NATIONAL ASSESSMENT PROGRAM – LITERACY AND NUMERACY

Online Assessment Research Development Study 2014 Cognitive Interviews: Challenging Items (Numeracy)

2014

NAPLAN NATIONAL ASSESSMENT PROGRAM Literacy and Numeracy

acara AUSTRALIAN CURRICULUM, ASSESSMENT AND REPORTING AUTHORITY

Online Assessment Research
Development Study 2014 Cognitive Interviews: Challenging Items (Numeracy)

Project management
ACARA's online research program is directed by Dr Goran Lazendic. The National Assessment and Surveys Online Program is managed by Julie-Anne Justus, John Skehan, Dominik Swierk, Justine Tyrrell and Rosemary Vo. Item and test development is managed by Anna Cohen (reading), Stephen Phillip (numeracy) and Ross Bindon (conventions of language).

Report author(s)
Ajay Ramful, Tom Lowrie and Tracy Logan, Faculty of Education, Science, Technology and Mathematics, University of Canberra, working with ACARA staff (Dr Goran Lazendic and Stephen Phillip), delivered the student and school interviews and developed the report entitled *NAPLAN Online 2014 Development Study: Cognitive interview research activity: Perceived difficulty of challenging Numeracy items.*

National Assessment and Surveys Online Program
The National Assessment and Surveys Online Program, funded by the Australian Government, is designed to deliver national assessments and surveys online. ACARA is responsible for planning and implementing a clearly defined assessment and reporting research agenda that will allow reporting to the Education Council on issues and options for delivering NAPLAN online. A key aspect of the program is ACARA's expanded assessment and reporting research agenda, incorporating a comprehensive investigation into assessment instruments and programs using online technology.

Acknowledgements

The appropriate citation for this report is:
Ramful, A, Lowrie, T and Logan, T, *NAPLAN Online 2014 Development Study: Cognitive interview research activity: Perceived difficulty of challenging Numeracy items.* Faculty of Education, Science, Technology and Mathematics, University of Canberra, 22 January 2015. Prepared for the Australian Curriculum, Assessment and Reporting Authority.

Project report

NAPLAN Online 2014 Development Study: Cognitive interviews research activity 2a: Perceived difficulty of challenging Numeracy items

Client

Australian Curriculum, Assessment and Reporting Authority

Date

22 January 2015

Contact

Professor Tom Lowrie
Centenary Professor
Faculty of Education, Science, Technology and Mathematics
University of Canberra
ACT 2601 Australia

Tel: +61 (0)2 6201 2568
E-mail: Thomas.Lowrie@canberra.edu.au

Report developed by Ajay Ramful, Tom Lowrie and Tracy Logan, University of Canberra, Australia.

UNIVERSITY OF CANBERRA

Table of Contents

Executive Summary

Background

NAPLAN Online 2014 Development Study: Cognitive interviews research activity 2a: Perceived difficulty of challenging Numeracy items

This report acknowledges ACARA's intent to establish a rigorous world-class curriculum and assessment program, and their research agenda associated with movement of this assessment program to online delivery. To this end, this report describes the findings of a research project that examined mathematically capable students' capacity to engage with challenging numeracy tasks presented in a digital (online) form.

Fundamentally, the project was concerned with how mathematically capable students engaged with mathematics items that were designed to challenge and elicit the use of higher-order thinking skills. Within the tailored-test design of the online assessment, the focus of this report is limited to items within Testlet F. Such items are not present in current NAPLAN tests and as such the requirement was to comprehensively investigate the functioning of these numeracy items.

Cognitive interviews were used to capture a rich data source of students' mathematics engagement across Testlet F by analysing students' cognitive and behavioural engagement on challenging numeracy items. Additionally, we determined specific aspects of the Testlet that enabled or constrained access to these items in order to make better sense of behavioural patterns.

Our analysis was undertaken utilising two frameworks. The first of these was Mayer's (2002) *A Taxonomy for Computer-Based Assessment of Problem Solving*, which was employed to identify whether the design of the items provided adequate testing context for these capable students. The second was Peressini and Webb's (1999) Analytic Mathematics Scoring Framework, which provided constructs to analyse students' knowledge, thinking skills and solution strategies in relation to the numeracy items.

In order to ascertain the relative difficulty these students encountered when solving questions from Testlet F, data were compared to that of a parallel study that included a more diverse cohort of students. The cohort reported on here would be considered representative of those students in the general population.

The main goal of the research project was to determine the extent to which mathematically capable students could solve numeracy items that were designed to be challenging. These students were able to access and engage with most of the questions presented in Testlet F. In fact, the knowledge and strategies demonstrated by the students indicated that the questions in Testlet F were fundamentally appropriate for students with higher levels of mathematics ability.

The findings are especially worthwhile in understanding and valuing the higher-order thinking skills that mathematically capable students employ when faced with challenging numeracy items.

Priority Areas

In order to address the scope and intent of the research design, three priority areas were identified. These priorities formed the basis of methodological design and data analysis.

Priority 1: Establish the extent to which the proposed challenging items in Testlet F provide adequate testing context for highly capable students.

Priority 2: Examine the performance of such students on Testlet F of the tailored test.

Priority 3: Monitor the students' knowledge, thinking skills and strategy use, and how these relate to the intended assessment outcomes envisaged by the test developers.

Research questions

There were seven research questions posed for the project.

Priority 1:

1.1 Are there design considerations which inhibit or enable these students to engage with these items in a meaningful way?

1.2 What design elements most impact on student access and performance?

Priority 2

2.1 How do highly capable students perform on the Testlet F items?

2.2 What performance characteristics are common among these students?

Priority 3

3.1 What types of problem-solving knowledge, skills and strategies do students utilise when solving challenging items?

3.2 What understandings and strategy use most influence performance?

3.3 What perceptions do the students have of Testlet F items? How do these identify with the intended assessment outcomes?

Key Findings (KF) and Recommendations (R)

Priority 1: Establish the extent to which the proposed challenging items in Testlet F provide adequate testing context for highly capable students.

KF1.1 These items constructed for Testlet F were designed to evoke high-order thinking. This is achieved by creating items that require the application and/or analysis of conventional mathematical concepts and procedures. In general, the language used to formulate the questions was clear. The main characteristic of poor design was associated with inconsistencies in graphic clarity and font size.

The items in Testlet F cover a broad range of areas across the mathematics curriculum. Further, Mayer's framework (2002) showed that for each of the four year levels, the majority of items require conceptual knowledge in contrast to purely mechanical procedural knowledge. In terms of cognitive processing, most of the items require higher-order skills such as analysis and application. Our detailed analysis of the structure of the problems showed that some of the tasks were however more procedural than conceptual.

R1.1 The majority of items across the four year levels tend to be concentrated on Number and Algebra and Measurement and Geometry. A more balanced Testlet F should include more items from the strand Statistics and Probability.

KF1.2 Some inadvertent effects were observed emanating from the visual weight of the graphics in the items in a digital environment. For instance, at Year 3 level, some of the graphics imposed additional demands for the students. Counting the small subdivisions on the screen was quite time consuming both for the Year 3 and Year 5 students. Some of the students counted the subdivisions by pointing their pencils to the screen. Similarly, finding the area of an irregular polygon by counting squares on the computer screen may be demanding. One has to mentally or visually split the region and keep track of the different areas as in the last item in Year 9. Further, the fonts used in some of the items were not always homogeneous, prompting students to focus on particular aspects of the problem and ignoring other important ones.

R1.2 It is recommended that the display of information in item design be tightly scrutinised to ensure that inadvertent effects do not bring additional load in the item. Counting subdivisions on the computer screen may be time consuming and it suggested that some form of online markers be made available to the students to help them benchmark their counting (on the computer screen).

KF1.3 Many students did not notice that a calculator was available for some of the items. The interviewers often had to tell the participants that a calculator was included (whenever the calculator first appeared in the Testlet). Another technical issue is that when the calculator opens, it covers the question or part of the question. We had to tell the students that the calculator could be moved. Further, the keys on the calculator as well as its functioning did create some issues.

R1.3 The calculator should be made more visible on the screen and its keys and operational capacities improved. As a more general recommendation, we would suggest that a tutorial be included (in the form of a demonstration) at the start of the test to remind students about the availability of resources in the online environment. For instance, the demonstration may illustrate how to move the calculator. A more direct suggestion is to allow the calculator to pop up on the side of the screen for questions where it is allowed to be used. The tutorial may serve other functions such as informing students whether they can move backwards to check their answers. In some of the problems, the answer space accepts only decimals and not fractions and the tutorial can be very useful here.

Priority 2: Examine the performance of such students on Testlet F of the tailored test.

KF2.1 These capable students demonstrated both flexibility and fluency in their problem solving. These skills were necessary to engage with these tasks since most items required students to use application and analysis processes. It was evident that some of the Year 3 students lacked the capacity to think flexibly when compared to the other three cohorts. This is unsurprising, given the students maturity.

R2.1 We commend the fact that the task designers were able to construct items that required higher-order cognitive processes. We encourage ACARA to maintain this design "stance" and not simply increase task complexity via increased content difficulty.

Priority 3: Monitor the students' knowledge, thinking skills and strategy use, and how these relate to the intended assessment outcomes envisaged by the test developers.

KF3.1 The students employed a range of problem-solving strategies to solve the items. The students used diagrams, sketches and tables, looked for patterns and other non-algorithmic approaches to make their way through the problems. Such analytical skills are crucial in dealing with novel situations, characteristic of several Testlet F items. The students particularly exploited their knowledge of numbers in either written or mental forms to open solution paths. The Testlet F items prompted them to coordinate multiple pieces of information in the various multi-stage problems. Often, the same problem was solved in different ways by the students. Four types of higher-level mathematical reasoning were particularly apparent from the students' responses, namely inductive/deductive reasoning, proportional reasoning, spatial reasoning and to a lesser extent, algebraic reasoning.

R3.1 Test designers may not always look at the mode of reasoning in designing items. They may be more inclined to focus on content and the structural nuances that elicit higher-order thinking. We would suggest that Testlet F include more items associated with proportional reasoning and pre-algebraic/algebraic reasoning. These forms of reasoning are known to be demanding for elementary and middle school students.

KF3.2 Typically these students possessed sound generic problem-solving skills but may have lacked foundational knowledge. Relatively elementary concepts such as decimals, diagonals or performing a division at times prevented the students from getting the correct answer. Consequently, errors occurred when the students lacked flexibility to use higher-order reasoning such as inductive, proportional and spatial reasoning. We appreciate the tension between introducing more difficult content knowledge and exposing students to high-order thinking.

R3.2 From a research perspective, it would be prudent of ACARA to explicitly assess foundations skills in easier testlets. The application of these concepts could then be assessed via higher-order problem solving in Testlet F. Such a design would provide both diagnostic and research-based opportunities for a range of stakeholders.

KF3.3 Our observations indicated that many students encountered challenges in solving the spatial tasks. This is an important finding especially since NAPLAN items seem to give much consideration to this form of reasoning.

R3.3 ACARA be encouraged to ensure that the numeracy assessment frameworks point to the importance of visual and spatial reasoning in solving graphics-based NAPLAN tasks. This is increasingly important given the move to online assessment.

