	2
	Very Small	212	25	1272	4	6	7	62	3	1
	Total	858	100	29504	100	88	100	1878	100	0
SA	Large	313	52	15489	82	67	78	1487	83	-1
	Mod Small	129	22	2451	13	14	16	277	15	-2
	Very Small	158	26	948	5	5	6	26	1	4
	Total	600	100	18888	100	86	100	1790	100	0
Tas.	Large	115	54	5051	81	47	78	1071	89	-8
	Mod Small	36	17	684	11	7	12	100	8	3
	Very Small	63	29	504	8	6	10	27	2	6
	Total	214	100	6239	100	60	100	1198	100	0
ACT	Large	77	78	3968	92	49	89	1126	92	0
	Mod Small	14	14	266	6	4	7	84	7	-1
	Very Small	8	8	56	1	2	4	11	1	0
	Total	99	100	4290	100	55	100	1221	100	0
NT	Large	56	35	2391	73	26	67	529	82	-9
	Mod Small	22	14	418	13	6	15	79	12	1
	Very Small	80	51	480	15	7	18	41	6	8
	Total	158	100	3289	100	39	100	649	100	0
Aust.	Large	4323	58	237656	87	487	81	11015	89	-1
	Mod Small	1206	16	22593	8	65	11	1084	9	0
	Very Small	1948	26	11822	4	47	8	311	3	2
	Total	7477	100	272071	100	599	100	12410	100	0

Table A3.6 provides a breakdown of the sample at the school level according to jurisdiction.

Table A3.6 School participation rates by jurisdiction

State/Territory	Number of schools sampled	Number of schools that participated	School participation (per cent)
NSW	93	91	97.8
Vic.	92	91	98.9
Qld	91	89	97.8
WA	92	88	95.7
SA	94	86	91.5
Tas.	62	60	96.8
ACT	55	55	100.0
NT	49	39	79.6
Aust.	**628**	**599**	**95.4**

Table A3.7 provides a breakdown of the achieved sample in comparison with the number of Year 6 students in each jurisdiction.

Table A3.7 Achieved sample by student participation

State/ Territory	Student population	Number of selected students	Number of students who participated	Within school exclusions	Within school exclusions (per cent)	Within-school student participation (per cent)
NSW	87087	2185	1911	23	1.1	87.5
Vic.	65396	2162	1930	19	0.9	89.3
Qld	57378	2177	1833	19	0.9	84.2
WA	29504	2126	1878	18	0.8	88.3
SA	18888	2178	1790	49	2.2	82.2
Tas.	6239	1366	1198	28	2.0	87.7
ACT	4290	1366	1221	28	2.0	89.4
NT	3289	920	649	10	1.1	70.5
Aust.	**272071**	**14480**	**12410**	**194**	**1.3**	**85.7**

Table A3.8 provides a breakdown of the achieved sample across states and territories according to gender, Indigenous status, students' language background and school geographic location.

Table A3.8 Percentage distribution of Year 6 sample characteristics by jurisdiction

	State/Territory (per cent)								Aust. (per cent)
	NSW	Vic.	Qld	WA	SA	Tas.	ACT	NT	
Student gender									

Female	47.1	47.6	50.7	52.2	47.0	50.5	48.4	50.7	49.1
Male	52.9	52.4	49.3	47.8	52.3	49.5	51.6	49.3	50.8
Missing	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.1
Indigenous status									
Indigenous	5.2	1.7	8.4	3.7	3.4	7.3	2.6	29.1	5.9
Non-Indigenous	93.2	98.2	91.2	93.3	94.9	87.9	90.5	70.7	92.0
Missing	1.5	0.1	0.4	2.9	1.7	4.8	6.9	0.2	2.2
Language background									
English speaking background	71.0	74.6	86.6	61.8	80.4	87.8	64.3	43.0	73.3
Language background other than English	29.0	24.2	13.0	36.1	18.7	12.2	35.6	54.1	25.8
Missing	0.1	1.1	0.3	2.1	0.9	0.0	0.1	2.9	0.8
Geographic location									
Metropolitan areas	72.7	79.2	71.5	70.3	73.9	44.7	100.0	0.0	69.5
Provincial areas	27.3	20.8	24.8	23.8	23.5	54.8	0.0	59.8	26.5
Remote and very remote areas	0.1	0.0	3.8	6.0	2.7	0.5	0.0	40.2	4.0
Number of students	**1911**	**1930**	**1833**	**1878**	**1790**	**1198**	**1221**	**649**	**12410**

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

Table A3.9 provides a breakdown of the number of students in the achieved sample by Indigenous status across the three geographic location categories.

Table A3.9 Achieved sample size by Indigenous status and geographic location

Geographic location	Number of students by Indigenous status			Total
	Indigenous	Non-Indigenous	Missing	
Metropolitan areas	255	8141	229	8625
Provincial areas	294	2959	35	3288
Remote and very remote areas	178	316	3	497
Aust.	**727**	**11416**	**267**	**12410**

APPENDIX 4 PROGRAMMING NOTES ON SAMPLING

SPSS syntax for sample selection

```
*============================================

*============================================

        NAP—SL 2015 SAMPLE PROCEDURE

*============================================

*============================================.

*SPSS version 22.

*============================================

        PPS SAMPLE MACRO

*============================================.

*-----------------------------------------------------------------------------------

 This macro will select sample schools for a particular stratum

 The following arguments are required:

 ~~~enrsize is equal to average enrolment size for modsmall and verysmall strata

   otherwise, set enrsize equal to 999 for large school strata

 ~~~strata is the name of the current stratum

 ~~~randm is a random number

 ~~~const is the sampling interval

*-----------------------------------------------------------------------------------.

DEFINE !SAMPLE (enrsize = !DEFAULT(999) !TOKENS(1)

                / strata = !TOKENS(1)

                / randm = !TOKENS(1)

                / const = !TOKENS(1)).


DATASET CLOSE ALL.

GET FILE='SampleFrameRandomOrder.sav'.


*------------------------------------------------------------------.
```

```
*  EXPLICIT STRATIFICATION.

*-------------------------------------------------------------------.

select if (RTRIM(Stratum)=!strata).

exe.


*-------------------------------------------------------------------.

*  IMPLICIT STRATIFICATION.

*-------------------------------------------------------------------.

*all implicit stratification variables need to be numeric ordinal categories. sequential
numbering is not required using this methodology.

*~~~assign SectorId as the first implicit stratification variable (assumes numeric categories,
not necessarily in sequential order).

COMPUTE imp_0 = SectorId.

SORT CASES BY imp_0 (A).

RANK VARIABLES = imp_0 (A) /RANK /PRINT=YES /TIES=CONDENSE.


*~~~add NAPLANDATA as implicit stratification variable.

DO IF (MOD(Rimp_0,2) > 0).

  compute imp_1 = (Rimp_0*100) + (NAPLANDATA*-1).

ELSE.

  compute imp_1 = (Rimp_0*100) + NAPLANDATA.

END IF.

SORT CASES BY imp_1(A).

RANK VARIABLES = imp_1(A) /RANK /PRINT=YES /TIES=CONDENSE.


*~~~add GeoId as implicit stratification variable.

DO IF (MOD(Rimp_1,2) > 0).

  compute imp_2 = (Rimp_1*100) + (GeoId*-1).

ELSE.
```

```
  compute imp_2 = (Rimp_1*100) + GeoId.
END IF.
SORT CASES BY imp_2(A).
RANK VARIABLES = imp_2(A) /RANK /PRINT=YES /TIES=CONDENSE.

*~~~add gr06 as implicit stratification variable.
DO IF (MOD(Rimp_2,2) > 0).
  compute imp_3 = (Rimp_2*1000) + (gr06*-1).
ELSE.
  compute imp_3 = (Rimp_2*1000) + gr06.
END IF.
SORT CASES BY imp_3 (A).
RANK VARIABLES = imp_3 (A) /RANK /PRINT=YES /TIES=CONDENSE.

SORT CASES BY imp_0 (A) imp_1 (A) imp_2 (A) imp_3 (A).
*file is now implicitly stratified.

*-----------------------------------------------------------------.
*  MEASURE OF SIZE (MOS) ADJUSTMENTS.
*-----------------------------------------------------------------.
*=======SMALL SCHOOLS=======.
!IF (!enrsize = 999)!THEN.
  *do nothing.
!ELSE.
 * for small schools set MOS equal to avg enr size for the explicit stratum.
        compute tmpgr06 = gr06.
        compute gr06 = !enrsize.
!IFEND.
```

```
*=======SET VERY LARGE SCHOOLS EQUAL TO THE SAMPLING INTERVAL=======.
if (gr06>!const) gr06 = !const.
exe.

*--------------------------------------------------------------------------------------------------------------.
*  SELECT SCHOOLS WITH PROBABILITY PROPORITIONAL TO SIZE (PPS)
*--------------------------------------------------------------------------------------------------------------.
compute ranstart = !randm.
compute interval = !const.
compute case = $casenum.
exe.

if ($casenum = 1) ticket1 = 1.
if ($casenum = 1) ticket2 = gr06.
if ($casenum > 1) ticket1 = lag(ticket2) + 1.
if ($casenum > 1) ticket2 = lag(ticket2) + gr06.
if ($casenum = 1) selector = ranstart.
if ($casenum > 1) selector = lag(selector).
string select (a3).
compute select = '___'.
if (ticket1 <= selector and selector <= ticket2) select = 'YES'.
if (select = 'YES') selector = selector + interval.
*HANDLE FOR LARGE SCHOOLS.
if (select = 'YES' and selector < ticket2) select = 'SOS'.
exe.

if ($casenum = 1) wintickt=ranstart.
```

```
if ($casenum > 1) wintickt=lag(selector).

exe.


*=======SELECT REPLACEMENT SCHOOLS=======.

DO IF ((lag(select)='YES' or lag(select)='SOS') and select = '___').

        compute select = 'R_1'.

        compute replaceid = lag(schoolid).

END IF.

DO IF ((lag(select,2)='YES' or lag(select,2)='SOS') and select = '___' and
lag($casenum,2)=1).

        compute select = 'R_2'.

        compute replaceid = lag(schoolid,2).

END IF.

SORT CASES BY case (D) .

DO IF ((lag(select)='YES' or lag(select)='SOS') and select = '___').

        compute select = 'R_2'.

        compute replaceid = lag(schoolid).

END IF.

DO IF ((lag(select,2)='YES' or lag(select,2)='SOS') and select = '___' and
lag($casenum,2)=1).

        compute select = 'R_1'.

        compute replaceid = lag(schoolid,2).

END IF.

SORT CASES BY case (A) .

if (select = 'YES' or select = 'SOS') replaceid = schoolid.

exe.

SAVE OUTFILE=!QUOTE(!CONCAT('All_',!UNQUOTE(!strata) , '.sav')).


*=======KEEP SAMPLED AND REPLACEMENT SCHOOLS=======.
```

```
set width = 120.

set length = 1000.

title Schools Selected from the Specified Stratum !strata.

select if (select='YES' or select='SOS').

list var=SchoolName stratum gr06 ticket1 ticket2 wintickt select / format = numbered.

title.

SAVE OUTFILE=!QUOTE(!CONCAT('Sample_',!UNQUOTE(!strata) , '.sav')).


!ENDDEFINE.


*DRAW SAMPLES.

!SAMPLE strata='ACT_Large_C' const=83 randm=58 enrsize=999.

!SAMPLE strata='ACT_Large_G' const=81 randm=59 enrsize=999.

!SAMPLE strata='ACT_Large_I' const=77 randm=50 enrsize=999.

!SAMPLE strata='ACT_ModSmall' const=67 randm=46 enrsize=19.

!SAMPLE strata='ACT_VerySmall' const=53 randm=2 enrsize=7.

!SAMPLE strata='NSW_Large_C' const=1042 randm=98 enrsize=999.

!SAMPLE strata='NSW_Large_G' const=1013 randm=104 enrsize=999.

!SAMPLE strata='NSW_Large_I' const=1021 randm=366 enrsize=999.

!SAMPLE strata='NSW_ModSmall' const=753 randm=252 enrsize=19.

!SAMPLE strata='NSW_VerySmall' const=427 randm=56 enrsize=6.

!SAMPLE strata='NT_Large_C' const=90 randm=83 enrsize=999.

!SAMPLE strata='NT_Large_G' const=83 randm=50 enrsize=999.

!SAMPLE strata='NT_Large_I' const=81 randm=48 enrsize=999.

!SAMPLE strata='NT_ModSmall' const=61 randm=23 enrsize=19.

!SAMPLE strata='NT_VerySmall' const=38 randm=15 enrsize=6.

!SAMPLE strata='QLD_Large_C' const=684 randm=579 enrsize=999.

!SAMPLE strata='QLD_Large_G' const=673 randm=210 enrsize=999.
```

!SAMPLE strata='QLD_Large_I' const=643 randm=134 enrsize=999.

!SAMPLE strata='QLD_ModSmall' const=475 randm=86 enrsize=18.

!SAMPLE strata='QLD_VerySmall' const=315 randm=130 enrsize=6.

!SAMPLE strata='SA_Large_C' const=216 randm=172 enrsize=999.