KF3.4 Estimation and systematic guess-and-check strategies were observed across all four year levels. On the one hand, the use of these more primitive strategies suggests that the students were exposed to challenging tasks. On the other, these fall back and non-deterministic strategies also suggest that these students are yet to develop more advanced mathematical abilities.

R3.4 Findings suggest that these students are able to produce correct answers despite not drawing on the higher-order problem-solving skills the tasks were designed to elicit. Consequently, reporting may need to be based on understanding of concepts and possible strategies used by students to solve the tasks, including the more primitive strategies that these students often fall back to.

KF3.5 During the cognitive interviews, we could observe that the students took much time to work out some of the items and to complete the Testlet as a whole. For some of the questions they took over five minutes.

R3.5 It should be recognised that students will attempt these higher-order questions after completing two testlets. Thus the number of items to be included in Testlet F needs to be judiciously thought through, as well as the time allocated. We recommend that less is best.

KF3.6 We had to remind some of the students that paper-and-pencil was available as they tried to solve the problems mentally. Some of the students tend not to use paper-and-pencil. On the other hand, we encountered students who would write their stepwise methods for each problem on paper. Moreover, at times, some students tend to rush through the problems and do not spend time verifying the soundness of their answers. In the cognitive interviews, often students changed their incorrect answers when we asked them to explain their thinking.

R3.6 Solving mathematics tasks in an online environment requires sophisticated levels of digital literacy. Students need to be exposed to these skills in the classroom otherwise, assessment results could be problematic. It is necessary for students to develop fluency within the digital assessment environment.

KF3.7 In general, the students regarded the items to be ‘easy’ or ‘moderately difficult’. Not surprisingly, the students had a positive affect towards mathematics. Interestingly, this was the case even though their answers were not always correct. Perception is closely associated to success. It is encouraging to see that students rated the majority of the items within their reach.

KF3.8 At Year 3 and 5, students tend to prefer doing the test on computer while at Year 7 and 9 there was a preference for paper-and-pencil mode. The students felt that they have more flexibility in working around the problem on paper than on computer. They can concentrate more on paper-and-pencil, and interpreting the problem on the computer requires more effort. Those who preferred a computer-based test did so because they think that it is easier to key-in and change the answer on the computer than on paper.

R3.8 It is important to be mindful of students’ preferences for test mode since it may inadvertently influence performance. A growing body of research literature (including our own work) shows that test mode does influence performance (e.g., Bennett et al., 2008; Lowrie & Logan, 2015). We suggest that ACARA continue research in this area, especially differences between student performance in pencil-and-paper and digital forms.

Background

Terms of Reference

This project investigated the cognitive and behavioural engagement of students with NAPLAN Numeracy items delivered within the new tailored (multi-stage) test design, with a particular focus on Testlet F, as requested in the Purpose Statement.

This investigation:

a. Established the extent to which the proposed challenging items in Testlet F provide adequate testing context for highly capable students.

b. Examined the performance of such students on Testlet F of the tailored test.

c. Monitored the students' knowledge, thinking skills and strategy use, and how these relate to the intended assessment outcomes envisaged by the test developers.

Research Design

To understand the ways in which the high-performing students engaged with the Testlet F items and to identify patterns of behaviour, a total of 52 cognitive interviews were conducted in Years 3, 5, 7 and 9. Our methodological approach involved both quantitative and qualitative analysis of students' responses. The quantitative analysis examined the extent to which the proposed challenging items provided adequate testing context for highly capable students as well as the performance of the students on those items. The qualitative analysis scrutinised the knowledge, thinking skills and strategies used by the high-performing students from a cognitive perspective. Similarly, the one-to-one, hour-long interviews enabled us to map the behavioural engagement of the students in an online environment.

Participants

Five schools participated in the research (see Table 1). All schools were situated in the ACT region and were from both public and private administrations.

Table 1: Participating Schools, Number of Students and Data Collection Schedule

Date	School	No. of students N = 52			
		Year 3	Year 5	Year 7	Year 9
12/09/2014	School A	6	6	—	—
16/10/2014	School B	5	5	—	—
20/10/2014	School C	—	—	5	5
22/10/2014	School D	5	5	—	—
29/10/2014	School E	—	—	5	5

Data collection instruments

Students were individually interviewed and their responses were video recorded to allow for retrospective analysis. The videos provided evidence as to the level of engagement the students had with the items, both in terms of strategies used to solve the mathematics tasks, as well as the constraints that they encountered or the errors that they made.

We also designed an accompanying open-ended questionnaire/observational grid based on the items in Testlet F to capture as much information as possible during the one-hour interview. Prior to the design of the instrument, we analysed each item in Testlet F with regard to its mathematical requirement. The open-ended questionnaire/observational grid consisted of four parts:

1. Profiling information: This first section attempted to gather data on the students' profile with regard to past performance in mathematics and type of mathematical experiences that they have at school and at home.
2. Specific questions based on individual items: The second section probes the strategies used by the students and records any relevant mathematical behaviour (e.g., visualisation actions used to solve the spatial tasks).
3. Attitudinal questions: The third section attempted to capture how the students evaluate the difficulty of the item. Each student was asked to evaluate the individual Testlet F item after s/he had solved it.
4. Students' perception of computer-based tests: After completing Testlet F, the student was required to explain how taking a mathematics test was similar or different on the computer compared to paper-and-pencil.

The interviews were conducted by four experienced interviewers who had been working with children across primary and secondary schools. Students' working out was kept as evidence of their thinking process and the struggles that they may have experienced. Although one hour was allocated to the interviews, often it would take longer. We chose to give students as much time as possible to understand the challenges that they encountered and to gather as much information about their views on the Testlet F items in a computer-based environment. Following one set of interviews, members of our team held discussions about the salient findings which enabled us to identify patterns of behaviour.

Reliability and validity

This study involved a substantial amount of qualitative data. Fifty-two hour-long cognitive interviews were conducted on an individual basis. Given that four experienced interviewers were involved, the peer debriefing allowed collective discussion about patterns and trends in the data. Such corroboration of data increased the credibility of the qualitative findings. The data has been systematically assembled to derive interpretations into structurally coherent and corroborating wholes.

Analysis of data: Analytical frameworks

The video records were analysed with respect to the engagement of the students with the items in an online environment. We specifically noted the strategies that the students used to solve the items. The data recorded in the open-ended questionnaire/observational grid were compiled in an Excel file to draw out patterns of behaviour and performance on the items across the four years. Two analytical frameworks were used to scrutinise the data. The first framework, Mayer's (2002) *A Taxonomy for Computer-Based Assessment of Problem Solving*, is based on content analysis and looks at the nature of tasks in terms of types of knowledge and cognitive processes. The second framework, Peressini and Webb's (1999) Analytic Mathematics Scoring Framework, provided constructs to analyse students' responses to the mathematical tasks. The dimensions of each framework are outlined below.

Mayer's framework for analysing the design of tasks

Mayer's framework consisted of four types of knowledge and six types of cognitive processes. We interpret the framework in mathematical problem solving as follows:

Four types of knowledge	
Factual knowledge	Knowledge of terminology and specific facts. For instance, knowing addition and multiplication facts.
Procedural knowledge	Knowledge of algorithms and procedures (e.g., knowing that in the multiplication of two fractions, numerators and denominators are multiplied).
Conceptual knowledge	Knowledge of interrelationship among concepts, knowledge of principles and models (e.g., recognising that repeated addition is the same as multiplication).
Metacognitive knowledge	"It includes knowing strategies for how to accomplish tasks, knowing about the demands of various tasks, and knowing one's capabilities for accomplishing various tasks" (Mayer, 2002, p. 626).

Six types of cognitive processes	
Remember	Recognising and/or recalling information from long term memory.
Understand	The meaning and sense making associated with interpreting, classifying, inferring and comparing.
Apply	The application of executing or implementing a procedure in a problem situation.
Analyse	Involves differentiating, organising or attributing essential information and working with the relation among these parts to solve the problem.
Evaluate	The verification of the soundness of an approach used to solve a problem.
Create	Assembling parts of a problem situation together to find the solution.

Peressini and Webb's framework for analysing students' responses

We adapted Peressini and Webb's (1999) framework to analyse the processes used by the students in solving the tasks. The framework consists of three main categories of constructs, namely

Foundational Knowledge, Solution Process, and Communication. This report focuses on the first two dimensions of the framework, namely Foundational Knowledge and Solution Process, in line with our objectives in Priority 3. These two main components are outlined below.

I. *Foundational Knowledge* – Foundation of mathematical working knowledge possessed by the individual that is brought to bear on the assessment situation:

A. Concepts, facts, and definitions

B. Procedures and algorithms

C. Misconceptions

The above constructs prompt us to assess the extent to which the sampled students possess the requisite mathematical knowledge to access the tasks in Testlet F. Thus, as we analysed the responses of the students we paid particular attention to their knowledge of numbers (including fractions, decimals, percentages), their fluency in applying the arithmetic operations, in using ratios and so forth, as they solve the challenging items. Similarly, we assess their knowledge of geometrical concepts and observe in what ways these permit or hinder access to the problems.

II. *Solution Process* – The analytical skills and reasoning abilities that the student demonstrated in solving the assessment task.

In assessing the extent to which the items in Testlet F are accessible to students, it was important to analyse how they interpreted the items and what types of problem-solving strategies they used in solving the tasks. To this effect, we adapted the following constructs from Peressini and Webb (1999).

A. Analytical skills: These refer to the ways in which the problem solver understands a problem in terms of such things as identifying the quantities and their interrelationships, drawing a sketch, diagram or table, examining special cases, decoding a graphic. It also involves solution procedures such as using a heuristic, estimation, systematic guess-and-check strategy, working backwards. Analytical skills also include the verification of the solution after solving the problem.

B. Reasoning abilities: The modes of reasoning the student exhibits while approaching and solving the tasks (includes tendencies associated with the student's reasoning capabilities).

It was also important to look at the type of reasoning abilities as this potentially showed the type of cognitive functioning that allowed or constrained students from solving the higher-level tasks. In particular, we focused on the following forms of reasoning: Inductive (patterns), Deductive, Spatial, Proportional and Abstracting.

Higher-order mathematical knowledge and skills

The following dimensions of higher-order thinking (and by implication high-order tasks) were adapted from Resnick (1987). In assessing whether the tasks elicit higher-order knowledge and skills, we look at the cognitive demand that they impose on the problem solver. The cognitive demand of tasks can be viewed in terms of whether they involve merely the memorisation of learnt facts, rules, formulae or procedures that are algorithmic as opposed to tasks that involve complex and non-algorithmic thinking where the pathway to the problem is not explicit (Smith & Stein, 1998). The following table highlights the dimensions of higher-order thinking utilised in this report:

Cognitive demand	Description
Prompting non-algorithmic approach	High-order tasks require students to improvise ways of approaching problem situations beyond the application of learnt procedures. In other words, the path to the solution is not fully specified in advance. They have an inherent level of productive ambiguity.
Prompting the coordination of multiple pieces of information, requiring students to appreciate subtleties in the problem statement	Higher-order thinking is elicited when students have to process multiple pieces of information to set relationship(s) among problem parameters. In the process, it also requires students to make the fine distinctions in the problem statement.
Providing opportunities for deploying multiple strategies	High-order problems can often be solved in multiple ways. Thus, they create opportunities for students to harness different strategies to navigate situations that do not have a predefined trajectory.
Prompting higher level of mathematical reasoning	The following forms of reasoning serve to describe the type of processes that students may use in the solution of mathematical problems and characterise the level of mathematical thinking.
	Inductive/deductive reasoning: Inductive reasoning requires students to generalise an observation. Thus, it may elicit higher-order thinking. Complementarily, deductive reasoning involves thinking from a general statement or formula to reach a particular conclusion.
	Proportional reasoning: In simple terms, proportional reasoning involves the ability to compare ratios and to predict or produce equivalent ratios (Lamon, 2007). Tasks which involve proportional reasoning are potentially capable of eliciting some degree of higher-order thinking.
	Spatial reasoning: Among others, spatial reasoning is basically related to visualising shapes or objects, to carry out mental transformations, to position oneself relative to other objects in the environment. It is known to be a demanding form of thinking.
	Algebraic reasoning: This form of reasoning involves working with a place holder, symbols or notations in articulating relationships among quantities. More importantly, it involves working with unknowns.

When students are out of resources as a result of the cognitive demand in a problem-solving situation, they are led to exploit more intuitive strategies such as systematic guess-and-check and estimation procedures. Thus, the use of these intuitive strategies may also suggest that the tasks are demanding in that the routine mathematical procedures may not be sufficient.

The following sections address the three research priority areas.

Priority 1: Establish the extent to which the proposed challenging items in Testlet F provide adequate testing context for highly capable students.

Research question 1.1: Are there design considerations which inhibit or enable these students to engage with these items in a meaningful way?

Research question 1.2: What design elements most impact on student access and performance?

Item Design in Terms of Knowledge and Cognitive Processes

The design of items can be an influential variable in students' problem-solving success. The problem context, the language used (e.g., syntactic and semantic aspects), the numbers used, the graphic included, the presentation of the task, the order of problem complexity among others influence the sense that a student may make of a given mathematical situation (Diezmann & Lowrie, 2012). Mathematics education research has clearly established the range of factors that influence students' success in solving problems. Thus, in understanding the design of the items in Testlet F, we considered the assessment purpose of each item from the content, numeracy element and structure perspectives. We then utilised Mayer's matrix to understand how the items elicited higher-order thinking. We also considered the task representation and design features such as the graphics and how they were situated in the digital environment, along with any literacy demands. The following section presents our findings according to year level.

Year 3

Table 2 presents the Year 3 items in Testlet F in terms of their curricular strand, related numeracy element and structure of the tasks. Seven items are from the strand 'Number and Algebra' while five items are from 'Measurement and Geometry', with no tasks from 'Statistics and Probability'. In terms of mathematical topics, the items cover a range of concepts: types of numbers, fractions, pattern, multiplication and division, length measurement, time, net and area. The word problems are set in a meaningful context. Although the concepts may be familiar to Year 3 students, it is the structure of the problems that direct them to a higher level (compared to what is conventionally covered at the Year 3 level). For instance, in a normal Year 3 mathematics curriculum, students are more likely to read time rather than work with time intervals in a digital and analogue clock as in Question 2. Similarly, the Fibonacci sequence, with unknown first term in Question 9, makes the problem challenging for a Year 3 student who has to think in reverse to deduce that the first number can deterministically be obtained by subtraction.

Table 2. Year 3 Items Classified by Content Strand, ACARA Numeracy Elements and Item Description

Item	Strand	Numeracy continuum	Concepts assessed and structure of problem
1	Number and Algebra	Estimating and calculating with whole numbers	Two-step word problem involving multiplication (or repeated addition)
2	Measurement and Geometry	Using measurement	Reading and subtracting time on a digital and analogue clock
3	Measurement and Geometry	Using spatial reasoning	Matching a relatively complex net of a cube to its 3D representation
4	Number and Algebra	Recognising and using pattern and relationships	Finding a missing number in a number pattern involving a constant difference and relatively small numbers
5	Measurement and Geometry	Using spatial reasoning	Rotation/Turning a grid involving four objects
6	Measurement and Geometry	Using measurement	Reading a ruler (height chart) involving metre and centimetre where the starting point of the ruler is not shown
7	Number and Algebra	Estimating and calculating with whole numbers	Identifying the length of a subdivision on a number line scaled from 0 to 1000.
8	Number and Algebra	Estimating and calculating with whole numbers	Successive computations involving an even number, halving, addition, multiplication and division
9	Number and Algebra	Recognising and using pattern and relationships	Number pattern involving a Fibonacci sequence with first term unknown
10	Number and Algebra	Using fractions, decimals, percentages, ratios and rates	Operating a simple fraction (quarter) on a quantity and finding its equivalence to half of an unknown quantity
11	Measurement and Geometry	Using measurement; Using spatial reasoning	Finding area on grid involving squares and triangles
12	Number and Algebra	Estimating and calculating with whole numbers	Two-step word problem involving multiplication (or repeated addition) and division

Based on Mayer's (2002) framework, most of the items required conceptual knowledge (see Table 3). Similarly, in terms of cognitive processes that may be required to solve the items, the majority of the items fell in the category 'Apply' and 'Analyse'. This shows that from a theoretical perspective, the items in Testlet F are at a high processing level for Year 3 students.

Table 3. Classification of Year 3 Items Within Mayer's Matrix

Types of knowledge	Type of cognitive process					
	Remember	Understand	Apply	Analyse	Evaluate	Create
Factual						
Procedural			8			
Conceptual			1,6,9,	2,3,4,5,6, 7,10,11,1 2		
Metacognitive						

The structure of the tasks

Number and Algebra

The seven items (Questions 1, 4, 7, 8, 9, 10 and 12) in the Number and Algebra strand elicit higher levels of thinking. For example, Question 1 is pitched to a higher level than a conventional multiplication task since it involves the coordination of three quantities. A routine Year 3 multiplication problem generally involves only two quantities. Question 12 is an excellent item to see if students can relate the multiplication and division operations to a problem situation. Although students are familiar with reading number lines and skip counting, Question 7 puts them in a challenging situation where the start and end numbers on the number lines are not known, and they have to find the size of one partition. Similarly, the item designers have included at least five concepts in Question 8 to pitch it at a high level. In Question 9, the Fibonacci sequence problem has been made challenging by requiring the students to find a preceding term rather than a succeeding term. Question 10 requires more than taking a half and a quarter of a quantity but requires the coordination between a quarter and a half of two quantities.

Measurement and Geometry

Question 2 assesses not only the reading of time on two different types of clocks but also the calculation of time intervals. Similarly, the tasks involving spatial visualisation are challenging as they involve relatively complex designs. For example, the nets given in Question 3 require advanced visual-mental folding. Such types of problems may not be conventionally encountered in school textbooks. Moreover, the measurement item (Question 6) is known to be a demanding situation from the mathematics education literature. Students are known to experience difficulties in reading scales that do not start from zero. In fact, in this item both the start and end points are not given. The area item (Question 11) requires the discernment of salient features and an understanding of the relationship between these features.

Year 5

With respect to the Year 5 items, seven were from the Number and Algebra strand while six items were from Measurement and Geometry, and one item was from Statistics and Probability (see Table 4).

Table 4. Year 5 Items Classified by Content Strand, ACARA Numeracy Elements and Item Description

Question	Strand	Numeracy continuum	Concepts assessed and structure of problem
1	Number and Algebra	Estimating and calculating with whole numbers	Two-step word problem involving addition and subtraction of age; requires working backward
2	Measurement and Geometry	Using spatial reasoning	Reflection/Flip
3	Number and Algebra	Recognising and using patterns and relationships	Pattern patterns
4	Number and Algebra	Estimating and calculating with whole numbers	Skip counting in steps of 50
5	Number and Algebra	Estimating and calculating with whole numbers	Multi-step word problem involving arithmetic operations
6	Measurement and Geometry	Using spatial reasoning	Constructing a solid from a net and counting the edges
7	Measurement and Geometry	Using spatial reasoning	Folding a net and adding numbers on opposite sides
8	Number and Algebra	Estimating and calculating with whole numbers	Word problem involving two additive comparisons
9	Number and Algebra	Estimating and calculating with whole numbers	Word problem involving the multiplicative comparison of two sets of quantities
10	Measurement and Geometry	Using measurement; Using spatial reasoning	Comparison of two volumes
11	Number and Algebra	Using fractions, decimals, percentages, ratios and rates	Proportion problem involving decimals
12	Statistics and Probability	Estimating and calculating with whole numbers	Word problem involving an unknown quantity
13	Measurement and Geometry	Estimating and calculating with whole numbers	Scaling problem
14	Measurement and Geometry	Using spatial reasoning	Counting number of edges on a 3D pyramid with three faces hidden

From Mayer's framework (see Table 5), all the tasks at the Year 5 level required conceptual knowledge and in terms of cognitive processes, they involve application and analysis.

Table 5 Classification of Year 5 Items Within Mayer's Matrix

Types of knowledge	Type of cognitive process					
	Remember	Understand	Apply	Analyse	Evaluate	Create
Factual						
Procedural						
Conceptual			3,4,11, 13,	1,2,5,6,7,8, 9,10,12,14		
Metacognitive						

The structure of the tasks

Number and Algebra

Question 1 requires students to coordinate two quantities at two different points in time. Further, it involves working backwards. These characteristics fix the item at a relatively higher level in comparison to a traditional Year 5 task. Questions 3 and 4 are similar to Questions 9 and 7 at the Year 3 level (respectively), and we have already commented on their order of difficulty earlier. Question 5 is a multi-step word problem where the inference that there are three small packets of crayons is critical in the solution of the problem. In Question 8, none of the quantities have a specific value but rather the additive relationships among the quantities are given. The aim is to find the value of the quantities. Thus, Question 8 is a higher-order item in that the problem solver has to work with differences rather than actual values. Question 9 is similar to Question 12 at the Year 3 level, and we have already commented on what makes it demanding. Question 11 brings to forth students' ability to reason proportionally. What makes this item demanding for Year 5s is the numerical feature of the quantities, that is, the use of decimals.

Measurement and Geometry

Question 2 requires students to visualise the relative positions of three related objects when flipped. In so doing it tests the sharpness with which students manipulate images. Question 6 involves the visualisation of a heptagonal-based prism, an object which may not be necessarily familiar to Year 5 students. It requires mentally folding the two-dimensional net to count the edges in a three-dimensional object. Question 7 involves not only folding the net of a cube but also adding the numbers on the opposite sides. These two aspects of the problem set it at a higher-order level. Question 10 involves the comparison of the volume of two boxes to determine how many of the small boxes fit in the larger box. It allows the acute assessment of students' reasoning about volumes in terms of the comparison of the appropriate dimensions from two given objects. It may be intuitive to compare the given dimensions as per the visual display of the two objects. Question 13 requires scaling a real object on a grid. This type of problem may not be conventionally encountered by Year 5 students in the normal mathematics curriculum. Question 14 is set at a higher level by hiding three faces of the heptagonal-based pyramid. Thus, students are required to visualise those hidden parts.