!SAMPLE strata='SA_Large_G' const=219 randm=158 enrsize=999.

!SAMPLE strata='SA_Large_I' const=219 randm=107 enrsize=999.

!SAMPLE strata='SA_ModSmall' const=163 randm=124 enrsize=19.

!SAMPLE strata='SA_VerySmall' const=113 randm=81 enrsize=6.

!SAMPLE strata='TAS_Large_C' const=106 randm=69 enrsize=999.

!SAMPLE strata='TAS_Large_G' const=107 randm=61 enrsize=999.

!SAMPLE strata='TAS_Large_I' const=111 randm=65 enrsize=999.

!SAMPLE strata='TAS_ModSmall' const=85 randm=77 enrsize=19.

!SAMPLE strata='TAS_VerySmall' const=70 randm=48 enrsize=8.

!SAMPLE strata='VIC_Large_C' const=749 randm=486 enrsize=999.

!SAMPLE strata='VIC_Large_G' const=764 randm=616 enrsize=999.

!SAMPLE strata='VIC_Large_I' const=736 randm=291 enrsize=999.

!SAMPLE strata='VIC_ModSmall' const=607 randm=107 enrsize=19.

!SAMPLE strata='VIC_VerySmall' const=372 randm=62 enrsize=6.

!SAMPLE strata='WA_Large_C' const=354 randm=320 enrsize=999.

!SAMPLE strata='WA_Large_G' const=342 randm=325 enrsize=999.

!SAMPLE strata='WA_Large_I' const=346 randm=327 enrsize=999.

!SAMPLE strata='WA_ModSmall' const=243 randm=59 enrsize=18.

!SAMPLE strata='WA_VerySmall' const=154 randm=107 enrsize=6.

APPENDIX 5 EXTRACT FROM TEST ADMINISTRATION MANUAL (INCLUDING SUPPLYING STUDENT PARTICIPATION INFORMATION)

This is an extract of the test administration manual used by schools to administer 2015 NAP—SL. Phone numbers, URLs, page references and some images have been removed as they are no longer current.

1. Introduction

Thank you for your support in administering the National Assessment Program – Science Literacy (NAP–SL) online sample assessment in 2015. This Test Administration Manual provides you with all the instructions you will need to administer the test.

Sample assessments are conducted nationally every three years. Your school is one of approximately 630 schools selected to take part in this online sample assessment.

You can confirm your school's nominated test date, under the 'School Information' tab on the NAP–SL school portal.

Your role in the administration of the test is critical.

It is essential that you are very familiar with all the procedures in this manual **BEFORE** the test**.**

Standardised administration of the test contributes significantly to the fairness and reliability of the assessment.

Follow all instructions carefully.

It is important to maintain a positive and calm atmosphere throughout the test.

1.1 Test schedule overview

On test day, students will be required to complete the following components of the online assessment:

- Practice questions
- Objective test
- Inquiry task
- Student survey.

Each test session is to be conducted in the order noted above. The system will not allow students to participate in the tasks out of sequence.

It is expected that the assessment will take just over 2 hours to administer, including time to read the instructions. You will need to factor in enough time for students to get seated, log into your school's computer system and then log onto the test website. This time may vary from school to school**.**

Allow enough time for students to get seated and logged into your school's computer system. This time will vary from school to school and may take longer than you expect.	
Step 1 and 2: Practice questions & Objective test	Time allowed
Test Administrator reads instructions and students complete the practice questions	10 minutes
Students undertake the test	60 minutes
Ending the session	5 minutes
Approximate time:	1 hour 15 minutes
STUDENTS MUST LOG OUT OF THE NAP–SL TEST BEFORE TAKING A BREAK. Allow a break of approximately 15–20 minutes before starting the Inquiry task.	
Step 3: Inquiry task	Time allowed
Students read the instructions page	1 minute
Students undertake the test	35 minutes
Ending the session	5 minutes
Approximate time:	40 minutes
Step 4: Student survey	Time allowed
Students read the instructions page	1 minute
Students undertake the survey	10 minutes
Ending the session	5 minutes
Approximate time:	15 minutes
Total assessment session approximate time:	2 hours 10 minutes

The test delivery system automatically times each student's attempt in each section. It is recommended students start at the same time to assist with administering each session of the assessment.

2. Test administrator's responsibilities

2.1 Preparing for the online assessment

Your school has nominated you as the Test Administrator (TA) to administer the test. You will be required to:

Prior to test day

- review this manual
- ensure that headphones are available and in working order (either a full set or sufficient spares if students bring their own)
- ensure that the Technical Readiness Test (TRT) has been conducted prior to test day on all devices to be used (including spares)
- print the list of selected students with usernames and passwords, available on the NAP–SL school portal from 6 October
- cut the student list into strips to hand individual usernames and passwords to students on test day
- ensure that there are arrangements in place for non-participating students
- ensure you are aware of your school's process for logging students into your school's computer system (school logins are not the same as the NAP–SL logins available on the portal)
- ensure there are sufficient computers/devices for the students to take the test
- ensure that there is a room with an internet connection where the computers/devices can be set up (if not the computer lab)
- ensure that there is a whiteboard or chalkboard to display the following web address (this is different from the NAP–SL school portal)
- consider placing a shortcut to the assessment system's web address on each computer/device to make it easier for students
- conduct the Pre-practice Session offered as part of the TRT to familiarise yourself and the students with some of the features of the test
- arrange the test session times
- review the troubleshooting guide in the back of this manual
- review the handling of emergencies and problem situations
- review procedures for dismissing students at the end of the session, particularly if the selected Year 6 students are not your regular class
- ensure arrangements are in place for the national sample Test Observer (if applicable)
- the day before the test, remind the students about the assessment and to bring their headphones (if not provided by your school) and their reading glasses if needed
- ask students to bring a book to read quietly in case they finish the test early.

For schools with unreliable or poor internet connection, a backup test delivery solution will be provided using USB sticks. Schools must inform the NAP–SL helpdesk prior to the assessment period if this method of test delivery is required. It is expected that fewer than 5% of schools participating in NAP–SL will require the offline solution. Additional information can be provided by the NAP–SL helpdesk should your school require the backup test delivery solution.

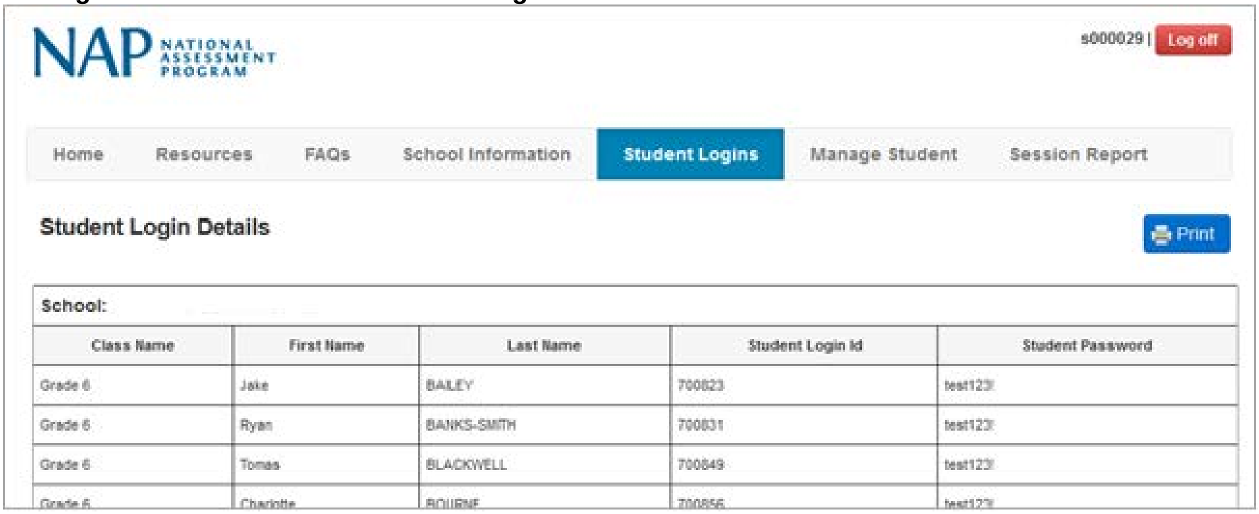
A random sample of students was selected for participation in the assessment program. The random sampling technique ensures that the selected sample is representative of all students across Australia. Therefore, it is important that ALL selected students participate in the assessment program.

How to access and print student login details

Log onto the NAP–SL school portal, (this is different from the NAP–SL test delivery system), click on the ‘Student Logins’ tab as shown below.

This page will display the list of students with their login ID and password for the test.

Figure A5.1 Portal view of student logins



To print:

1. Click on the blue ‘Print’ button.
2. Follow your browser/printer’s print settings to print the list.

Student login details should **NOT** be provided to students prior to the testing session.

We suggest that you cut the list into strips for individual students (see below) and distribute these at the commencement of the test session.

Figure A5.2 Example student login strip

Grade 6	Jake	BAILEY	700823	test123!

On test day

- confirm that the assessment space is still available
- ensure that the NAP–SL assessment web address is clearly displayed in the room
- confirm that all computers/devices and internet connection are working
- confirm all necessary equipment (keyboard, monitor, mouse, headphone socket) is working
- confirm that enough headphones are available (either a full set or sufficient spares if students bring their own)
- ensure that you have the students’ names, usernames and passwords printed and ready to hand out
- ensure that you have ready access to the troubleshooting guide
- ensure you have ready access to the NAP–SL Helpdesk contact details at Educational Assessment Australia (EAA)

- know who to contact in your school in the event of any school-specific technical problems
- ensure that the Test Administrator's script in this manual, is accessible electronically or you have a print out of this section
- record student participation information in the NAP–SL school portal
- obtain a copy of the Session Report from the NAP–SL school portal
- complete the Session Report online through the NAP–SL school portal.

Arrangements for national sample test observers (if applicable)

In order to ensure the consistency of the data collection procedures, observers will visit approximately 5% of participating schools. Your school will have been contacted already if it is to be visited by an observer. Test observers will not interfere in any way. They will only observe that the assessment has been conducted in a consistent manner, following the directions set out in this manual. The test observers will have the necessary *Working with Children* clearance or authorisation.

If your school has been advised that a test observer has been allocated, please ensure the necessary arrangements are in place to accommodate them on the test day.

2.2 Supervising the test sessions

You are responsible for monitoring the assessment sessions and the following points need to be observed:

- once the assessment instructions have begun, no more students may be admitted to the session
- only Year 6 students who have been selected are to participate
- students should not leave the session unless it is absolutely necessary
- if a student cannot complete the session (for example if they become ill), please record this information in the 'Manage Student' tab within the NAP–SL school portal either during the test session or immediately after the test session
- school observers should be limited to necessary staff members and the national sample Test Observer (if applicable)
- to ensure standardised assessment procedures and security of the assessment, NO copying of test materials, screenshots, photographs or video recordings of the sessions are allowed.

As the Test Administrator, you must also:

- read out the instructions and teacher script exactly as written
- encourage student participation
- actively monitor student conduct
- walk around the room
    - check that students are not talking
    - check that students are working on their own
    - check that students are not accessing external websites.

Assisting students

It is expected that professional and ethical behaviour will be demonstrated in all aspects of the test administration. If required and appropriate for your students, you may:

- make sure that all students undertake the practice questions and that they understand how to provide their answers
- read and clarify general instructions
- remind students of the standard response types, e.g. for multiple choice, click on the circle next to the option you wish to select
- advise students during the Objective test to only leave a question if they are unsure of the answer and move onto another question. If they have time at the end of the first session they can return to any unanswered questions
- **NOTE** that this is not possible for the **Inquiry task** as answers are locked once a student moves onto a new question
- students who have completely finished may be permitted to read a book or engage in another activity that does not distract the other students
- if a student needs extra time, contact the NAP–SL Helpdesk while the student is still logged on, using the toll-free number.

During the test, Test Administrators must NOT:

- give hints or examples
- indicate to students whether answers are correct or incorrect
- remind students about related work completed in class.

Test Administrators **MUST READ ALOUD TO STUDENTS ALL INSTRUCTIONS PROVIDED IN THE SHADED BOXES** in the script.

Practice questions

Some practice questions are provided at the outset of the test. These questions will help to:

- familiarise students with the navigation features of the online test
- provide examples of each kind of question and response types.

This is not an opportunity for teaching. Move through these questions as quickly and as efficiently as possible.

2.3 Student participation information

Participation details need to be updated in the school portal. This includes updating details for any selected students who:

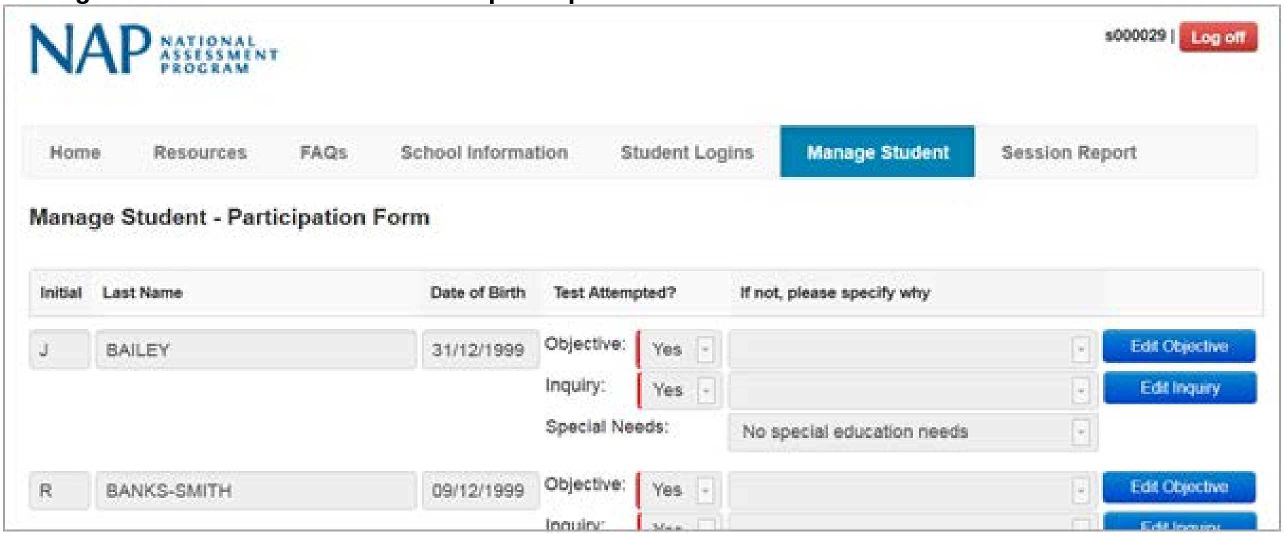
- did not attempt the Objective test (and the reason why)
- did not attempt the Inquiry task (and the reason why)
- have special education needs.

In addition, the Test Administrator should note any other information relating to participation, for example, where a student attempted the test but had to leave part way through due to illness or other reason.

How to access student participation information

Log onto the NAP–SL school portal (this is different from the NAP–SL test delivery system) and click on the 'Manage Student' tab (see below). This page will be used to record student participation information for your students.

Figure A5.3 Portal view of student participation form



The first initial, the last name and the date of birth of each student will be shown on this page. The details are set, by default, to show that each student attempted the objective test, attempted the inquiry task and that the student has NO special education needs.

A student should remain classified as having attempted the test if they logged into the test delivery system for that component.

How to record student special education needs

Note: this section can be completed in advance of the testing session.

The details are set, by default, to show that each student has NO special education needs.

For each student who has special education needs, complete the following steps:

- Click on the blue button.
- **Note**: this can be completed when you complete student non-participation details.
- Select the special education need from the menu.
- Click on the green button.

The options for the 'Special Needs' field are defined below:

No special education needs	The student has no special education needs.
Functional disability	The student has a moderate to severe permanent physical disability.
Intellectual disability	The student has a mental or emotional disability and has either been tested as cognitively delayed or is considered in the professional opinion of qualified staff to be cognitively delayed.
Limited test language proficiency	The student is not a native speaker of the language of the assessment (i.e. English) and has limited proficiency in this language.

The intent of NAP–SL is to be as **inclusive** as possible, therefore students with special needs are encouraged to participate. However, it is acknowledged that some students with limited proficiency in the language of the assessment, or those who have a physical, mental or emotional disability, may not be able to participate.

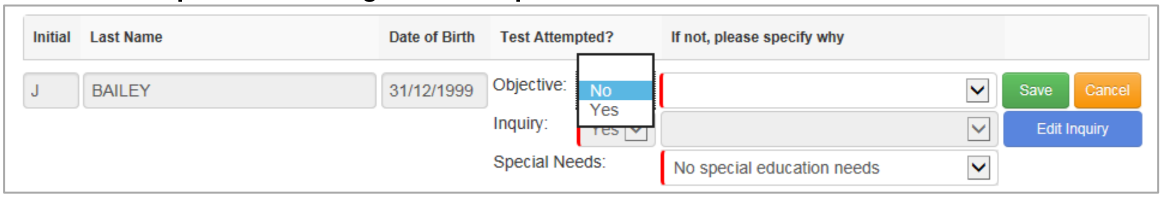
How to record student non-participation

Note: This section is not to be confused with the 'Special Needs' field. If a student did not participate in a section or all of the assessment due to their special needs, this should be recorded in the *Test Attempted* drop down menu(s) as well.

For each student who **did not attempt** the Objective test, complete the following:

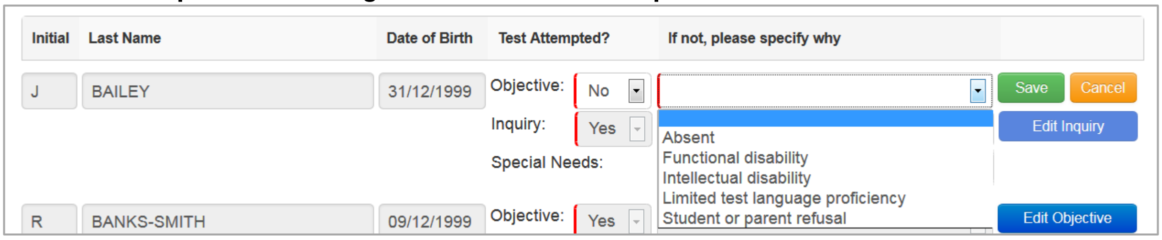
1. Click on the blue button at the end of the row for the relevant student. The menu for the objective test will be editable.
2. Select 'No' as shown.

Figure A5.4 Example of recording non-attempt

3. 

4. Select the reason as shown.

Figure A5.5 Example of recording reason for non-attempt

5. 

The reasons for a student not attempting the assessment component are defined below:

Absent	Student was absent for the test.
Functional disability	Student has a moderate to severe permanent physical disability such that he/she cannot perform in the testing situation. Functionally disabled students who can respond to the assessment should be included.
Intellectual disability	Student has a mental or emotional disability and is cognitively delayed such that he/she cannot perform in the testing situation. This includes students who are emotionally or mentally unable to follow even the general instructions of the assessment. Students should NOT be excluded from participating solely because of poor academic performance or disciplinary problems.

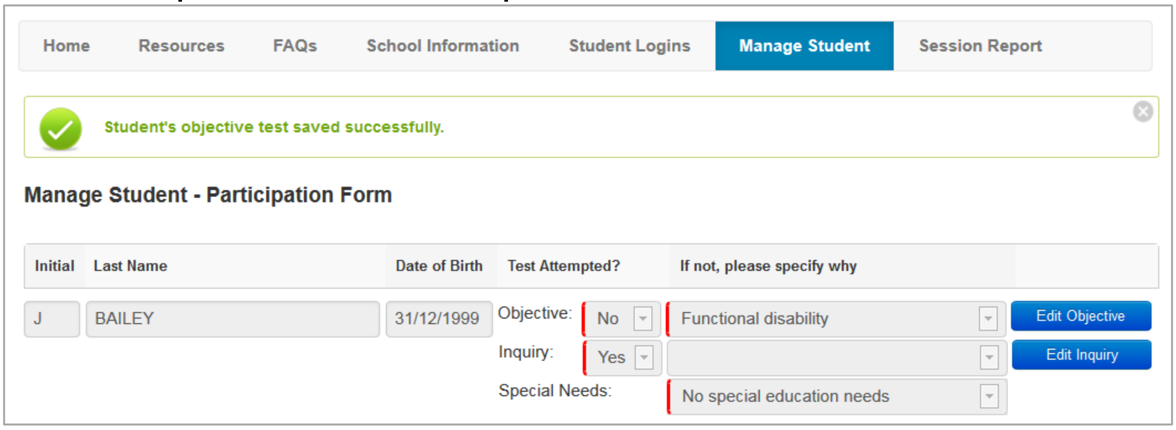
Limited test language proficiency	The student is unable to read or speak the language of the assessment (i.e. English) and would be unable to overcome the language barrier in the testing situation. Typically a student who has received less than one year of instruction in the language of the assessment may be excluded.
Student or parent refusal	Parent/caregiver requested that student not participate OR student refused to participate.

It is important that these criteria be followed strictly for the study to be comparable within states and territories across the country.

The following rule should be applied: **when in doubt, allow the student to participate.**

6. Click on the green button. Ensure that the message at the top of the screen shows that the information has been saved successfully. Once saved, the fields will turn grey and cannot be changed.

Figure A5.6 Example of saved fields on the portal



7.

Repeat steps 1–4 for all students who **did not attempt** the **Objective test**.

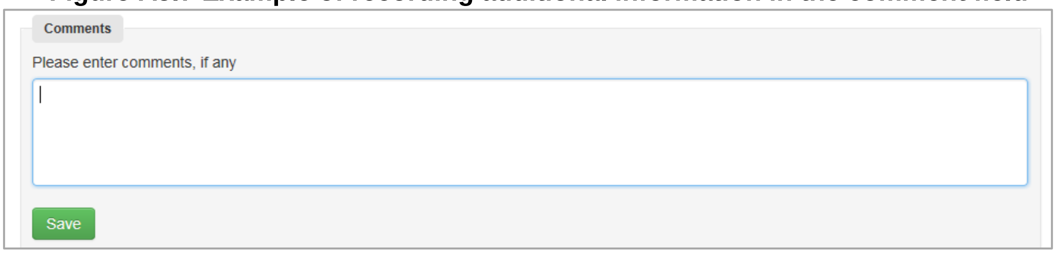
Repeat steps 1–4 for all students who **did not attempt** the **Inquiry task** (by clicking on the relevant *Inquiry* task buttons).

Note: while updating the non-participation field for a student, you can also update the student's special needs field, if applicable.

How to record additional participation information

At the bottom of this page there is a 'Comments' field as shown below.

Figure A5.7 Example of recording additional information in the comment field



This field should be used to record any additional participation information. For example:

- if a student attempted the test but had to leave part way through (due to illness, an appointment etc.)
- if a student experienced computer difficulties during the test
- if a student logged into the test, attempted a few questions but then refused to complete the remainder of the test.

To enter additional participation information:

- Enter the name of the student and relevant participation information into the Comments field.
- Click on the green button.

2.4 After the test session

Please ensure that you have accurate records of student participation as per the previous section.

For any student's name that was incorrectly spelt, please email the correct details to the NAP–SL Helpdesk.

Catch up session if more than 20% of students are absent on the test day

A follow-up session should be scheduled if more than 20% of the selected students are absent from the test session (not counting student refusals, excluded or ineligible students).

- select a date and time that will maximise the number of students attending the follow-up session. The follow-up session should be administered as soon as possible after the originally scheduled test date. Make every attempt to ensure that all previously absent students attend the follow-up session.
- you should follow the same procedures in the follow-up session as for the main assessment session. The same time allocation will apply to any follow-up session.

Session report

Ensure you have either online access to the Session Report during the assessment, or a print-out to record timing and other important information relating to the test session.

To access the Session Report, log into the NAP–SL school portal, and click on the 'Session Report' tab.

You may choose to print a copy of the Session Report prior to the test session, and complete the responses on the hardcopy and then transfer the responses back online before the end of the testing period.

To complete the Session Report:

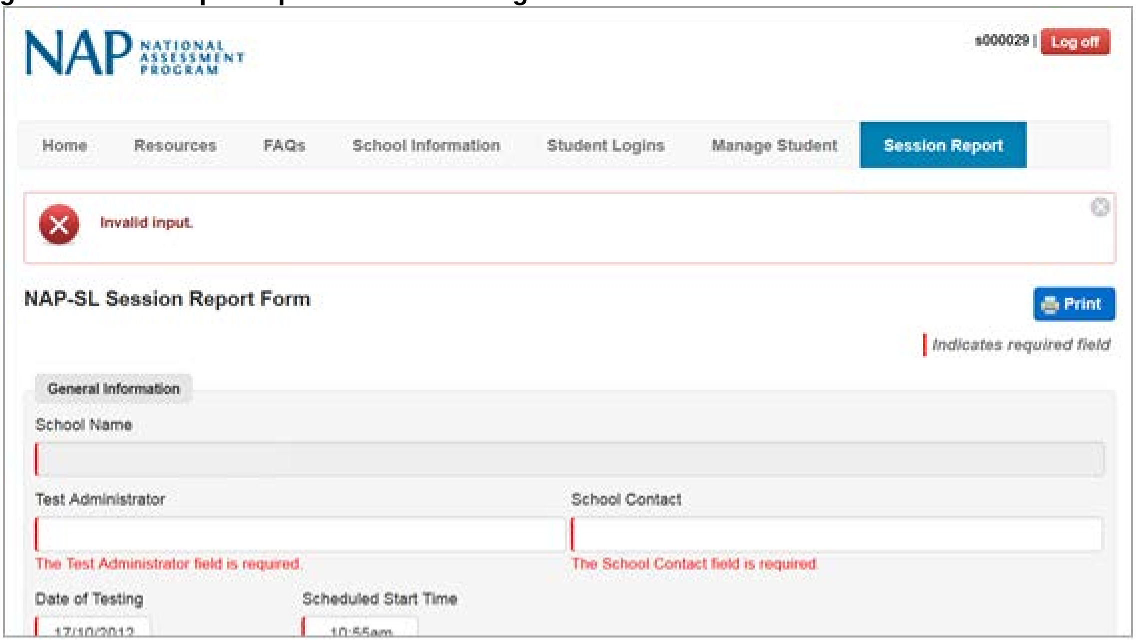
1. Complete all of the components of the report either from the pull down menu and selecting 'yes'/'no' or by typing into the field. Most fields are mandatory.
2. Click on the green button at the bottom of the screen once complete.