Statistics and Probability

Only one item, Question 12, could be classified under the Statistics and Probability strand. This item is demanding as it involves working with an unknown quantity.

Year 7

The distribution of the items by strand for Year 7 is as follows: nine items from Number and Algebra, five items from Measurement and Geometry and two items from Statistics and Probability (see Table 6).

Table 6. Year 7 Items Classified by Content Strand, ACARA Numeracy Elements and Item Description

Question	Strand	Numeracy continuum	Concepts assessed and structure of problem
1	Number and Algebra	Estimating and calculating with whole numbers	Find the number of factors in a number (50)
2	Number and Algebra	Estimating and calculating with whole numbers	Rounding to the nearest 10, 100, 1000 and 10 000
3	Measurement and Geometry	Using spatial reasoning	Finding cross section
4	Statistics and Probability	Using fractions, decimals, percentages, ratios and rates	Ratio; Alternatively, computing number of favourable outcomes given probability;
5	Measurement and Geometry	Using measurement; Using spatial reasoning	Identifying position on a grid map and measuring distance on a scale
6	Number and Algebra	Using fractions, decimals, percentages, ratios and rates	Comparing fractions
7	Number and Algebra	Recognising and using patterns and relationships	Number patterns
8	Measurement and Geometry	Using measurement, Using spatial reasoning	Perimeter involving unknown sides
9	Measurement and Geometry	Using spatial reasoning	Rotation/Turn
10	Statistics and Probability	Interpreting statistical information	Finding most likely event
11	Number and Algebra	Estimating and calculating with whole numbers	Word problem involving prime number and factors
12	Number and Algebra	Using fractions, decimals, percentages, ratios and rates	Rate, money, rounding to the nearest five cents
13	Measurement and Geometry	Using measurement	Finding length using geometrical properties
14	Number and Algebra	Using fractions, decimals, percentages, ratios and rates	Addition and division involving decimals
15	Number and Algebra	Using fractions, decimals, percentages, ratios and rates	Ratio; multiplicative comparison
16	Number and Algebra	Recognising and using patterns and relationships	Patterns from geometric representation

Mayer’s framework identified the majority of the Year 7 items as analysing conceptual knowledge (see Table 7). In contrast to the younger grade levels, a number of the Year 7 items were classified as applying procedural knowledge; however, in these instances, the procedural knowledge required was moderately complex.

Table 7. Classification of Year 7 Items Within Mayer’s Matrix

Types of knowledge	Type of cognitive process					
	Remember	Understand	Apply	Analyse	Evaluate	Create
Factual						
Procedural			1,3,9, 14			
Conceptual			2	4,5,6,7,8, 10,11,12, 13,15,16		
Metacognitive						

The structure of the tasks

Number and Algebra

Question 1 is a routine task. Question 2 is pitched at a higher level in that it tests whether students are able to round numbers to the nearest 10, 100, 1000 and 10 000 in one item. A conventional Year 7 question would have only asked the problem solver to round the number to only one level, say nearest 100 or 1000. Question 6 involves the comparison of four fractions with different denominators, three of which (3/4, 19/24, 5/8) can be readily put on the same denominator to identify 19/24 as the largest of the three. However, the comparison of 19/24 and 13/16 may be slightly more demanding.

In general, Year 7 students may not be working extensively with linear relationships and may not have much exposure to identify patterns of the type set in Question 7. Thus, those students who can reason inductively in such a problem would show a higher level of numerical competence. Although Question 11 involves two familiar concepts (prime numbers and factors introduced at the primary level), it creates sufficient challenge for the Year 7 students through contextualisation in a word problem. Question 12 tests students’ ability to work with rates and provides opportunities to observe if they can reason proportionally. Question 14 is a procedural item, verifying if students can compute addition and division of two-digit decimals by including a distractor (where one of the numbers in the denominator is the same as the numerator). This item tests procedural competency more than it elicits higher-order thinking. Multiplicative reasoning is known to be demanding for middle school reasoning. Question 15 is an ideal item to tests students’ understanding of multiplicative comparison or ratio.

Measurement and Geometry

Access to Question 3 may not necessarily require higher-order thinking in as much as it requires familiarity in working with cross sections. Question 5 requires two essential steps: (i) identifying the position of an object on a grid map, and (ii) measuring distance on a scale. The combination of these two aspects set Question 5 at a higher-order level. Question 8 provides opportunities for students to elicit their knowledge beyond their conventional experiences of perimeter. This item requires inference beyond the information provided in the diagram. Question 9 tests students' sense of spatial orientation. Question 13 is aimed at a higher level as it involves the calculation of length by using a design which requires knowledge of the properties of equilateral triangles. The last item (Question 16) elicits students' ability to construct a numerical pattern from a geometrical pattern and to reason inductively.

Statistics and Probability

Question 4 may either be interpreted from a ratio perspective or from a probability perspective. In terms of ratio, it involves interpreting 1 in 9 chance as 1:9. In terms of probability, it requires computing the number of favourable outcomes given the probability. Question 10 may be a novel situation for Year 7 as students may not have been exposed much to combined events. But this item allows them to extrapolate their knowledge of probability.

Year 9

The distribution of the items by strand for Year 9 is as follows: eight items from Number and Algebra, seven items from Measurement and Geometry and one item from Statistics and Probability (see Table 8).

Table 8. Year 9 Items Classified by Content Strand, ACARA Numeracy Elements and Item Description

Question	Strand	Numeracy continuum	Concepts assessed and structure of problem
1	Measurement and Geometry	Using spatial information	Counting number of edges, where some of them are hidden
2	Number and Algebra	Recognising and using patterns and relationships	Linear relationship
3	Number and Algebra	Using fractions, decimals, percentages, ratios and rates	Rate involving large numbers
4	Number and Algebra	Using fractions, decimals, percentages, ratios and rates	Speed in a two-stage journey
5	Measurement and Geometry	Using spatial information	Properties of quadrilaterals
6	Number and Algebra	Using fractions, decimals, percentages, ratios and rates	Multiplication of decimals
7	Statistics and Probability	Using fractions, decimals, percentages, ratios and rates	Ratio; Alternatively, computing number of favourable outcomes given probability;
8	Number and Algebra	Recognising and using patterns and relationships	Number patterns
9	Measurement and Geometry	Using measurement	Volume of cylinders
10	Number and Algebra	Recognising and using patterns and relationships	Finding unknown value of a quantity
11	Number and Algebra	Using fractions, decimals, percentages, ratios and rates	Division of decimals
12	Measurement and Geometry	Using measurement; Using spatial reasoning	Symmetry and calculation of distance on a grid
13	Number and Algebra	Using fractions, decimals, percentages, ratios and rates	Percentages
14	Measurement and Geometry	Using measurement	Finding angle in a complex figure
15	Measurement and Geometry	Using measurement; Using spatial reasoning	Volume of prism
16	Measurement and Geometry	Using measurement	Volume of cube

The majority of the Year 9 items were either applying or analysing conceptual knowledge (see Table 9).

Table 9. Classification of Year 9 Items Within Mayer's Matrix

Types of knowledge	**Type of cognitive process**					
	Remember	**Understand**	**Apply**	**Analyse**	**Evaluate**	**Create**
Factual						
Procedural		5	11			
Conceptual			6,10,13	1,2,3,4,7,8,9,12, 14,15,16		
Metacognitive						

The structure of the tasks

Number and Algebra

Questions 2 and 8 require students to determine the linear relationship between two sets of numbers. It provides opportunities to observe whether students can reason inductively. Question 2 allows the student to show his/her flexibility to relate numbers. Question 8 is similar to Question 7 at Year 7 as commented earlier. Question 3 is a multiplicative comparison problem involving large numbers and an indivisibility relationship between the two given numbers. Further, it requires rounding the result of the division to the nearest cent. A student who can answer this question certainly shows that s/he has a sound understanding of multiplicative situations. Rates, particularly speed, are known to be a hurdle for many students. By casting a problem situation involving speed across a two-stage journey, Question 4 provides a test case to verify the robustness of students' concept of speed. Questions 6 and 11 test students' ability to manipulate decimals. Although Question 6 appears to be a more procedural question, it tests students' sense of decimal multiplication by asking them to situate the answer in an interval rather than giving the exact value. Thus, it has a more conceptual characteristic and is a suitable item for Testlet F. Question 11 is a procedural task and tests students' proficiency with two-digit decimal division, which we know may be problematic for secondary school students. Question 10 involves a distractor in the presentation of the algebraic task in that students may be prompted to cancel the unknown term from either side of the equation. Thus, it creates a situation where the problem solver's ability to work with unknowns may be tested. However, given that it uses a small numerical value, students may be tempted to use guess-and-check. Question 13 is an application of percentage and is not necessarily a higher-order task for Year 9 students.

Measurement and Geometry

The spatial tasks include Questions 1, 12, 15 and 16. Question 1 involves counting the edges of a prism, where some of the edges are hidden. It requires students to visualise the hidden edges. Question 12 involves two main aspects, reading a grid scale and locating a point so that a dot configuration has a line of symmetry. Locating the line of symmetry may not be obvious to students. Question 15 requires students to find the area of the base of a prism on a grid to compute the volume of the prism. Although this problem may not involve higher-order reasoning, it does require students to work with an irregular polygon. The last item, Question 16, is a higher-order item as it involves the realisation that the base should be a cube and how such a smallest cube may be constructed. It may also entail elements of visualisation.

Question 5 is a particularly elementary problem for a Year 9 student and it does not require higher-order thinking. Question 9 involves the comparison of the volumes of two cylinders. It requires the application of the formula for finding the volume of a cylinder and as such it may not be a high-order problem. Rather than asking students to find a missing angle in one isolated isosceles triangle, Question 14 embeds 10 such triangles together in a visually complex configuration. In so doing, it sets the order of the problem to the next level.

Statistics and Probability

Question 7 at Year 9 is the same as Question 4 at Year 7.

Summary

Across all the year levels, the majority of items fell under the Number and Algebra and Measurement and Geometry strands of the Australian Curriculum. The number of Statistics and Probability items was sparse, which was surprising to us given the strand's importance in the curriculum. Furthermore, we would anticipate that the content associated with this strand would be precisely what highly capable students should be engaging with. There was a relatively even mix of the ACARA numeracy continuum elements present across all the year levels. Due to the lack of Statistics and Probability items, the numeracy element *Interpreting statistical information* was underrepresented.

Mayer's (2002) framework provided a matrix to understand how the items were designed to elicit higher-order knowledge and cognitive processes. The majority of the items across all year levels related to applying or analysing (process) conceptual knowledge. The processes and knowledge fell toward the middle of the higher-order thinking continuum. Therefore, these items could be deemed to be at a high level of processing for the respective year levels and provided opportunities for students to utilise various higher-order thinking skills.

Task Representation and Design Features

In this section, we comment on the design of the Testlet F items with respect to the presentation of the task in terms of its formulation, graphics used and accessibility in a digital environment. As in the previous sections, we present the data on a year-by-year basis.