If all of the mandatory fields have been completed, the following will appear at the top of the screen, confirming the information entered has been saved.

Figure A5.8 Portal feedback message

If any information is missing, no changes will be saved. The following message will appear at the top of the screen and an error message will appear below any fields that contain incorrect or incomplete data. These are the fields that require further attention.

Figure A5.9 Example of portal error message



Once all required fields have been corrected, remember to click on the green submit button. Repeat this step if needed until the screen shows the information has been saved successfully.

Maintaining security

Because some of the NAP–SL materials will be used in the future to measure change between assessment cycles, their security is important. To maintain security, details of the questions must not be shared or recorded in any form including copying, screenshots, photographs or video recordings of the sessions.

3. Test administrator's instructions and script

To ensure test administration consistency, the Test Administrator's script must be read **WORD FOR WORD.**

This script appears in the **shaded boxes** on the following pages. The unshaded sections of text are the instructions and background information for the Test Administrator. Please follow these instructions carefully to ensure the assessment is conducted in the same way throughout Australia and that all students have the same instructions.

Get the students ready to begin the test

In most schools, students have school-specific login details. Help students to log into their computers and ensure that they have access to the internet. These login details are different from the NAP–SL specific login details you have printed.

If you find that the start of the session is delayed because a small number of students are having technical problems accessing your school system or if some computers/devices are not working, then you may ask the rest of the students to take out their reading materials and to read quietly until you have resolved these issues.

Once all students are logged into their computers/devices and can access the internet using the appropriate TRT compliant web-browser, ask the students to type in the web-address (URL) into the address bar of their browser and press enter.

Once entered correctly, students should see the login screen.

Help any students who are having problems. Appendix 3 [not included in this extract] includes common technical issues you may encounter with a suggested course of action.

3.1 Starting the test

Ensure all students are settled and paying attention. Hand out the pre-printed student login details, usernames and password (the password is the same for all students at your school).

Record the start time on the NAP–SL Session Report.

Say:

> You have been selected to take part in an important national education assessment. This assessment is called the National Assessment Program – Science Literacy. Its goal is to find out what students your age know about science. The results of the study will help education departments and governments determine what students are learning. We ask that you do the very best that you can.

Get the students to type in their Username and Password for the test. Once logged in they will see the student dashboard. INSIST THAT STUDENTS DO NOT CLICK ON ANY OF THE BUTTONS ON THIS SCREEN.

Check to see that students can see their name on the top right hand side of the screen.

Ask students to wait for instructions and pay attention to you.

Say:

> To make sure that all students doing this test receive exactly the same instructions, I will be reading them to you.
>
> If you have difficulty in reading the questions during the session, please raise your hand and I will help you.

3.2 Step 1: Practice questions

Read each of the practice questions in turn following the given script [Note: the script for each of the practice questions is not included in this extract].

Respond to any questions.

Say:

You may now click on the 'Finish' button.

Students should now see the 'Summary' page.

Check that all students completed all 7 questions.

Explain to students that the Summary page can be accessed at any point during the test.

Ask students to click on the 'Return to practice questions' button. Explain to students that this will allow them to go back and check answers to questions.

Ask students to return to the Summary page by clicking on the 'Summary' button on the top right hand side of the screen.

Say:

You have now finished the practice questions. You may click on the 'Finish' button.

Say:

> When you finish the Summary page you will see a confirmation box to alert you that you are leaving the practice questions. This is the opportunity to confirm that you would like to finish and go back to the dashboard by clicking 'Yes'. There is also an opportunity to return to the practice questions to check your answers by clicking the 'No' button.
>
> You may click on the 'Yes' button to now return to the dashboard but **DO NOT CLICK ON THE OBJECTIVE TEST BUTTON**.

3.3 Step 2: Objective test

All students should now be back at the Student dashboard.

Say:

Listen to my instructions **before** clicking on the Objective test button.

> You will have 1 hour to complete the Objective test.
>
> You need to work on your own to answer the questions.
>
> Remember to open and close the panel on the left hand side of each question. It includes

> information that will help you answer the questions.
>
> If you make a mistake and want to change your answer, you may do so.
>
> If you don't know the answer to a question, flag it and try the next one. If you have time, go back to the questions you didn't finish and attempt to complete them.
>
> Do not start until I tell you to. You will not need your headphones for this section.
>
> When you have finished this section, we will take a break before you start the next section, which is called the Inquiry task.
>
> Are there any questions?

Answer any student questions.

Say:

> Now click on the Objective test button on the dashboard. You will see the introduction page letting you know how many questions are in your test.
>
> To help you keep track of the time available, there is a timer located on the bottom left-hand side of the screen.
>
> Please remember, you can go back to any question and change your answer if you need to.
>
> Now begin the test by clicking the 'Next' button. Use your time carefully and do as much as you can.

Record the end time for the Introduction and the start time for the Objective test on the NAP–SL Session Report.

Note: Even though we have allocated 10 minutes for the Introduction (which includes starting the test, the practice questions and the introductory information relating to the Objective test) and 60 minutes for the Objective test we are interested to know the exact time each component takes to complete as this will vary from school to school.

Monitor the students.

After 30 minutes, say:

> Check the timer on your screen. You should have about 30 minutes to go. Don't forget that when you finish, you can go back through the test and answer any questions you may have missed if there is still time remaining.

After approximately 55 minutes, say:

> Check your timer; you should have about 5 minutes to go.

Any students who were delayed starting the session or whose session was interrupted may still have some time left, please allow them to finish.

Students who have finished should read quietly.

Once all students have finished the Objective test, **record the end time on the NAP–SL Session Report** and say:

> The time on your test should now have finished.
>
> After clicking on the 'Finish' button on the Summary screen you will receive two confirmation boxes. You should click 'Yes' in the first confirmation box and then 'ok' in the second to be returned to the dashboard.
>
> Click on the 'Log out' button on the top right hand corner. But do **NOT** close the web browser. We will now have a break.

The break does not have to be timed exactly and is at the discretion of the teacher. It should align with normal school policies or breaks. It should be no less than 15 minutes.

PLEASE ENSURE STUDENTS LOG OUT OF THE NAP–SL TEST BEFORE THE BREAK. The NAP—SL test delivery system will remember their progress.

When students log out, they **must** click on the 'Log out' button found in the top right hand corner. However, they **must NOT** just close the web browser.

Note: if your school computers have a timeout feature, students who have not logged out of the NAP–SL test while on their break, may be locked out of the system. If this occurs contact the NAP–SL Helpdesk immediately on 1800 231 085 to unlock the students' accounts.

3.4 Steps 3 and 4: Inquiry task and student survey

Returning from the break

Ensure students are seated in the same places they were in for the Objective test. Ask the students to log back into the system.

Check that all students are back on the Student dashboard before proceeding.

Introducing the inquiry task and the student survey

Ensure that students do not click on the Inquiry task button on the dashboard until you have delivered the following instructions.

Say:

> In this part of the test, there are two sections to complete. You will have 35 minutes to complete the first section which is the Inquiry task. After that there is a short survey to complete.

> For the Inquiry task, there are a number of videos you will need to watch. You will need your headphones but do not put them on yet.
>
> The black banner at the top of the question will let you know when there is a new video to watch.
>
> Watch the videos carefully in the panel on the left. They include information that will help you answer the questions. You may watch the videos in full screen mode and re-watch the videos if you need to.
>
> In this section, once you have moved on from a question, you cannot go back and change your answer. However, you can look back at the questions you have done.
>
> Do not start until I tell you to. Are there any questions?

Answer any student questions.

Say:

> In the final section of the test, you will be asked to complete a survey.
>
> This survey is to find out your opinions and ideas about science.
>
> Please read each sentence carefully and answer as accurately as you can. You may ask for help to read the questions or if you are not sure how to show your answer.
>
> Remember: there are no right or wrong answers when answering the survey. Your answers should be the ones you think are best for you. Are there any questions?
>
> Now click on the Inquiry task button. You will need to enter a PIN to access this section of the test.
>
> When you have finished this section of the test you can go back to the dashboard and complete the survey by clicking on the Student survey button.
>
> Type in this PIN: [PIN] and click OK.
>
> Read through the Introduction page, put on your headphones and then click ‘Next’ to start the test.
>
> Use your time carefully and do as much as you can.

Record the start time for the Inquiry task on the NAP–SL Session Report.

Monitor the students.

After 15 minutes, say:

You have about 20 minutes to go.

After approximately 30 minutes, say:

Check the timer; you should have about 5 minutes to go.

Any students who were delayed starting the session or whose session was interrupted may still have some time left. Allow them to finish.

At the end of the Inquiry task, students will be presented with two confirmation boxes before being returned to the dashboard.

After all students have finished, record the end time of the Inquiry task on the NAP–SL Session Report.

Students should commence the survey as soon as they have completed the Inquiry task.

Record the start time of the Student survey on the NAP–SL Session Report.

Once all remaining students have finished the Inquiry task, say:

The time on your Inquiry task should now have finished.

Ensure that all students find their way back to the Student dashboard.

Students should start the survey as soon as they have finished the Inquiry task.

No PIN is required.

Note: Survey items may be read to students, but no further explanations should be given. You must not assist students with their answers.

Extra time for the survey can be provided if needed as this section of the test is not automatically timed.

After all students have finished the survey, record the end time on the NAP–SL Session Report.

Once all students have completed the survey, say:

Please stop. Thank you for helping with this important study.

Concluding the test session

Once each student has completed the survey, a message will appear advising that responses have been successfully submitted and that students have been automatically logged off.

Please check that each student has been successfully logged out of the NAP–SL test delivery system.

Dismiss the students according to the policy of the school.

Complete the questions in the NAP–SL Session Report.

APPENDIX 6 VARIABLES IN FILE

Table A6.1 File Name: NAPSL 2015_PV_2016-06-02.sav

Variable names	Description
studentid	unique student identifier
SchoolID	unique school identifier
form	test form (1 to 7)
Q1 to Q107	test items 1 to 107
State	state / territory
Sector	school sector
GeoLocation	school geographic location
Sex	student gender
ATSI	student Indigenous status
LBOTE	student language background
wle	ConQuest weighted likelihood estimate (WLE)
school_mean_wle	School mean ConQuest weighted likelihood estimate (WLE)
FinalWeight	final sample weight
dum_stateNSW	state regression variable 1
dum_stateNT	state regression variable 2
dum_stateQLD	state regression variable 3
dum_stateSA	state regression variable 4
dum_stateTAS	state regression variable 5
dum_stateVIC	state regression variable 6
dum_stateWA	state regression variable 7
dum_sectorG	sector regression variable 1
dum_sectorI	sector regression variable 2
dum_geolocation2	geographic regression variable 1
dum_geolocation3	geographic regression variable 2
dum_sex2	sex regression variable 1
dum_sex9	sex regression variable 2
dum_atsi1	ATSI regression variable 1
dum_atsi9	ATSI regression variable 2
dum_lbote1	LBOTE regression variable 1
dum_lbote9	LBOTE regression variable 2
PV15.1 to PV15.10	2015 plausible values calibrated free (1 to 10)
EAP	EAP
EAPError	EAP SE value
PV1 to PV10	2015 plausible values located on historic (2006) scale (1 to 10)
RW1 to RW309	replicate weights (1 to 309)

APPENDIX 7 CONQUEST CONTROL FILE FOR PRODUCING PLAUSIBLE VALUES

File Name: NAPSL _pv.cqc

```
reset;

Title population estimates - Trailing Missing as Incorrect;

data NAPSL _final.dat;

format responses 1-107, pid 109-116, booklet 123
  FinalWeight 132-139
  schoolmeanWLE 140-147
  stateNSW 148
  stateNT 149
  stateQLD 150
  stateSA 151
  stateTAS 152
  stateVIC 153
  stateWA 154
  sectorG 155
  sectorI 156
  geoProvincial 157
  geoRemote   158
  gender2 159
  gender9 160
  atsi1  161
  atsi9  162
  lbote1  163
  lbote9  164;

codes 0,1,2,3,4,5,9,A,B;

label << NAPSL .lab;
```

```
set n_plausible=10;

caseweight FinalWeight;


key
11334121111111143111311112111144411111141341211434131111111111131141121133
1113411221221112114211112213111111 ! 1;

key
XXXXXXXXXXXXXXXXXXX2XXXXXXXXXXXXXXXXX2XXXXXXXXXXXXXXXXXXXXXXX
XX2XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX2 ! 2;

model booklet + item + item*step;

regression

  schoolmeanWLE

  stateNSW

  stateNT

  stateQLD

  stateSA

  stateTAS

  stateVIC

  stateWA

  sectorG

  sectorI

  geoProvincial

  geoRemote

  gender2

  gender9

  atsi1

  atsi9

  lbote1

  lbote9  ;
```

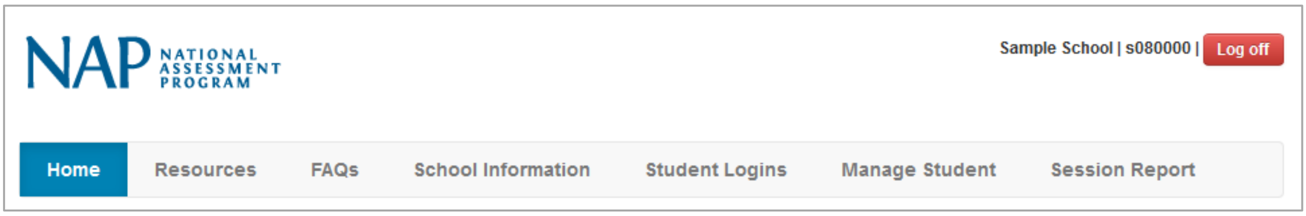
```
import anchor_param << calibration.anc;

estimate ! iterations=2000, fit=no, nodes=200;

show !estimate=latent >> popPV.shw;

itanal >> popPV.itn;

show cases !estimate=latent >> popPV.pls;

show cases !estimate=wle >> popPV.WLE;
```

APPENDIX 8 NAP—SL SCHOOL PORTAL OVERVIEW

A school portal was developed for the 2015 NAP sample assessments – science literacy (NAP—SL). It was designed to provide schools with up-to-date information and ensure secure information sharing between each school and Educational Assessment Australia (EAA).

General navigation instructions

The following navigation bar will appear across the top of each page when you log on. This bar will be customised to show only the pages relevant for each stage. As a result, the number of pages will vary for different stages. The following lists the pages for stages 4 and 5.



Home

The *Home* page contains key information about the school portal and most importantly a timeline for the assessment program.

Resources

The *Resources* page contains general information and materials for the assessment program. This will include links to user guides, administration manuals and copies of email correspondence sent to schools.

FAQs

The *FAQs* page contains answers to frequently asked questions relevant to the current stage of the project. This page will be updated regularly throughout the assessment program.

School Information

The *School Information* page contains the details that we currently have on file for the school. These details can be viewed and/or updated at any time.

Student Logins

The *Student Logins* page provides the list of the selected students and their individual login details. This list can be printed for use on test day.

Manage Student

The *Manage Student* page will be used during Stage 5 for schools to provide information about each selected student's participation on the day of the test.

Session Report

The *Session Report* page contains a survey regarding the testing day. Test administrators must complete the form to provide valuable feedback about the test administration.

APPENDIX 9 2015 NAP—SL : STUDENT SURVEY

The survey was administered after the inquiry task on the same online platform as the rest of the assessment. Students were given the following instructions before starting the survey.

In the final section of the test, you will be asked to complete a survey.

This survey is to find out your opinions and ideas about science.

Please read each sentence carefully and answer as accurately as you can. You may ask for help to read the questions or if you are not sure how to show your answer.

Remember: there are no right or wrong answers when answering the survey. Your answers should be the ones you think are best for you. Are there any questions?

Each group of items appeared together on the screen one group at a time.

Table A9.1 Student survey questions and response options

Item	Text	Response options			
Group 1: How much do you agree with the statements below?					
1	I would like to learn more science at school.	Strongly Agree	Agree	Disagree	Strongly Disagree
2	I think it would be interesting to be a scientist.	Strongly Agree	Agree	Disagree	Strongly Disagree
3	I enjoy doing science.	Strongly Agree	Agree	Disagree	Strongly Disagree
4	I enjoy learning new things in science.	Strongly Agree	Agree	Disagree	Strongly Disagree
Group 2: How much do you agree with the statements below?					
5	I learn science topics quickly.	Strongly Agree	Agree	Disagree	Strongly Disagree
6	I can usually give good answers to science questions.	Strongly Agree	Agree	Disagree	Strongly Disagree
7	I can understand new ideas about science easily.	Strongly Agree	Agree	Disagree	Strongly Disagree
Group 3: How much do you agree with the statements below?					
8	Science is an everyday part of my life.	Strongly Agree	Agree	Disagree	Strongly Disagree
9	Science is important for lots of jobs.	Strongly Agree	Agree	Disagree	Strongly Disagree

Item	Text	Response options			
10	Science is important: it changes how we live.	Strongly Agree	Agree	Disagree	Strongly Disagree
11	Scientific information helps people make good decisions.	Strongly Agree	Agree	Disagree	Strongly Disagree
Group 4: How much do you agree with the statements below?					
12	Science is about remembering facts.	Strongly Agree	Agree	Disagree	Strongly Disagree
13	Science is about doing experiments.	Strongly Agree	Agree	Disagree	Strongly Disagree
14	Science is finding out about how things work.	Strongly Agree	Agree	Disagree	Strongly Disagree
15	Science is about solving problems.	Strongly Agree	Agree	Disagree	Strongly Disagree
16	Science is quite easy for most people to understand.	Strongly Agree	Agree	Disagree	Strongly Disagree
Group 5: How often do you do these things outside of school?					
17	I view TV programs, DVDs or websites about science topics at home.	Frequently (More than 2 times a week)	Often (1 or 2 times a week)	Sometimes (less than once a week)	Never
18	I read books, newspapers or magazine articles about science topics.	Frequently (More than 2 times a week)	Often (1 or 2 times a week)	Sometimes (less than once a week)	Never
19	I talk about science ideas with my friends and family.	Frequently (More than 2 times a week)	Often (1 or 2 times a week)	Sometimes (less than once a week)	Never
Group 6: How often do you do these things at school?					
20	I view TV programs, DVDs or websites about science topics at school.	Frequently (More than 2 times a week)	Often (1 or 2 times a week)	Sometimes (less than once a week)	Never
21	We read books, newspapers or magazine articles about science topics at school.	Frequently (More than 2 times a week)	Often (1 or 2 times a week)	Sometimes (less than once a week)	Never
Group 7: How often do you do these things at school?					
22	During science lessons I get to plan and carry out my own investigations.	Always	Mostly	Sometimes	Never
23	When our class investigates things in science, we work in groups to carry out the investigation.	Always	Mostly	Sometimes	Never
24	I use a computer for research or to present my science ideas and findings.	Always	Mostly	Sometimes	Never

Item	Text	Response options			
25	Our class has in-depth discussions about science ideas.	Always	Mostly	Sometimes	Never
Group 8: Which of these science topics have you studied at school?					
26	Earth and space (Earth and space sciences) - for example, weather, soil, rocks, gravity, using Earth's resources, the planets, sun and moon.	Yes	No	I don't know	
27	Energy and force (Physical sciences) - for example, how toys and other machines work, electricity, heat, light, sound, magnets.	Yes	No	I don't know	
28	Living things (Biological sciences) - for example, living and non living things, how animals and plants survive in their environment, life cycles, interdependence.	Yes	No	I don't know	
29	Matter (Chemical sciences) - for example, the different properties of materials such as plastics and metals, the different uses of materials, changes to materials (solids, liquids and gases).	Yes	No	I don't know	
Group 9: How often do you have science lessons at school?					
30	How often do you have science lessons at school?	More than once a week	Once a week	Less than once a week	Hardly Ever
Group 10: Do you agree with the statements below?					
31	My classroom teacher teaches science to our class.	Yes	No		
32	I think my teacher enjoys teaching science.	Yes	No		
33	Our teacher invites visitors to school to talk to us about science topics.	Yes	No		
34	Our class goes on excursions related to the science topics we are learning about.	Yes	No		
Group 11: How much do you agree with the statements below? Science is about…					
35	making observations about the world.	Strongly Agree	Agree	Disagree	Strongly Disagree
36	asking questions about objects and events.	Strongly Agree	Agree	Disagree	Strongly Disagree
37	making predictions and testing them.	Strongly Agree	Agree	Disagree	Strongly Disagree
38	describing patterns and relationships.	Strongly Agree	Agree	Disagree	Strongly Disagree
39	using evidence to develop explanations.	Strongly Agree	Agree	Disagree	Strongly Disagree

Item	Text	Response options			
Group 12: How much do you agree with the statements below?					
40	People from many different countries have made important contributions to science.	Strongly Agree	Agree	Disagree	Strongly Disagree
41	Women and men are involved in science.	Strongly Agree	Agree	Disagree	Strongly Disagree
42	People from all cultural backgrounds in Australia are involved in science.	Strongly Agree	Agree	Disagree	Strongly Disagree
43	People of all ages are involved in science.	Strongly Agree	Agree	Disagree	Strongly Disagree

APPENDIX 10 STUDENT SURVEY 2015 RESULTS

The following table provides a comparison of survey items that appeared in both the 2012 and the 2015 student surveys. The percentages have been weighted to account for sampling affects across states and territories and so differ slightly from the percentages shown in parts of the *2015 NAP—SL Public Report*.

The significance of the difference has been calculated taking into account a Bonferroni adjustment.

Table A10.1 Comparison of student responses to common questions in the 2009, 2012 and 2015 student surveys

Group and number	Survey question		2012 Per cent	SE	2015 Per cent	SE	2012 to 2015 difference (percentage points)	Significant?	Change from 2012
G01	**How much do you agree with the statements below?**								
1	I would like to learn more science at school.	Strongly Disagree	3.73	0.30	2.18	0.20	-1.55	YES	decrease
		Disagree	15.08	0.56	10.69	0.45	-4.39	YES	decrease
		Agree	57.10	0.71	58.09	0.74	0.99	NO	
		Strongly Agree	24.08	0.67	28.20	0.77	4.12	YES	increase
2	I think it would be interesting to be a scientist.	Strongly Disagree	12.33	0.51	7.27	0.37	-5.07	YES	decrease
		Disagree	28.93	0.60	23.46	0.56	-5.47	YES	decrease
		Agree	44.67	0.75	51.17	0.63	6.50	YES	increase
		Strongly Agree	14.06	0.43	18.10	0.49	4.04	YES	increase
3	I enjoy doing science.	Strongly Disagree	3.32	0.27	2.57	0.21	-0.75	NO	
		Disagree	11.34	0.45	10.53	0.44	-0.81	NO	
		Agree	56.68	0.66	56.38	0.72	-0.30	NO	
		Strongly Agree	28.67	0.70	30.53	0.80	1.86	NO	
4	I enjoy learning new things in science.	Strongly Disagree	2.41	0.25	1.90	0.18	-0.51	NO	
		Disagree	8.48	0.40	5.92	0.33	-2.55	YES	decrease
		Agree	53.49	0.67	49.51	0.78	-3.98	YES	decrease
		Strongly Agree	35.63	0.71	42.67	0.83	7.04	YES	increase

Group and number	Survey question		2012 Per cent	SE	2015 Per cent	SE	2012 to 2015 difference (percentage points)	Significant?	Change from 2012
G02	**How much do you agree with the statements below?**								
5	I learn science topics quickly.	Strongly Disagree	6.26	0.35	3.76	0.25	-2.50	YES	decrease
		Disagree	34.88	0.64	27.25	0.63	-7.63	YES	decrease
		Agree	48.39	0.70	54.95	0.71	6.56	YES	increase
		Strongly Agree	10.48	0.37	14.04	0.45	3.56	YES	increase
6	I can usually give good answers to science questions.	Strongly Disagree	4.81	0.29	3.02	0.23	-1.80	YES	decrease
		Disagree	29.82	0.64	23.13	0.56	-6.69	YES	decrease
		Agree	55.11	0.77	60.35	0.68	5.24	YES	increase
		Strongly Agree	10.26	0.38	13.51	0.45	3.25	YES	increase
7	I can understand new ideas about science easily.	Strongly Disagree	4.79	0.32	2.66	0.21	-2.13	YES	decrease
		Disagree	28.47	0.64	22.15	0.54	-6.32	YES	decrease
		Agree	51.71	0.72	55.43	0.65	3.72	YES	increase
		Strongly Agree	15.04	0.46	19.77	0.56	4.73	YES	increase
G03	**How much do you agree with the statements below?**								
8	Science is an everyday part of my life.	Strongly Disagree	19.22	0.65	12.25	0.41	-6.97	YES	decrease
		Disagree	43.62	0.71	41.43	0.61	-2.19	NO	
		Agree	24.66	0.58	32.23	0.61	7.57	YES	increase
		Strongly Agree	12.50	0.60	14.09	0.52	1.60	NO	
9	Science is important for lots of jobs.	Strongly Disagree	2.96	0.26	1.86	0.16	-1.11	YES	decrease
		Disagree	17.40	0.58	14.62	0.45	-2.78	YES	decrease
		Agree	52.72	0.57	52.18	0.61	-0.54	NO	
		Strongly Agree	26.92	0.64	31.35	0.63	4.43	YES	increase
10	Science is important: it changes how	Strongly Disagree	2.87	0.25	1.50	0.16	-1.37	YES	decrease