Year 3

While the majority of the items have been well designed, some of them drew our attention as we interviewed the Year 3 students. It can be observed that 9 out of the 12 items involved a graphic in Testlet F. It is important to be mindful about the possible inadvertent effects that graphics may induce in an item. For instance, the analogue clock presented in Question 2 did not include numbers but only marks. One of the students read the time in the analogue clock as three o'clock (and gave the answer as 45 minutes by subtracting the incorrectly read time from 3:45) possibly because the numbers were not included. It is suggested that numbers be included in items involving analogue clocks, especially for Year 3 students.

The response from a student to Question 5 drew our attention to the colours used in this graphic item. The student described the orange dot as a green dot. It is suggested that shading (or other means of differentiation) rather than colours be used as this may inadvertently affect the student's response. The graphic in Question 6 also raised some concern. It was quite difficult for the students to read the small subdivisions on the computer screen. Some of the students used their pencil to count the small subdivisions on the screen. They had to move close to the screen to be able to focus on the subdivisions. It is suggested that for such types of measurement items, the subdivisions be made relatively conspicuous so that repeated counting does not consume examination time. On a paper-and-pencil test, it may be easier to use the pencil and count the small subdivisions..

Additionally, we could observe that the students did not pay much attention to information written above the table in Question 4. There is a remarkable difference in the font size of the numbers presented in the table and the wordings surrounding the table. We wonder whether this difference in presentation may have contributed to the students ignoring the wordings; although in this item, the information given above the table is not overly influential to the solution of the problem. A similar observation can be made in Question 9, where the font of the numbers was much bigger than the instructions given above the Fibonacci sequence. Several students focused on the numbers without paying attention to what was written above them, which in fact, contained the rule to solve the problem.

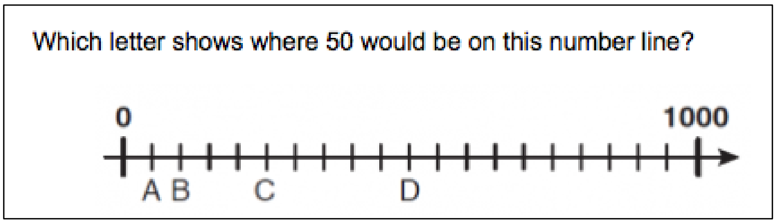
Another item that drew our attention is Question 8. Some students missed this item because one of the intermediate steps was incorrect. Regarding the design of this item, we would suggest that even for this category of high-achieving students, it might be fair, from an assessment perspective, to have a fewer number of steps.

Year 5

The interviews with the Year 5 students elicited some of the ambiguities that the design of the items inadvertently created. For example, in Question 2, many students hesitated whether the answer should have been A (involving flipping) or D (not involving flipping). Some of the students were not clear whether a flip was to be done and consequently chose D as the answer. We feel that the indication in the problem is not precise enough and may involve the two different interpretations that the students made. Some students even asked whether it was the right or left shoe. Further, a rectangle is presented rather than a shoe. Some students did not even realise that a flipping action was involved: "I don't really know (if my answer is correct). I just tried to match it up because this is what I thought it meant." Further, the picture in the question is bigger than those given in the options. "I tried to match the **bigger** picture (given in the question). I found D (option) the closest."

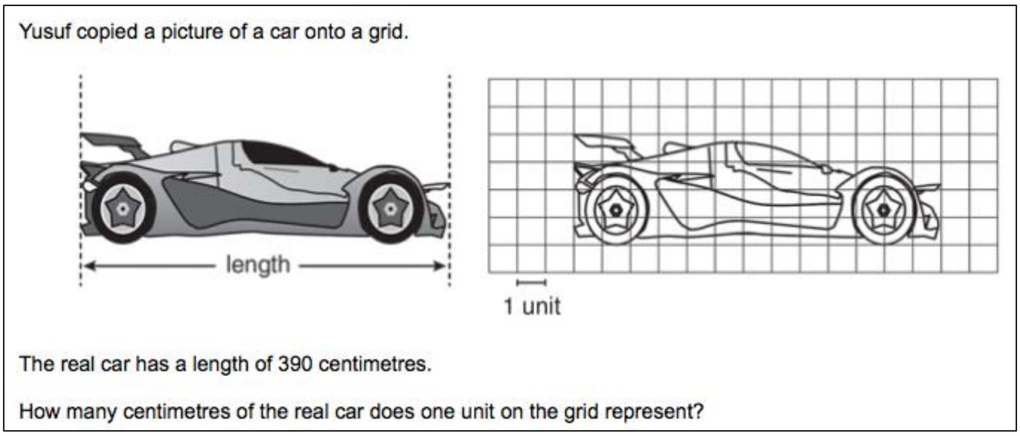
Question 3 at Year 5 is similar to Question 9 at Year 3. As we mentioned earlier, the numbers in the pattern were presented in much bigger font than the instructions describing the pattern, possibly making it appear that this was the most important information in the problem. Consequently, some of the students just focused on the numbers and spent much time figuring out what should be the first missing number in the pattern. One of the students mentioned that the question mark (embedded in the blue layout) in the first term of the pattern looked like the numeral 2. It is suggested that an empty bar (or some other place holder) be included, rather than question mark.

In Question 4, a few students had difficulties counting the small subdivisions on the number line on the screen: "The number line was small. I got mixed up." It appears that in a digital environment, it requires much more effort to count along subdivisions and grids. We could observe students counting and recounting the subdivisions in Questions 4 and 13. This may be more so in a stressful examination environment.



Question 4 from the Year 5 items

In Question 13, a few students experienced some challenge in keeping track of the subdivisions in 'counting' the length of the car on the grid on the computer screen. Some got 14 units rather than 13 units. Students had to come close to the screen to count. As one student pointed out: "Except the grids, we couldn't see the car…the car made the grid like. People with glasses really can't see it properly… it was blurry." It is suggested that larger grids be used.



Question 13 from the Year 5 items

In Question 10, the orientation of the two boxes side-by-side appear to have prompted the students to divide corresponding sides, without considering that two of the sides were equal. It appears that the diagram may have been a distractor here, tunnelling students' attention on corresponding lengths. It remains to be verified how such students would solve this problem with a diagram in a different orientation or even without a diagram.

In Question 11, the notation for grams (g) does not have a consistent font; one is larger than the other. This did not raise any issue but it may be good practice to maintain consistency of fonts. A few students used the answer that they obtained in the first part to answer the question in the second part intuitively. For instance, one student looked at the 'resemblance' of the numbers to answer the question, i.e., since the amount of sodium given in the first column involved hundredth (0.03 g), the number above it should be 0.05 g (in the hundredth as well). Similarly, she wrote the mass of sodium as 3 g based on the integer value '5' g in the second column. Another student looked at the pattern in the answer in the first row in attempting to find the missing value in the second column: "I just worked that out looking at the top one because they are similar."

Our working experiences with primary students suggest that they tend to make a number of associations in a problem intuitively often in a flawed fashion. Thus, in the design of items such as Question 11, it may be better to present such information differently rather than a table where the structure of the numbers can be influential. Furthermore, in the same question, one of the students attempted to enter 1/2 rather than 0.5 in the table and the program did not accept it. He then tried to type 'half'. The student then asked the interviewer how he could input 1/2. In other words, the program accepts only decimals. This observation was also made in Question 13. It is recommended that such types of information be shared with the students in a tutorial before taking the online examination.

Year 7
As in Years 3 and 5, the items in Year 7 were well designed from a presentation point of view. Relatively fewer issues were noted. In Question 8, for the sake of consistency, it is suggested that the size of the symbol for metre (m) be uniformly maintained. Further, it is suggested that the fonts be made bigger.

Year 9
Similarly, in Year 9, the design of the items did not cause any direct inadvertent issues for the students. In Question 4, it is suggested to avoid using similar numbers for two different quantities. The speed is 6 km/h and the distance for the second part of the journey is 6 km. Apparently, this led to some confusion. In Question 8, none of the students used the triangles above the table. In fact, they become so small as the number of triangles increase, it is almost impossible to use them appropriately. As one of the students mentioned: "The diagram makes it more difficult." We wonder whether it is important to put all the triangles in this item, or only the first few. In Question 12, five dots are given in the problem, only three of which are labelled. It is not clear why the other two dots were not labelled. In fact, the absence of labels for the two points did bring some confusion for a few students. In Question 15, it was challenging for some of the students to visualise splitting the given area in the question and to mentally keep track of the subdivision on the screen. On pencil-and-paper, the area can be subdivided and marked for ease of counting.

Literacy Demands

Year 3
For Year 3 students of this mathematics ability, language was not particularly an issue. The students could easily read the tasks. For some of the items, although the students could read the problem, often it was the variation in the question (which required higher-order mathematical thinking) that led them to provide an incorrect answer. At times, it was the challenge of getting the solution procedure in the problem that made the students indicate that the wording was not clear. For example, one of the students made the following comments in Question 10: "I understand 'a quarter of 120' but I don't get it that (when) it says 'is the same as a half of'." At other times, it was the coordination of the language (problem statement) and the graphic associated to the problem that created some confusion. For example, in Question 6, one of the students mentioned he was not

confident about the correctness of his answer because “I am not sure if we just add how many, how much taller (than) a metre he is or like the whole thing.” Given that a complete height chart was not given, he experienced some ambiguity in terms of what the question was really looking for. During the interview, when a student missed the problem, we asked him/her to read the problem aloud to determine if linguistic elements were a hindrance. All the students could read the given problems with understanding.

Year 5

As for the Year 5s, the students could easily read the items but it was primarily when they did not know what method to use that they indicated that the language was not clear: For example in Question 13, one student mentioned that the following sentence is not clear: “How many (centimetres) of the real car does one unit on the grid represent?”

Year 7

Similarly, the Year 7 students could read the items fluently and made sense of the problem situation. The few comments that we find worth reporting is in relation to Questions 2 and 5. In Question 2, the wording of the problem, specifically, “Which two ways of rounding give the same answer?” was not clear for two students. One student mentioned that the problem is quite wordy: “Lots to read and I have to read each one.” In Question 5, two students mentioned that they did not understand what “due North” and “due West” meant. We wonder if ‘due’ is a common mathematical terminology at Year 7 and whether it is necessary to use this word in the item.

Year 9

As was the case for the previous three year levels, there was no apparent linguistic issue on the formulation of the problems. Sometimes, it is because of inadequate mathematical knowledge or when the students had forgotten about particular terms that they mentioned that the question was not clear. For example, in Question 3, three students had difficulty understanding what the problem was asking. One student mentioned that the terms “another country” is confusing. Maybe the name of a country could have been included instead. We were surprised to observe that 7 out of the 10 students answered Question 5 incorrectly. The students were unsure about the meaning of ‘diagonals that cross at right angles’. A few just focused on the right angles and chose option A as the answer. Some just guessed the answer. The following are some examples of their comments:

Students’ verbatim comments

“We don’t do this kind of geometry stuff.”

“Don’t understand what the question is asking for.”

“Does not know what the term ‘diagonal’ mean exactly.”