Group and number	Survey question		2012 Per cent	SE	2015 Per cent	SE	2012 to 2015 difference (percentage points)	Significant?	Change from 2012
	we live.	Disagree	10.89	0.43	7.47	0.34	-3.42	YES	decrease
		Agree	45.20	0.70	41.67	0.71	-3.53	YES	decrease
		Strongly Agree	41.04	0.78	49.36	0.81	8.31	YES	increase
11	Scientific information helps people make good decisions.	Strongly Disagree	3.55	0.30	2.29	0.19	-1.26	YES	decrease
		Disagree	18.01	0.52	13.58	0.44	-4.42	YES	decrease
		Agree	50.58	0.59	50.81	0.70	0.23	NO	
		Strongly Agree	27.87	0.61	33.32	0.64	5.45	YES	increase
G04	**How much do you agree with the statements below?**								
12	Science is about remembering facts.	Strongly Disagree	6.75	0.32	5.30	0.30	-1.45	YES	decrease
		Disagree	36.27	0.70	36.19	0.67	-0.08	NO	
		Agree	45.83	0.62	46.86	0.65	1.03	NO	
		Strongly Agree	11.15	0.45	11.65	0.45	0.50	NO	
13	Science is about doing experiments.	Strongly Disagree	2.19	0.21	2.06	0.19	-0.14	NO	
		Disagree	15.00	0.46	18.94	0.52	3.94	YES	increase
		Agree	49.88	0.60	49.98	0.63	0.10	NO	
		Strongly Agree	32.92	0.62	29.02	0.60	-3.90	YES	decrease
14	Science is finding out about how things work.	Strongly Disagree	1.18	0.20	0.59	0.09	-0.59	YES	decrease
		Disagree	4.78	0.26	3.95	0.24	-0.83	NO	
		Agree	52.54	0.65	50.24	0.67	-2.30	NO	
		Strongly Agree	41.50	0.62	45.22	0.70	3.72	YES	increase
15	Science is about solving problems.	Strongly Disagree	4.21	0.30	3.23	0.23	-0.97	YES	decrease
		Disagree	22.31	0.55	16.29	0.48	-6.02	YES	decrease
		Agree	50.59	0.62	53.02	0.65	2.43	YES	increase
		Strongly Agree	22.90	0.55	27.46	0.67	4.56	YES	increase

Group and number	Survey question		2012 Per cent	SE	2015 Per cent	SE	2012 to 2015 difference (percentage points)	Significant?	Change from 2012
16	Science is quite easy for most people to understand.	Strongly Disagree	10.27	0.41	8.98	0.39	-1.29	NO	
		Disagree	43.08	0.67	46.29	0.71	3.21	YES	increase
		Agree	37.50	0.62	36.46	0.65	-1.04	NO	
		Strongly Agree	9.14	0.37	8.26	0.40	-0.89	NO	
G05	**How often do you do these things outside of school?**								
17	I view TV programs, DVDs or websites about science topics at home.	Never	30.60	0.71	28.05	0.63	-2.54	YES	decrease
		Sometimes	41.16	0.63	40.80	0.61	-0.36	NO	
		Often	18.89	0.55	19.63	0.54	0.74	NO	
		Frequently	9.35	0.39	11.51	0.42	2.16	YES	increase
18	I read books, newspapers or magazine articles about science topics.	Never	40.95	0.82	36.02	0.69	-4.93	YES	decrease
		Sometimes	36.77	0.68	36.56	0.68	-0.21	NO	
		Often	15.82	0.47	18.66	0.51	2.84	YES	increase
		Frequently	6.46	0.32	8.76	0.37	2.30	YES	increase
19	I talk about science ideas with my friends and family.	Never	43.60	0.83	35.40	0.68	-8.20	YES	decrease
		Sometimes	33.84	0.66	33.01	0.56	-0.83	NO	
		Often	14.99	0.43	20.37	0.50	5.38	YES	increase
		Frequently	7.57	0.32	11.22	0.45	3.65	YES	increase
G06	**How often do you do these things at school?**								
20	I view TV programs, DVDs or websites about science topics at school.	Never	19.94	0.79	15.79	0.51	-4.15	YES	decrease
		Sometimes	47.80	0.85	41.89	0.75	-5.91	YES	decrease
		Often	26.23	0.83	33.19	0.75	6.96	YES	increase
		Frequently	6.03	0.48	9.14	0.42	3.10	YES	increase
21	We read books, newspapers or	Never	23.50	0.83	24.03	0.61	0.54	NO	

Group and number	Survey question		2012 Per cent	SE	2015 Per cent	SE	2012 to 2015 difference (percentage points)	Significant?	Change from 2012
	magazine articles about science topics at school.	Sometimes	45.71	0.76	42.02	0.64	-3.69	YES	decrease
		Often	23.96	0.75	26.22	0.63	2.26	NO	
		Frequently	6.83	0.47	7.73	0.38	0.90	NO	
G07	**How often do you do these things at school?**								
22	During science lessons I get to plan and carry out my own investigations.	Never	23.33	0.90	13.71	0.55	-9.62	YES	decrease
		Sometimes	49.66	0.74	48.34	0.67	-1.33	NO	
		Mostly	21.93	0.71	30.09	0.61	8.16	YES	increase
		Always	5.07	0.42	7.86	0.34	2.78	YES	increase
23	When our class investigates things in science, we work in groups to carry out the investigation.	Never	7.45	0.60	3.95	0.31	-3.50	YES	decrease
		Sometimes	32.15	0.85	25.08	0.66	-7.07	YES	decrease
		Mostly	45.33	0.87	48.82	0.76	3.49	YES	increase
		Always	15.07	0.60	22.15	0.87	7.08	YES	increase
24	I use a computer for research or to present my science ideas and findings.	Never	16.99	0.78	11.86	0.70	-5.13	YES	decrease
		Sometimes	39.70	0.78	35.74	0.74	-3.96	YES	decrease
		Mostly	30.90	0.74	35.70	0.82	4.80	YES	increase
		Always	12.41	0.55	16.71	0.66	4.29	YES	increase
25	Our class has in-depth discussions about science ideas.	Never	15.49	0.82	10.23	0.53	-5.27	YES	decrease
		Sometimes	44.64	0.79	38.67	0.74	-5.97	YES	decrease
		Mostly	26.75	0.73	32.27	0.67	5.52	YES	increase
		Always	13.12	0.67	18.83	0.61	5.72	YES	increase
G08	**Which of these science topics have you studied at school?**								
26	Earth and Space (Earth and space sciences)	Yes	76.64	0.87	76.58	0.74	-0.06	NO	
		No	9.34	0.52	8.40	0.43	-0.94	NO	

Group and number	Survey question		2012 Per cent	SE	2015 Per cent	SE	2012 to 2015 difference (percentage points)	Significant?	Change from 2012
		I don't know	14.02	0.55	15.02	0.52	1.00	NO	
27	Energy and Force (Physical sciences)	Yes	74.87	0.96	70.36	0.91	-4.51	YES	decrease
		No	12.09	0.59	14.00	0.62	1.91	NO	
		I don't know	13.04	0.59	15.64	0.56	2.60	YES	increase
28	Living Things (Biological sciences)	Yes	72.19	0.89	68.51	0.85	-3.68	YES	decrease
		No	12.34	0.56	14.84	0.60	2.50	YES	increase
		I don't know	15.47	0.56	16.65	0.58	1.18	NO	
29	Matter (Chemical sciences)	Yes	63.44	0.98	67.43	0.97	3.99	YES	increase
		No	15.43	0.57	14.67	0.67	-0.76	NO	
		I don't know	21.14	0.69	17.90	0.59	-3.23	YES	decrease
G09	**How often do you have science lessons at school?**								
30	How often do you have science lessons at school?	hardly ever	19.81	1.23	14.99	0.91	-4.82	YES	decrease
		less than once a week	18.03	0.93	17.13	0.87	-0.91	NO	
		once a week	39.71	1.50	45.76	1.51	6.05	YES	increase
		more than once a week	22.44	1.23	22.12	1.22	-0.32	NO	
G10	**Do you agree with the statements below?**								
31	My classroom teacher teaches science to our class.	No	29.30	1.55	26.69	1.28	-2.62	NO	
		Yes	70.70	1.55	73.31	1.28	2.62	NO	
32	I think my teacher enjoys teaching science.	No	17.58	0.93	14.92	0.66	-2.66	YES	decrease
		Yes	82.42	0.93	85.08	0.66	2.66	YES	increase
33	Our teacher invites visitors to school to talk to us about science topics.	No	69.11	1.37	69.40	0.99	0.29	NO	
		Yes	30.89	1.37	30.60	0.99	-0.29	NO	

Group and number	Survey question		2012 Per cent	SE	2015 Per cent	SE	2012 to 2015 difference (percentage points)	Significant?	Change from 2012
34	Our class goes on excursions related to the science topics we are learning about.	No	59.64	1.38	61.21	1.07	1.57	NO	
		Yes	40.36	1.38	38.79	1.07	-1.57	NO	

Notes: Percentages may not add to 100 per cent due to rounding. In Groups 5 and 6 (items 17 to 21) 'Frequently' means more than 2 times a week. 'Often' means 1 or 2 times a week. 'Sometimes' means less than once a week. Group 8 (items 26 to 29) were amended in 2015 to include the equivalent strand names from the Australian Curriculum: science.

APPENDIX 11 ITEM RUNNING ORDER FOR MAIN STUDY

Below is a list of items by test forms:

- Forms 1 to 7 refer to the objective test forms
- InqTask 1 and InqTask 2 refer to the two inquiry tasks
- Practice and TRTPrePractice refer to practice items
- Survey refers to the survey item groups.

Table A11. 1 Item running order for the main study by form

Form	Form Q#	Form Code	Cluster	Cluster Q#	Cluster Code	IARS code	Item Name
Form1	1	F1Q01	1	1	C1Q01	x00000606	NSL06H149.49
Form1	2	F1Q02	1	2	C1Q02	x00000607	NSL06H149.50
Form1	3	F1Q03	1	3	C1Q03	x00000599	NSL09H308.08
Form1	4	F1Q04	1	4	C1Q04	x00000600	NSL09H308.09
Form1	5	F1Q05	1	5	C1Q05	x00000601	NSL09H308.10
Form1	6	F1Q06	1	6	C1Q06	x00000587	NSL06H084.84
Form1	7	F1Q07	1	7	C1Q07	x00000588	NSL06H084.85
Form1	8	F1Q08	1	8	C1Q08	x00000589	NSL06H084.86
Form1	9	F1Q09	1	9	C1Q09	x00000590	NSL06H084.87
Form1	10	F1Q10	1	10	C1Q10	x00000591	NSL06H084.88
Form1	11	F1Q11	1	11	C1Q11	x00000602	NSL09H405.05
Form1	12	F1Q12	1	12	C1Q12	x00000603	NSL09H405.06
Form1	13	F1Q13	1	13	C1Q13	x00000626	NSL12H503.03
Form1	14	F1Q14	1	14	C1Q14	x00000627	NSL12H503.06
Form1	15	F1Q15	2	1	C2Q01	x00000595	NSL09H360.60
Form1	16	F1Q16	2	2	C2Q02	x00000596	NSL09H360.62
Form1	17	F1Q17	2	3	C2Q03	x00000640	NSL12H559.59
Form1	18	F1Q18	2	4	C2Q04	x00000641	NSL12H559.61
Form1	19	F1Q19	2	5	C2Q05	x00000642	NSL12H559.62
Form1	20	F1Q20	2	6	C2Q06	x00002817	NSL12H564.64C
Form1	21	F1Q21	2	7	C2Q07	x00000645	NSL12H564.65
Form1	22	F1Q22	2	8	C2Q08	x00003646	NSL12H564.66C
Form1	23	F1Q23	2	9	C2Q09	x00000648	NSL12H564.69
Form1	24	F1Q24	2	10	C2Q10	x00000649	NSL12H564.70
Form1	25	F1Q25	2	11	C2Q11	x00000592	NSL06H041.41
Form1	26	F1Q26	2	12	C2Q12	x00000593	NSL06H041.44
Form1	27	F1Q27	4	1	C4Q01	x00000628	NSL12H517.17
Form1	28	F1Q28	4	2	C4Q02	x00000629	NSL12H517.18
Form1	29	F1Q29	4	3	C4Q03	x00000630	NSL12H517.19
Form1	30	F1Q30	4	4	C4Q04	x00000631	NSL12H517.21
Form1	31	F1Q31	4	5	C4Q05	x00000632	NSL12H517.22
Form1	32	F1Q32	4	6	C4Q06	x00002815	NSL12H551.51C
Form1	33	F1Q33	4	7	C4Q07	x00000637	NSL12H551.52
Form1	34	F1Q34	4	8	C4Q08	x00000638	NSL12H551.53
Form1	35	F1Q35	4	9	C4Q09	x00000639	NSL12H551.54
Form1	36	F1Q36	4	10	C4Q10	x00000581	NSL06H021.21
Form1	37	F1Q37	4	11	C4Q11	x00000582	NSL06H021.22
Form1	38	F1Q38	4	12	C4Q12	x00000583	NSL06H021.23
Form2	1	F2Q01	2	1	C2Q01	x00000595	NSL09H360.60

Form	Form Q#	Form Code	Cluster	Cluster Q#	Cluster Code	IARS code	Item Name
Form2	2	F2Q02	2	2	C2Q02	x00000596	NSL09H360.62
Form2	3	F2Q03	2	3	C2Q03	x00000640	NSL12H559.59
Form2	4	F2Q04	2	4	C2Q04	x00000641	NSL12H559.61
Form2	5	F2Q05	2	5	C2Q05	x00000642	NSL12H559.62
Form2	6	F2Q06	2	6	C2Q06	x00002817	NSL12H564.64C
Form2	7	F2Q07	2	7	C2Q07	x00000645	NSL12H564.65
Form2	8	F2Q08	2	8	C2Q08	x00003646	NSL12H564.66C
Form2	9	F2Q09	2	9	C2Q09	x00000648	NSL12H564.69
Form2	10	F2Q10	2	10	C2Q10	x00000649	NSL12H564.70
Form2	11	F2Q11	2	11	C2Q11	x00000592	NSL06H041.41
Form2	12	F2Q12	2	12	C2Q12	x00000593	NSL06H041.44
Form2	13	F2Q13	3	1	C3Q01	x00000075	NSL15E_C006.2
Form2	14	F2Q14	3	2	C3Q02	x00000073	NSL15E_C006.1
Form2	15	F2Q15	3	3	C3Q03	x00000482	NSL15E_W002.1
Form2	16	F2Q16	3	4	C3Q04	x00000483	NSL15E_W002.2
Form2	17	F2Q17	3	5	C3Q05	x00000484	NSL15E_W002.3
Form2	18	F2Q18	3	6	C3Q06	x00000485	NSL15E_W002.4
Form2	19	F2Q19	3	7	C3Q07	x00000505	NSL15E_W002.5
Form2	20	F2Q20	3	8	C3Q08	x00003651	NSL15E_Z002.2C
Form2	21	F2Q21	3	9	C3Q09	x00000134	NSL15E_Z002.4
Form2	22	F2Q22	3	10	C3Q10	x00000204	NSL15E_Z007.2
Form2	23	F2Q23	3	11	C3Q11	x00000205	NSL15E_Z007.3
Form2	24	F2Q24	3	12	C3Q12	x00000207	NSL15E_Z007.4
Form2	25	F2Q25	5	1	C5Q01	x00000386	NSL15E_F001.1
Form2	26	F2Q26	5	2	C5Q02	x00000388	NSL15E_F001.2
Form2	27	F2Q27	5	3	C5Q03	x00002796	NSL15E_F001.4C
Form2	28	F2Q28	5	4	C5Q04	x00000372	NSL15E_M003.1
Form2	29	F2Q29	5	5	C5Q05	x00000373	NSL15E_M003.2
Form2	30	F2Q30	5	6	C5Q06	x00000538	NSL15E_A006.1
Form2	31	F2Q31	5	7	C5Q07	x00000540	NSL15E_A006.3
Form2	32	F2Q32	5	8	C5Q08	x00000475	NSL15E_W006.1
Form2	33	F2Q33	5	9	C5Q09	x00000476	NSL15E_W006.2
Form2	34	F2Q34	5	10	C5Q10	x00000478	NSL15E_W006.3
Form2	35	F2Q35	5	11	C5Q11	x00000397	NSL15E_F004.2
Form2	36	F2Q36	5	12	C5Q12	x00000398	NSL15E_F004.3
Form3	1	F3Q01	3	1	C3Q01	x00000075	NSL15E_C006.2
Form3	2	F3Q02	3	2	C3Q02	x00000073	NSL15E_C006.1
Form3	3	F3Q03	3	3	C3Q03	x00000482	NSL15E_W002.1
Form3	4	F3Q04	3	4	C3Q04	x00000483	NSL15E_W002.2
Form3	5	F3Q05	3	5	C3Q05	x00000484	NSL15E_W002.3
Form3	6	F3Q06	3	6	C3Q06	x00000485	NSL15E_W002.4
Form3	7	F3Q07	3	7	C3Q07	x00000505	NSL15E_W002.5
Form3	8	F3Q08	3	8	C3Q08	x00003651	NSL15E_Z002.2C
Form3	9	F3Q09	3	9	C3Q09	x00000134	NSL15E_Z002.4
Form3	10	F3Q10	3	10	C3Q10	x00000204	NSL15E_Z007.2
Form3	11	F3Q11	3	11	C3Q11	x00000205	NSL15E_Z007.3
Form3	12	F3Q12	3	12	C3Q12	x00000207	NSL15E_Z007.4
Form3	13	F3Q13	4	1	C4Q01	x00000628	NSL12H517.17
Form3	14	F3Q14	4	2	C4Q02	x00000629	NSL12H517.18
Form3	15	F3Q15	4	3	C4Q03	x00000630	NSL12H517.19
Form3	16	F3Q16	4	4	C4Q04	x00000631	NSL12H517.21
Form3	17	F3Q17	4	5	C4Q05	x00000632	NSL12H517.22

Form	Form Q#	Form Code	Cluster	Cluster Q#	Cluster Code	IARS code	Item Name
Form3	18	F3Q18	4	6	C4Q06	x00002815	NSL12H551.51C
Form3	19	F3Q19	4	7	C4Q07	x00000637	NSL12H551.52
Form3	20	F3Q20	4	8	C4Q08	x00000638	NSL12H551.53
Form3	21	F3Q21	4	9	C4Q09	x00000639	NSL12H551.54
Form3	22	F3Q22	4	10	C4Q10	x00000581	NSL06H021.21
Form3	23	F3Q23	4	11	C4Q11	x00000582	NSL06H021.22
Form3	24	F3Q24	4	12	C4Q12	x00000583	NSL06H021.23
Form3	25	F3Q25	6	1	C6Q01	x00000431	NSL15E_H009.1
Form3	26	F3Q26	6	2	C6Q02	x00000437	NSL15E_H009.3
Form3	27	F3Q27	6	3	C6Q03	x00000520	NSL15E_E001.1
Form3	28	F3Q28	6	4	C6Q04	x00000525	NSL15E_E001.4
Form3	29	F3Q29	6	5	C6Q05	x00000066	NSL15E_H004.1
Form3	30	F3Q30	6	6	C6Q06	x00000076	NSL15E_H004.2
Form3	31	F3Q31	6	7	C6Q07	x00000077	NSL15E_H004.3
Form3	32	F3Q32	6	8	C6Q08	x00000093	NSL15E_H002.1
Form3	33	F3Q33	6	9	C6Q09	x00000100	NSL15E_H002.2
Form3	34	F3Q34	6	10	C6Q10	x00000115	NSL15E_H002.3
Form3	35	F3Q35	6	11	C6Q11	x00000545	NSL15E_A007.1
Form3	36	F3Q36	6	12	C6Q12	x00000546	NSL15E_A007.2
Form3	37	F3Q37	6	13	C6Q13	x00000549	NSL15E_A007.8
Form4	1	F4Q01	4	1	C4Q01	x00000628	NSL12H517.17
Form4	2	F4Q02	4	2	C4Q02	x00000629	NSL12H517.18
Form4	3	F4Q03	4	3	C4Q03	x00000630	NSL12H517.19
Form4	4	F4Q04	4	4	C4Q04	x00000631	NSL12H517.21
Form4	5	F4Q05	4	5	C4Q05	x00000632	NSL12H517.22
Form4	6	F4Q06	4	6	C4Q06	x00002815	NSL12H551.51C
Form4	7	F4Q07	4	7	C4Q07	x00000637	NSL12H551.52
Form4	8	F4Q08	4	8	C4Q08	x00000638	NSL12H551.53
Form4	9	F4Q09	4	9	C4Q09	x00000639	NSL12H551.54
Form4	10	F4Q10	4	10	C4Q10	x00000581	NSL06H021.21
Form4	11	F4Q11	4	11	C4Q11	x00000582	NSL06H021.22
Form4	12	F4Q12	4	12	C4Q12	x00000583	NSL06H021.23
Form4	13	F4Q13	5	1	C5Q01	x00000386	NSL15E_F001.1
Form4	14	F4Q14	5	2	C5Q02	x00000388	NSL15E_F001.2
Form4	15	F4Q15	5	3	C5Q03	x00002796	NSL15E_F001.4C
Form4	16	F4Q16	5	4	C5Q04	x00000372	NSL15E_M003.1
Form4	17	F4Q17	5	5	C5Q05	x00000373	NSL15E_M003.2
Form4	18	F4Q18	5	6	C5Q06	x00000538	NSL15E_A006.1
Form4	19	F4Q19	5	7	C5Q07	x00000540	NSL15E_A006.3
Form4	20	F4Q20	5	8	C5Q08	x00000475	NSL15E_W006.1
Form4	21	F4Q21	5	9	C5Q09	x00000476	NSL15E_W006.2
Form4	22	F4Q22	5	10	C5Q10	x00000478	NSL15E_W006.3
Form4	23	F4Q23	5	11	C5Q11	x00000397	NSL15E_F004.2
Form4	24	F4Q24	5	12	C5Q12	x00000398	NSL15E_F004.3
Form4	25	F4Q25	7	1	C7Q01	x00002830	NSL15E_H007.1C
Form4	26	F4Q26	7	2	C7Q02	x00003649	NSL15E_H007.3C
Form4	27	F4Q27	7	3	C7Q03	x00000173	NSL15E_V001.1
Form4	28	F4Q28	7	4	C7Q04	x00000174	NSL15E_V001.2
Form4	29	F4Q29	7	5	C7Q05	x00000428	NSL15E_H008.4
Form4	30	F4Q30	7	6	C7Q06	x00000453	NSL15E_Z027.1
Form4	31	F4Q31	7	7	C7Q07	x00000454	NSL15E_Z027.2
Form4	32	F4Q32	7	8	C7Q08	x00000408	NSL15E_M004.1

Form	Form Q#	Form Code	Cluster	Cluster Q#	Cluster Code	IARS code	Item Name
Form4	33	F4Q33	7	9	C7Q09	x00000410	NSL15E_M004.3
Form4	34	F4Q34	7	10	C7Q10	x00000411	NSL15E_M004.4
Form4	35	F4Q35	7	11	C7Q11	x00000412	NSL15E_M004.5
Form4	36	F4Q36	7	12	C7Q12	x00000393	NSL15E_F003.3
Form4	37	F4Q37	7	13	C7Q13	x00000394	NSL15E_F003.4
Form5	1	F5Q01	5	1	C5Q01	x00000386	NSL15E_F001.1
Form5	2	F5Q02	5	2	C5Q02	x00000388	NSL15E_F001.2
Form5	3	F5Q03	5	3	C5Q03	x00002796	NSL15E_F001.4C
Form5	4	F5Q04	5	4	C5Q04	x00000372	NSL15E_M003.1
Form5	5	F5Q05	5	5	C5Q05	x00000373	NSL15E_M003.2
Form5	6	F5Q06	5	6	C5Q06	x00000538	NSL15E_A006.1
Form5	7	F5Q07	5	7	C5Q07	x00000540	NSL15E_A006.3
Form5	8	F5Q08	5	8	C5Q08	x00000475	NSL15E_W006.1
Form5	9	F5Q09	5	9	C5Q09	x00000476	NSL15E_W006.2
Form5	10	F5Q10	5	10	C5Q10	x00000478	NSL15E_W006.3
Form5	11	F5Q11	5	11	C5Q11	x00000397	NSL15E_F004.2
Form5	12	F5Q12	5	12	C5Q12	x00000398	NSL15E_F004.3
Form5	13	F5Q13	6	1	C6Q01	x00000431	NSL15E_H009.1
Form5	14	F5Q14	6	2	C6Q02	x00000437	NSL15E_H009.3
Form5	15	F5Q15	6	3	C6Q03	x00000520	NSL15E_E001.1
Form5	16	F5Q16	6	4	C6Q04	x00000525	NSL15E_E001.4
Form5	17	F5Q17	6	5	C6Q05	x00000066	NSL15E_H004.1
Form5	18	F5Q18	6	6	C6Q06	x00000076	NSL15E_H004.2
Form5	19	F5Q19	6	7	C6Q07	x00000077	NSL15E_H004.3
Form5	20	F5Q20	6	8	C6Q08	x00000093	NSL15E_H002.1
Form5	21	F5Q21	6	9	C6Q09	x00000100	NSL15E_H002.2
Form5	22	F5Q22	6	10	C6Q10	x00000115	NSL15E_H002.3
Form5	23	F5Q23	6	11	C6Q11	x00000545	NSL15E_A007.1
Form5	24	F5Q24	6	12	C6Q12	x00000546	NSL15E_A007.2
Form5	25	F5Q25	6	13	C6Q13	x00000549	NSL15E_A007.8
Form5	26	F5Q26	1	1	C1Q01	x00000606	NSL06H149.49
Form5	27	F5Q27	1	2	C1Q02	x00000607	NSL06H149.50
Form5	28	F5Q28	1	3	C1Q03	x00000599	NSL09H308.08
Form5	29	F5Q29	1	4	C1Q04	x00000600	NSL09H308.09
Form5	30	F5Q30	1	5	C1Q05	x00000601	NSL09H308.10
Form5	31	F5Q31	1	6	C1Q06	x00000587	NSL06H084.84
Form5	32	F5Q32	1	7	C1Q07	x00000588	NSL06H084.85
Form5	33	F5Q33	1	8	C1Q08	x00000589	NSL06H084.86
Form5	34	F5Q34	1	9	C1Q09	x00000590	NSL06H084.87
Form5	35	F5Q35	1	10	C1Q10	x00000591	NSL06H084.88
Form5	36	F5Q36	1	11	C1Q11	x00000602	NSL09H405.05
Form5	37	F5Q37	1	12	C1Q12	x00000603	NSL09H405.06
Form5	38	F5Q38	1	13	C1Q13	x00000626	NSL12H503.03
Form5	39	F5Q39	1	14	C1Q14	x00000627	NSL12H503.06
Form6	1	F6Q01	6	1	C6Q01	x00000431	NSL15E_H009.1
Form6	2	F6Q02	6	2	C6Q02	x00000437	NSL15E_H009.3
Form6	3	F6Q03	6	3	C6Q03	x00000520	NSL15E_E001.1
Form6	4	F6Q04	6	4	C6Q04	x00000525	NSL15E_E001.4
Form6	5	F6Q05	6	5	C6Q05	x00000066	NSL15E_H004.1
Form6	6	F6Q06	6	6	C6Q06	x00000076	NSL15E_H004.2
Form6	7	F6Q07	6	7	C6Q07	x00000077	NSL15E_H004.3
Form6	8	F6Q08	6	8	C6Q08	x00000093	NSL15E_H002.1