“We haven’t learned about diagonals crossing at right angles, never.”

Dealing with subtleties in a word problem

One of the characteristics of higher-order problems is the manner in which the task is posed. Students who can observe the subtleties in the wording of such problems are more capable of discerning the inherent quantities and their relationships and, consequently, have more chances of successfully solving the problems. For example, Question 1 involves a situation where the subtleties in the wordings posed some challenge for a number of Year 5 students. The statement of Question 1 reads as follows: "In 4 years time Jodie will be 20 years old. Steve will then be half Jodie's age. How old is Steve now?" During the interview, we could observe how they re-read the problem several times. As one of the students mentioned: "The words made it a bit trickier." This item also requires students to work backward. The most common incorrect answer was 8, obtained by subtracting 4 from 20 and dividing 16 by 2. Such subtleties in making sense of a word problem could also be observed in Question 9.

Issues with the calculator

- The calculator was not easily accessible for a number of students during problem solving. After prompting by the interviewer suggesting that a calculator is available, some students started to use it.
- When the calculator is opened, it covers the problem statement. The students did not know that the calculator could be dragged.
- The calculator lacks flexibility, for instance, intermediate numbers cannot be deleted. For example, if we want to delete the number 7 and replace it by 6 in $\pi \times 7^2 \times 25$, we have to delete everything and retype the numbers. This was observed in Question 9 in Year 9.
- The cube root of 64 was given as 3.999999… rather than the exact number 4.
- A few students mentioned that the division symbol looks like the addition symbol on the calculator.
- One student could not find the multiplication key on the calculator given that it is represented by '*' rather than '×'.

Priority 2: Examine the performance of such students on Testlet F of the tailored test.

Research question 2.1: How do highly capable students perform on the Testlet F items?

Research question 2.2: What performance characteristics are common among these students?

Student Performance on Testlet F

The cohort of students who participated in this study could be considered both capable and high performing. In comparison to those students who completed Testlet F in the main study, this project's participants had higher mean scores across all year levels (with three of the four at statistically significant levels, see Table 10). Hence, the findings of this report can certainly reflect the profile and characteristics of the students who are likely to encounter such Testlet F items in the future.

Table 10. Performance Comparisons Between the Main Study Cohort and the Cognitive Interview Cohort

Level	Mean (Main Study)	Mean (Cognitive Interviews)	*t*-test result
Year 3	0.37	0.62	$t(22) = -4.076, p \leq 0.001$
Year 5	0.39	0.54	$t(22) = -1.878, p = 0.072$
Year 7	0.41	0.66	$t(30) = -3.815, p \leq 0.001$
Year 9	0.38	0.63	$t(30) = -3.722, p \leq 0.001$

Comparative Performance Across Common Items

A number of items were presented across two year levels—both Year 3 and 5 and again in Year 7 and 9. Presented below are descriptions of the performance between the students at different year levels on these items.

Comparison of Year 3 and Year 5 on the three common items
Three items were identified as common across the Year 3 and 5 Testlets; Number-line, Fibonacci pattern, and Spoon-and-Fork. Unsurprisingly, more Year 5 students answered these items correctly than Year 3 students (see Table 11).

Table 11. Comparison of Similar Items Between Year 3 and Year 5

Item	Percentage correct	
	Year 3	**Year 5**
Number-line	69%	73%
Fibonacci pattern	31%	80%
Spoon-and-Fork	43%	67%

The Number-line item

At both year levels, computational and systematic guess-and-check strategies could be observed. The Year 5 students were more competent and agile in skip counting along the number line. At both year levels, the students misread the final number as 100 rather than 1000. In general, the Year 3 students rated this item as hard, while the Year 5 students rated it in the continuum easy to Ok.

Some of the Year 3 students considered this item as being hard partly because they had not done a similar problem before. At school, the scale tends to be limited to 100 and the number lines do have many such subdivisions. They may be more used to finding the length along a number line, rather than finding the size of a subdivision given the length, especially when the skip count involved jumps of 50. Further, they may not be acquainted with this format of a measurement problem where letters are used to designate particular positions of numbers. Some examples of their comments are below:

Students' verbatim comments

"In school, up to 100, not 1000. We have just numbers and we put them in the number line, not letters."

"More lines in the number line."

"Normally count by 10's or 1's in school."

Year 5 students were confident about their answers, although some of them were incorrect. The problem was seen as being different from school questions. For example, at school, they are used to counting by 1, 5, and 10, whereas here they have to count by 50. The quotes below highlight other differences:

Students' verbatim comments

"You normally have numbers in the line and complete blank and this is all blank."

"We don't really make the range so big. We don't use letters but numbers."

"We don't do really this kind of questions at school."

"Less strokes in class, about 10 to 12."

The Fibonacci pattern item

Clearly, the Year 5 students were more fluent in solving the Fibonacci pattern item. In fact, a majority of the Year 3 students answered this item incorrectly. It appears that the Year 3 students did not spend much time reading how the pattern was created but rather started to work with the numbers directly. Some of the Year 3 students gave the answer 8 since it is half of 16 (the next term in the pattern), on the basis of a more intuitive rule. One student used a systematic guess-and-check. He started with 10 to observe that 10 + 16 = 26. To adjust the sum to 23, he changed his initial guess to 10 – 3 = 7.

On the other hand, some of the Year 5 students knew that the sequence of numbers given in the question was called a Fibonacci sequence and were familiar with this pattern. One student even referred to the first unknown term by x: "Make the question mark (given in the problem) x; so $x + 16 = 23$." A few students identified the pattern by reading the item stem and could deduce that they had to subtract 16 from 23. However, a few students did use systematic guess-and-check. Thus, it is mainly the higher level of mathematical maturity of the Year 5 students that explain why they do better than the Year 3s.

The Spoon-and-Fork item

Question 12 was the second most demanding item for the Year 3s. Those who obtained the correct answer first added 15 + 15 + 15 or multiplied 15 by 3. However, not all of them could immediately deduce that they had to do division to find the number of spoons. Thus, they attempted to add particular numbers (three times) such as 7, 7.5, 8 and eventually 9 to deduce the length of the spoon. Those who could not relate this problem to a multiplication (or repeated addition) and corresponding division situation, visually compared a fork and a spoon. The length of the spoon was estimated to be more than half of the fork (i.e., greater than 7.5) and 8 was given as an answer.

On the other hand, some of the Year 5 students could readily see this problem situation as involving multiplication and division ($15 \times 3 = 45$; $45 \div 5 = 9$). Although some students added (or multiplied) the lengths of the spoons to get 45 cm, what was missing was the realisation that they had to divide 45 by 5.

Like the Year 3 students, some of the Year 5s did use the more intuitive visual comparison strategy. For example, a few students observed that the spoon was about half of the fork and gave the answer 1/2 of 15. Others went one step further and argued that it is a bit more than 7.5 and gave answers such as 8 cm and 9 cm, without any numerical justification. Another student gave the answer 13, justifying her answer: "The spoon is a little bit off the fork."

Comparison of Year 7 and Year 9 on the two common items

Two items were identified as common across the Year 7 and 9 Testlets: the Muffin problem and the Pattern problem. Unsurprisingly, more Year 9 students answered these items correctly than Year 7 students (see Table 12).

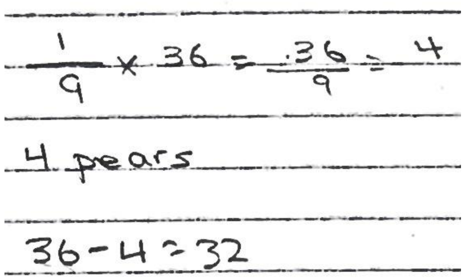
Table 12. Comparison of Similar Items Between Year 7 and Year 9 Students

Item	Percentage correct	
	Year 7	**Year 9**
Muffin	70%	90%
Pattern	40%	50%

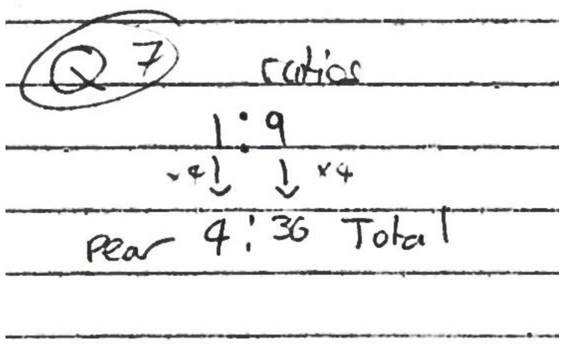
Muffin item

The strategies used by the Year 7 and Year 9 students were quite similar. Both years essentially interpreted the problem from a fraction perspective. The item was accessible to both year levels. This item was generally rated as easy by both groups of students.

Sample solution from Year 7 student: 1/9 × 36 = 36/9 = 4; so there are 4 pear muffins. Thus, there are 36 − 4 = 32 apple muffins.



Sample solution from Year 9 student: 1 in 9 chance is the same as 4 in 36 chance (1/9 = 4/36); 36 – 4 = 32. However, a few of the Year 9 students interpreted the problem as a ratio as follows—Pear:Apple = 1:9 = 4:36; 36 ÷ 9 = 4; 36 − 4 = 32.



Pattern item

Note that this item was among the hardest for the Year 7 students. Those who could solve it either deduced the rule between the numbers in the two rows of the table (i.e., divide by 2 and subtract 1) or extended each of the corresponding terms until they reached 6561 (the required cell) by adding the numbers in the successive columns. A few students made computational mistakes while one student mentioned that she cannot see the pattern and did not know how to start the problem. This question was seen as different from the patterns that they encounter at school.

With regard to the Year 9 students, approximately 50% of the students answered this item correctly. Like the Year 7 students, different strategies could be observed. For example, after comparing the corresponding numbers in the first and second rows, one student observed that if he doubled the number of white triangles (second row) and subtracted 1, this would equal the number of black triangles (first row). Then, using the 'opposite' sequence of operations, he computed $(6561 - 1) \div 2 = 3280$ on the calculator. Another student observed that if he divided the number of black triangles by 2 and 'round down' the results, this will yield the number of white triangles. One student used the rule: $(x \div 2) - 0.5 = y$. A few students had studied linear relationships and could easily identify the pattern. However, some of the students could not find the pattern and struggled with this item. They were not conversant with particular strategies that would allow them to reason inductively. Half of the students rated it as being a hard item while the other half considered it easy.

Characteristics of the High-Performing Students

Profile of Year 3 students

Participating schools were requested to provide students who were mathematically capable (top 20–30% in their class cohort). During the interviews, we collected additional information from the students about their mathematical achievements. Most of the Year 3 students mentioned that they got A's in mathematics or did well in NAPLAN, and have been doing well in the subject in previous terms. For example, one student mentioned: "A in Math in every single year." Among others, they like performing arithmetic operations, especially multiplication and solving challenging problems. Some of them showed remarkable proficiency with simple fractions and decimals. For example, in his response, one of the students mentioned that 0.5 multiplied by an odd number will yield a number that ends in 0.5. Considering their age level, some of the students had remarkable numerical flexibility. In one of the schools, the students were from an extended mathematics class, where they studied higher mathematics than their grade level. We could also observe that they were quite perseverant in nature, and did not give up when confronted with a challenging problem. In the other schools, although the students were from the top 20% of their class, their responses in the cognitive interviews showed that they were above average achievers but not necessarily high performers. Most of them work out mathematical problems on *Mathletics* at home. Some even work with their parents on particular mathematics books at home.

Profile of Year 5 students

The sample of Year 5 students consisted of 10 boys and 5 girls. The majority of students mentioned that they have been doing well in mathematics in previous years. Eight of the participants explicitly mentioned that they got Band 7 or Band 8 in their previous NAPLAN examinations. The students from one of the participating schools were from an extended mathematics group and they have been studying some higher mathematics than their grade level, such as Fibonacci sequences. During the one-to-one interviews, we could identify six students struggling with the problems; in fact, this is reflected in the final score in the Testlet. Some of the students had remarkably high profiles in Mathematics. For example, one of them was among the top 18% in the NSW region in the ICAS examination. Another one participated in the mathematics Olympiad and did very well. He was in the highest group of the extension classes in mathematics in Years 3 and 4. Most of them expressed that they like mathematics. For instance, one of the highest performing students mentioned: “I love maths. Maths is my best subject. I love problem solving and difficult problems.” Most of the students like performing arithmetic operations. Some of them work with their parents at home and solve problems on *Mathletics*.

Profile of Year 7 students

The sample of Year 7 students consisted of 5 boys and 5 girls. Most of the students scored A’s in mathematics and a few of them were in the top band in the last NAPLAN examinations. One of them took the ICAS examinations. Three of them mentioned that they work out mathematical problems with their parents at home. All of them like mathematics.

Profile of Year 9 students

The sample of Year 9 students consisted of 7 boys and 3 girls from two schools. Most of the students scored A’s in mathematics and a few of them were in the top band (8, 9 and 10) in the last NAPLAN examinations. One student mentioned that she attends extension classes where she studies higher-level mathematics. In one of the two schools, mathematics homework is sent to students through an online medium.

Priority 3: Monitor the students' knowledge, thinking skills and strategy use, and how these relate to the intended assessment outcomes envisaged by the test developers.

Problem-Solving and Higher-Order Thinking Skills

Research question 3.1: What types of problem-solving knowledge, skills and strategies do students utilise when solving challenging items?

In this section, we report how the students attempted to solve the tasks in terms of their problem-solving knowledge, skills and strategies. The majority of the items in Testlet F required students to analyse, synthesise, evaluate, or manipulate information. They go beyond the routine or mechanical application of previously learned materials such as stating mathematical facts or plugging numbers into a formula. Thus, they challenge students to reason. In this section, we highlight how the items elicit high-order knowledge and skills in terms of the following dimensions, consonant with our analytical framework:

1. Prompting non-algorithmic approach
2. Prompting the coordination of multiple pieces of information
3. Providing opportunities for deploying multiple strategies
4. Prompting higher level of mathematical reasoning: (a) Inductive/deductive reasoning; (b) Proportional reasoning; (c) Spatial reasoning; and (d) Algebraic reasoning

We also discuss knowledge elements that appeared to hinder access to the problems.

Prompting non-algorithmic approach

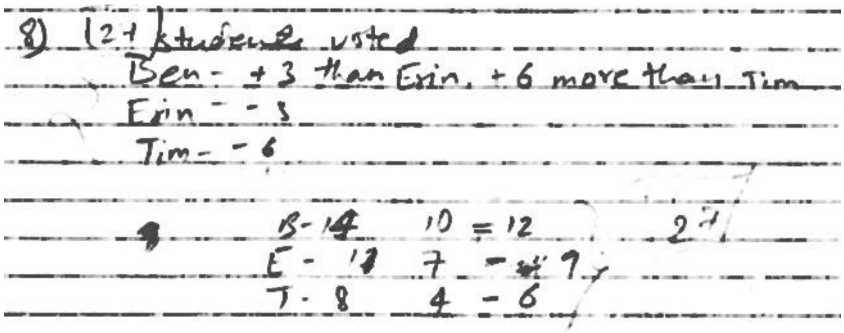
The higher-order nature of the problems prompted the students to devise creative ways to solve situations that they may not have encountered before. Thus, the items led them to devise problem-solving strategies that go beyond school-learned algorithms and procedures. In some of the items, we could observe how the students deployed quite insightful and sophisticated forms of reasoning.

Example 1: Exploiting the conceptual relationship between numbers – Year 3

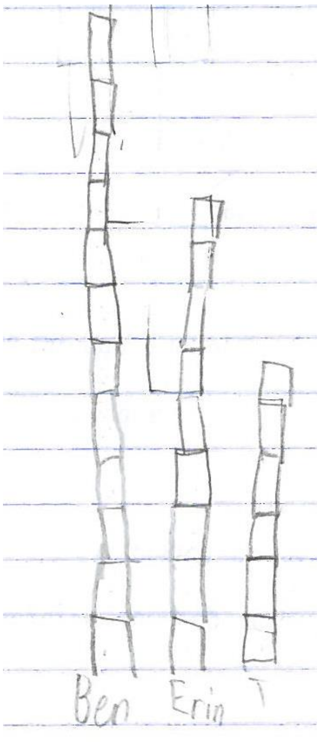
To solve Question 10, two students used the fact that 1/4 is 1/2 of 1/2 to correspondingly deduce that the answer can be obtained (directly) by taking 1/2 of 120. We consider this approach to be particularly sophisticated and non-routine as they used the knowledge that 1/4 of a quantity is 1/2 of 1/2 of that quantity. Students who used a routine approach first computed 1/4 of 120 to get 30 and doubled it to get 60.

Example 2: Making inferences beyond what is explicitly presented – Year 5

Some of the items in Testlet F prompted the problem solver to apply his/her mathematical knowledge in novel situations to make inferences beyond what was explicitly presented. Thus, the students had to improvise ways of approaching the problem beyond the conventionally taught procedures at school. Two items, namely Questions 8 and Question 12, were particularly constraining for the Year 5 students. With the unavailability of algebraic knowledge or tools, most of the Year 5 students referred to a systematic guess-and-check procedure to find the answer in Question 8. They did not have a deterministic procedure to solve this problem. After considering the differences between Ben, Erin and Tim, several students plugged in systematic guesses until they obtained the required answer 12 as shown by the diagram below. The student made the following successive adjustment until the total for Ben, Eric and Tim was 27, starting with 14 and adjusting his answer systematically to 10 and 12.



Another student looked at the differences between the three children and attempted to adjust them by using the rectangles shown below.



Similarly, Question 12 was challenging for a number of students. Only two of them used a truly deterministic procedure as follows: 54 – 12 = 42; 28 + 23 = 51; 51 – 42 = 9 or 28 + 23 + 12 = 63, 63 − 54 = 9. Several students used the guess-and-check method. Some students could not figure out how to solve the problem and, as a result, mentioned that it is not clear: "I can't figure out because it's not clear." Apparently, it was the literacy demand in the problem that prompted them to mention that the question was not clear.

The above two problems created some form of uncertainty for the students. In fact, one of the characteristics of high-order thinking (Resnick, 1987) is that it may involve uncertainty in that not everything that bears on the task at hand is known. It appears that such uncertainty was due to the unavailability of explicit information in the task rather than the foundational knowledge of the students. In other words, it is the nature of the tasks that creates uncertainty. As a consequence, students are prompted to reason and devise ways of approaching problems differently. We present two more examples where we could observe such productive scepticism in the students' mathematical behaviour as they started to solve the problems.

Example 3: Year 7

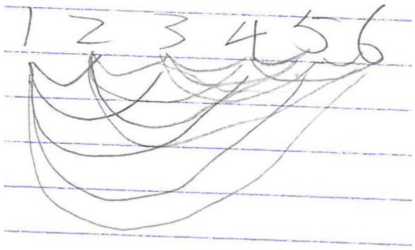
In Question 8, the majority of students estimated the lengths of the three sides where measurements were not given. They used the measures of the known sides as benchmarks in their estimation. They tried to allocate values to the three bottom unknown horizontal lengths to sum them to 9. One student said that the problem does not include enough information to be able to solve it. Only two students realised the equivalence of the given length (9 cm) and the three unknown lengths. These students interpreted the shape as a rectangle: "Move the line down to make a rectangle." Thus, they calculated the perimeter as (6 + 9) × 2 + (2 × 4) = 38 m.

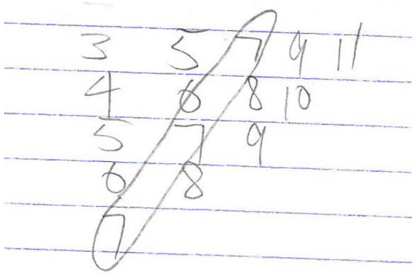
Example 4: Year 9

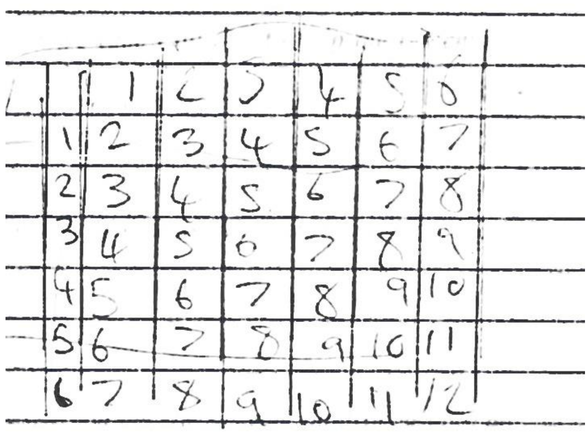
In Question 3, a few students were uncertain whether they had to divide 10 by 23 or 23 by 10. Thus, they were prompted to find a way to verify which of the two divisions should be used. For example, one student deduced that he had to divide 10 by 23 by rounding 23 million to 20 million. He then deduced that if there were 20 million people, then the amount would have been 50 cents (10 ÷ 20) per person; but since there are 23 million, the amount would be slightly less. Thus, the rounding procedure helped the student to surmount the challenge imposed by the indivisibility of 10 by 23 and prompt him to interpret the problem situation. Those who got it wrong divided 23 by 10. Sometimes they made computational mistakes as this question involves large numbers.

Example 5: Drawing a diagram or making a table – Year 7

In Question 10, some of the students made a diagram or table to be able to count all the possibilities when the spinner is turned twice (see diagrams on the following page). In the first diagram, the student listed all the possibilities when the spinner is turned two times and picked 7 as the maximum. Another student made a tabular representation (two-way table).







	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

The students mentioned that this problem situation was new to them:

Students' verbatim comments

"We don't do this type of questions at school."

"We don't do a lot of chance."

"At school, we do possibilities for spinning wheels but not spinning twice and adding it up."

"At school they give me more information like 'numbers could repeat' and questions would be 'most likely to land on'."

"Because it wasn't just chance. It was two steps—first adding totals and after finding all possibilities."

"School problems have easier language to understand."