Form	Form Q#	Form Code	Cluster	Cluster Q#	Cluster Code	IARS code	Item Name
Form6	9	F6Q09	6	9	C6Q09	x00000100	NSL15E_H002.2
Form6	10	F6Q10	6	10	C6Q10	x00000115	NSL15E_H002.3
Form6	11	F6Q11	6	11	C6Q11	x00000545	NSL15E_A007.1
Form6	12	F6Q12	6	12	C6Q12	x00000546	NSL15E_A007.2
Form6	13	F6Q13	6	13	C6Q13	x00000549	NSL15E_A007.8
Form6	14	F6Q14	7	1	C7Q01	x00002830	NSL15E_H007.1C
Form6	15	F6Q15	7	2	C7Q02	x00003649	NSL15E_H007.3C
Form6	16	F6Q16	7	3	C7Q03	x00000173	NSL15E_V001.1
Form6	17	F6Q17	7	4	C7Q04	x00000174	NSL15E_V001.2
Form6	18	F6Q18	7	5	C7Q05	x00000428	NSL15E_H008.4
Form6	19	F6Q19	7	6	C7Q06	x00000453	NSL15E_Z027.1
Form6	20	F6Q20	7	7	C7Q07	x00000454	NSL15E_Z027.2
Form6	21	F6Q21	7	8	C7Q08	x00000408	NSL15E_M004.1
Form6	22	F6Q22	7	9	C7Q09	x00000410	NSL15E_M004.3
Form6	23	F6Q23	7	10	C7Q10	x00000411	NSL15E_M004.4
Form6	24	F6Q24	7	11	C7Q11	x00000412	NSL15E_M004.5
Form6	25	F6Q25	7	12	C7Q12	x00000393	NSL15E_F003.3
Form6	26	F6Q26	7	13	C7Q13	x00000394	NSL15E_F003.4
Form6	27	F6Q27	2	1	C2Q01	x00000595	NSL09H360.60
Form6	28	F6Q28	2	2	C2Q02	x00000596	NSL09H360.62
Form6	29	F6Q29	2	3	C2Q03	x00000640	NSL12H559.59
Form6	30	F6Q30	2	4	C2Q04	x00000641	NSL12H559.61
Form6	31	F6Q31	2	5	C2Q05	x00000642	NSL12H559.62
Form6	32	F6Q32	2	6	C2Q06	x00002817	NSL12H564.64C
Form6	33	F6Q33	2	7	C2Q07	x00000645	NSL12H564.65
Form6	34	F6Q34	2	8	C2Q08	x00003646	NSL12H564.66C
Form6	35	F6Q35	2	9	C2Q09	x00000648	NSL12H564.69
Form6	36	F6Q36	2	10	C2Q10	x00000649	NSL12H564.70
Form6	37	F6Q37	2	11	C2Q11	x00000592	NSL06H041.41
Form6	38	F6Q38	2	12	C2Q12	x00000593	NSL06H041.44
Form7	1	F7Q01	7	1	C7Q01	x00002830	NSL15E_H007.1C
Form7	2	F7Q02	7	2	C7Q02	x00003649	NSL15E_H007.3C
Form7	3	F7Q03	7	3	C7Q03	x00000173	NSL15E_V001.1
Form7	4	F7Q04	7	4	C7Q04	x00000174	NSL15E_V001.2
Form7	5	F7Q05	7	5	C7Q05	x00000428	NSL15E_H008.4
Form7	6	F7Q06	7	6	C7Q06	x00000453	NSL15E_Z027.1
Form7	7	F7Q07	7	7	C7Q07	x00000454	NSL15E_Z027.2
Form7	8	F7Q08	7	8	C7Q08	x00000408	NSL15E_M004.1
Form7	9	F7Q09	7	9	C7Q09	x00000410	NSL15E_M004.3
Form7	10	F7Q10	7	10	C7Q10	x00000411	NSL15E_M004.4
Form7	11	F7Q11	7	11	C7Q11	x00000412	NSL15E_M004.5
Form7	12	F7Q12	7	12	C7Q12	x00000393	NSL15E_F003.3
Form7	13	F7Q13	7	13	C7Q13	x00000394	NSL15E_F003.4
Form7	14	F7Q14	1	1	C1Q01	x00000606	NSL06H149.49
Form7	15	F7Q15	1	2	C1Q02	x00000607	NSL06H149.50
Form7	16	F7Q16	1	3	C1Q03	x00000599	NSL09H308.08
Form7	17	F7Q17	1	4	C1Q04	x00000600	NSL09H308.09
Form7	18	F7Q18	1	5	C1Q05	x00000601	NSL09H308.10
Form7	19	F7Q19	1	6	C1Q06	x00000587	NSL06H084.84
Form7	20	F7Q20	1	7	C1Q07	x00000588	NSL06H084.85
Form7	21	F7Q21	1	8	C1Q08	x00000589	NSL06H084.86
Form7	22	F7Q22	1	9	C1Q09	x00000590	NSL06H084.87

Form	Form Q#	Form Code	Cluster	Cluster Q#	Cluster Code	IARS code	Item Name
Form7	23	F7Q23	1	10	C1Q10	x00000591	NSL06H084.88
Form7	24	F7Q24	1	11	C1Q11	x00000602	NSL09H405.05
Form7	25	F7Q25	1	12	C1Q12	x00000603	NSL09H405.06
Form7	26	F7Q26	1	13	C1Q13	x00000626	NSL12H503.03
Form7	27	F7Q27	1	14	C1Q14	x00000627	NSL12H503.06
Form7	28	F7Q28	3	1	C3Q01	x00000075	NSL15E_C006.2
Form7	29	F7Q29	3	2	C3Q02	x00000073	NSL15E_C006.1
Form7	30	F7Q30	3	3	C3Q03	x00000482	NSL15E_W002.1
Form7	31	F7Q31	3	4	C3Q04	x00000483	NSL15E_W002.2
Form7	32	F7Q32	3	5	C3Q05	x00000484	NSL15E_W002.3
Form7	33	F7Q33	3	6	C3Q06	x00000485	NSL15E_W002.4
Form7	34	F7Q34	3	7	C3Q07	x00000505	NSL15E_W002.5
Form7	35	F7Q35	3	8	C3Q08	x00003651	NSL15E_Z002.2C
Form7	36	F7Q36	3	9	C3Q09	x00000134	NSL15E_Z002.4
Form7	37	F7Q37	3	10	C3Q10	x00000204	NSL15E_Z007.2
Form7	38	F7Q38	3	11	C3Q11	x00000205	NSL15E_Z007.3
Form7	39	F7Q39	3	12	C3Q12	x00000207	NSL15E_Z007.4
InqTask1	1	I1Q01	C	1	CCQ01	x00000236	NSL15Esi3-1.1
InqTask1	2	I1Q02	C	2	CCQ02	x00000237	NSL15Esi3-1.2
InqTask1	3	I1Q03	C	3	CCQ03	x00000239	NSL15Esi3-1.4
InqTask1	4	I1Q04	C	4	CCQ04	x00000241	NSL15Esi3-2.1
InqTask1	5	I1Q05	C	5	CCQ05	x00000243	NSL15Esi3-3.1
InqTask1	6	I1Q06	C	6	CCQ06	x00000327	NSL15Esi3-4.1
InqTask1	7	I1Q07	C	7	CCQ07	x00000245	NSL15Esi3-5.1
InqTask1	8	I1Q08	C	8	CCQ08	x00000246	NSL15Esi3-5.2
InqTask1	9	I1Q09	C	9	CCQ09	x00000247	NSL15Esi3-5.3
InqTask1	10	I1Q10	C	10	CCQ10	x00000578	NSL15Esi3-5.5
InqTask2	1	I2Q01	D	1	CDQ01	x00000249	NSL15Esi5-1.1
InqTask2	2	I2Q02	D	2	CDQ02	x00000579	NSL15Esi5-1.2
InqTask2	3	I2Q03	D	3	CDQ03	x00000250	NSL15Esi5-2.1
InqTask2	4	I2Q04	D	4	CDQ04	x00000251	NSL15Esi5-3.1
InqTask2	5	I2Q05	D	5	CDQ05	x00000252	NSL15Esi5-4.1
InqTask2	6	I2Q06	D	6	CDQ06	x00000253	NSL15Esi5-4.2
InqTask2	7	I2Q07	D	7	CDQ07	x00000254	NSL15Esi5-4.3
InqTask2	8	I2Q08	D	8	CDQ08	x00000580	NSL15Esi5-4.4
InqTask2	9	I2Q09	D	9	CDQ09	x00003657	NSL15Esi5-5.1C
InqTask2	10	I2Q10	D	10	CDQ10	x00003659	NSL15Esi5-5.2C
InqTask2	11	I2Q11	D	11	CDQ11	x00000257	NSL15Esi5-5.3
Practice	1	PcQ01	P	1	CPQ01	x00003623	NSL15EPrP1.1
Practice	2	PcQ02	P	2	CPQ02	x00003624	NSL15EPrP1.2
Practice	3	PcQ03	P	3	CPQ03	x00003625	NSL15EPrP1.3
Practice	4	PcQ04	P	4	CPQ04	x00003628	NSL15EPrP1.4
Practice	5	PcQ05	P	5	CPQ05	x00003627	NSL15EPrP2.1
Practice	6	PcQ06	P	6	CPQ06	x00003630	NSL15EPrP2.2
Practice	7	PcQ07	P	7	CPQ07	x00003638	NSL15EPrP2.3
TRTPrePrac	1	TcQ01	T	1	CTQ01	x00003561	NSL15EPrT1.1
TRTPrePrac	2	TcQ02	T	2	CTQ02	x00003565	NSL15EPrT1.2
TRTPrePrac	3	TcQ03	T	3	CTQ03	x00003569	NSL15EPrT1.3
TRTPrePrac	4	TcQ04	T	4	CTQ04	x00003572	NSL15EPrT2.1
TRTPrePrac	5	TcQ05	T	5	CTQ05	x00003574	NSL15EPrT2.2
TRTPrePrac	6	TcQ06	T	6	CTQ06	x00003576	NSL15EPrT2.3
TRTPrePrac	7	TcQ07	T	7	CTQ07	x00003578	NSL15EPrT2.4

Form	Form Q#	Form Code	Cluster	Cluster Q#	Cluster Code	IARS code	Item Name
TRTPrePrac	8	TcQ08	T	8	CTQ08	x00003581	NSL15EPrT3.1
TRTPrePrac	9	TcQ09	T	9	CTQ09	x00003585	NSL15EPrT3.2
TRTPrePrac	10	TcQ10	T	10	CTQ10	x00003592	NSL15EPrT3.4
Survey	1	SyQ01	S	1	CSQ01	x00000358	NSL12ESG01-04
Survey	2	SyQ02	S	2	CSQ02	x00000359	NSL12ESG05-07
Survey	3	SyQ03	S	3	CSQ03	x00000360	NSL12ESG08-11
Survey	4	SyQ04	S	4	CSQ04	x00000361	NSL12ESG12-16
Survey	5	SyQ05	S	5	CSQ05	x00000362	NSL12ESG17-19
Survey	6	SyQ06	S	6	CSQ06	x00000363	NSL12ESG20-21
Survey	7	SyQ07	S	7	CSQ07	x00000365	NSL12ESG22-25
Survey	8	SyQ08	S	8	CSQ08	x00000366	NSL12ESG26-29
Survey	9	SyQ09	S	9	CSQ09	x00000367	NSL12ESG30-30
Survey	10	SyQ10	S	10	CSQ10	x00000368	NSL12ESG31-32
Survey	11	SyQ11	S	11	CSQ11	x00000552	NSL13ESG01-05
Survey	12	SyQ12	S	12	CSQ12	x00000554	NSL13ESG06-08