Prompting coordination of multiple pieces of information

Higher-order thinking is elicited when students have to process multiple pieces of information to set relationship(s) among problem parameters. In the process, it also requires students to make the fine distinctions in the problem statement. Research suggests that primary-aged students struggle when there are multiple steps or multiple pieces of information to consider in any one task. Hence, the ability to coordinate such information could be considered higher order.

Example 1: Year 3

Students who successfully solved Question 1 showed diligence in differentiating between the three quantities in the problem, namely rabbits, carrots and days. They demonstrated that they could coordinate the different quantities. Five qualitatively distinct strategies could be observed (see Table 13).

Table 13. Year 3 Students' Strategies for Solving Question 1

Distinct strategies for Year 3 students solving Question 1	
Strategy A	Figure out number of carrots per day (2 × 3) then multiply by the number of days (6 × 5 = 30)
Strategy B	Figure out number of carrots per day (2 × 3 = 6) then repeated counting, i.e., adding 6 + 6 = 12 (for 2 days); 12 + 12 = 24 (for 4 days); 24 + 6 = 30 (for 5 days)
Strategy C	3 × 5 = 15 and double it (since there are 2 rabbits)
Strategy D	5 × 3 = 15; 15 × 2 = 30
Strategy E	Counting by 6's to get 30

Those students who were incorrect on this item just multiplied the 3 (carrots) × 5 (days) for only one day and gave the answer 15. In other words, they coordinated only 2 rather than 3 quantities in the problem.

Example 2: Year 3

Item 8 involves five successive steps, where the solution from one step is fed forward in the next step. Although each of the steps involves relatively elementary knowledge of number facts and arithmetic operations, what makes this problem challenging for the Year 3 students is to carry forward the intermediate results. Some of the students missed one step and consequently got the final answer wrong. A highly capable student is expected to make such successive coordination.

Example 3: Year 5

Question 5 involves multiple pieces of information. The common strategy to solve the problem was to find the number of crayons in the large and small packets and to add the amount as follows: 3 × 6 = 18; 9 – 3 = 6; 6 × 10 = 60; 18 + 60 = 78. Those who missed the problem in the process of solving it either did not read it carefully or attempted to do it mentally. When we asked those students who answered incorrectly to explain their answers, they realised that they misread or miscalculated. The one student that could not solve it even with further questioning attempted the problem mentally and as a result did not coordinate the multiple pieces of information.

Example 4: Year 7

Question 11 requires students to apply their knowledge of prime numbers and factors in a problem situation rather than the mere listing of factors as in Question 1. Thus, they have to use two concepts together. The common strategy to solve this problem was to find the factors of 57 (1, 3, 19, 57) and to identify the prime numbers (3 and 19); then to add 2 (3 + 2 = 5) to get another prime number (5). A few of the students thought that 57 was a prime number and gave it as the answer.

Example 5: Year 9

Question 4 is a multi-step problem requiring successive inferences. It involves the coordination of speed, distance and time in a two-stage journey. Although the item was answered well, a number of students obtained the answer using estimation rather than the required method. The few students who knew the method to solve this problem proceeded as follows. First, they worked out the time taken for the first 10 km = 10/6 = 1.7 hrs. Then, they computed the speed for the remaining 6 km (16 − 10) as 6 ÷ (4 − 1.7) = 6 ÷ 2.3 = 2.6 km/hr. Some students could perform the calculation for the first part of the journey—i.e., Indra took 10 ÷ 6 = 1 2/3 hrs or 1 hr 40 mins—but were not sure how to proceed with the second part or made computational errors. A few students estimated the answer, although some of their intermediate steps were incorrect. One student used the answer options to work out the problem.

Providing opportunities to deploy multiple strategies

High-order problems can often be solved in multiple ways. Thus, they create opportunities for students to harness different strategies to navigate situations that do not have a predefined trajectory.

Example 1: Year 3

In Question 7, three distinct strategies could be observed in the solutions of the Year 3 students. In the first and most sophisticated strategy, efficiently used by only 2 students, the respondents divided 1000 by 20 to get 50. In the second strategy, 1000 is halved to 500 in the middle of the number line; then 500 is divided by 10 to get 50; or 500 is subdivided by 2 to get 250 at *C*. The third and most common strategy was to start with an estimation such as 10, 20, 30, 40 and 50 and count along the number line. Some of the students initially (incorrectly) took the end number as 100 instead of 1000.

Example 2: Year 5

In Question 3, three distinct strategies could be observed.

> Strategy A: Students identified pattern from given instruction and subtracted 16 from 23.
> Strategy B: Students reasoned by denoting the unknown as x; $x + 16 = 23$.
> Strategy C: Students used systematic guess-and-check.

Example 3: Year 7

In Question 12, two strategies were apparent from the students' work. In the first strategy, they found the cost of 1 kg of banana from each stall. Then they found the difference in price per kg and multiplied by 15 for 15 kg. Finally, they rounded their answer. In the second more efficient strategy, they multiplied \$8.48 × 5 and \$13.99 × 3, and found the difference.

A few students did not use the calculator and that resulted in computational mistakes in adding or multiplying the decimals, although they knew the procedure to solve the problem. One student made prior approximations (rounding) and ended up with 40 cents rather than 45 cents.

Example 4: Year 9

In Question 13, the common strategy was to work out the number of seeds that sprouted in each pot as follows: 75% of 12 = 9; 50% of 20 = 10; 50% of 18 = 9. Thus, the total number of seeds that sprouted is 9 + 10 + 9 = 28. Then they calculated the percentage by multiplying 28 by 2 to get 56%.

Another student used a similar procedure but went a step further to deduce that since 100% represents 50 seeds, 1% represents 0.5 seeds. Then he inferred that since 28 is 3 more than 25, 3 seeds should represent 6% and added it to 50% to get 56%.

Prompting higher level of mathematical reasoning

Inductive/deductive reasoning

A higher-order task may also involve the generalisation of an observation or the formulation of conjectures. The items involving patterns (or linear relationships) in Testlet F elicited the students' ability to reason inductively and allowed them to conjecture the general form of the patterns. While the Year 3 and 5 pattern items were centred on finding the missing element in a two-column table or Fibonacci sequence, the Year 7 and 9 pattern items involved conjecturing a more advanced rule.

Example 1: Year 3

Questions 4 and 9 provide a glimpse of Year 3 students' ability to work with number patterns to reason inductively. Two strategies could be identified from the students' responses in Question 4. The first strategy involves working with the difference between the successive numbers in the first column (15 − 11 = 4 and 21 − 15 = 6) and second column (8 − 4 = 4), and extending the rule to the second column (6 + 8 = 14); or 11 + 4 = 15; 15 + 6 = 21; 4 + 4 = 8, 8 + 6 = 14. The second strategy involves calculating the difference between the numbers in the two columns (11 − 4 = 7), and then subtracting 7 from 21 to get 14. In conventional lessons, students at the Year 3 level generally may not be working with two-column patterns.

Students' verbatim comments

"We don't normally work with tables at school."

"Not many age questions and neither differences."

"We don't see age questions. Only questions like 'find the difference'."

Example 2: Year 7

Two items allowed us to sense the ability of the Year 7 students to reason inductively. Question 7 was the hardest item in the Testlet for the Year 7 students. When we asked the students to explain their solution, two of them changed their answer. Thus, essentially only 2 out of the 10 students were successful in this task. The first student used the following expression:

No. of white triangles = (No. of black triangles − 1) ÷ 2 = (6561 − 1) ÷ 2 = 3280. The second student worked as follows: She observed the pattern in the first row. Each number is multiplied by 3. She continued three more terms: 243 × 3 = 729; 729 × 3 = 2187; 2187 × 3 = 6561. Then, she could observe that if the terms in one column are added, they yield the next term in the pattern.

Example 3: Year 7

In Question 16, half of the number of students could identify the pattern from the given graphic as follows: 400 dots was interpreted as 4 squares; 1 square = 100 dots (400 ÷ 4); dots on each side of square = 10 ($100 = 10^2$); the number of the shape = number of dots on one side minus 1 (10 – 1 = 9). A few students used systematic guess-and-check and were successful while others could not solve it.

Example 4: Year 9

In Question 8, some of the students showed that they could reason inductively. For example, after comparing the corresponding numbers in the first and second rows, one student observed that if he doubled the number of white triangles (second row) and subtracted 1, this would equal the number of black triangles (first row). Then using the ‘opposite’ sequence of operations, he computed (6561 − 1) ÷ 2 = 3280 on the calculator. Another student observed that if he divided the number of black triangles by 2 and ‘round down’ the results, this will yield the number of white triangles. One student used the rule: $(x \div 2) - 0.5 = y$. A few students had studied linear relationships and could easily identify the pattern. However, some of the students could not find the pattern and struggled with this item. They were not conversant with particular strategies that would allow them to reason inductively.

Example 5: Year 9

In Question 2, a few students could deduce the rule relating the numbers in the first and second column. They could express the relationship between the two quantities symbolically as $y = x \times 3 + 5$. Others continued the sequence in the first row to 15 and found the corresponding value of y for each value of x. This primitive approach shows that some of the students had not yet developed the necessary skills to reason inductively.

Proportional reasoning

Year 5

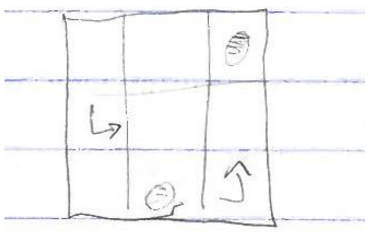
Proportional reasoning is important to solve Question 11. Those students who had well-articulated proportional reasoning ability could readily access the problem as follows: 10 g is one-tenth of 100 g; thus they computed one-tenth of 5 g to obtain 0.5 g, or since 100 is divided by 20 to get 5, 10 has to be divided by 20 too. For the second part, they knew that they had to apply the same multiplying factor (10) as in the first part. The incorrect answers were due to the inability to see the multiplicative (or proportional) relationship between the two quantities (10 g serving and 100 g serving). One student looked at the ‘resemblance’ of the numbers to answer the question, i.e., since the amount of sodium given in the first column involved hundredth (0.03 g), the number above it should be 0.05 g (in the hundredth as well). Similarly, she wrote the mass of sodium as 3 g based on the integer value ‘5’ g in the second column. Another student looked at the pattern in the answer in the first row in attempting to find the missing value in the second column: “I just worked that out looking at the top one because they are similar.”

Year 7

Question 12 revealed the extent to which the students could reason proportionally. Two strategies were apparent from the students’ work. The first strategy is commonly referred to as the unit-rate method, where they convert the cost to the lowest common factor—in this case, the cost of 1 kg of bananas and subtracted the difference. The second, more efficient strategy is commonly referred to as factor-of-change strategy, where they multiplied $8.48 × 5 and $13.99 × 3, and found the difference.

Spatial reasoning
Year 3

The two spatial items (Questions 3 and 5) in Year 3 required quite advanced forms of manipulation of images. The visualisation task in Questions 3 was taxing for some of the students as could be witnessed by the time they took to answer it. The main strategy was to visualise folding the nets given in the multiple choice options, one after the other (i.e., eliminating answers A, B and C), with accompanying gesture. In Question 5, a few students mentally rotated the grid. They pictured in their mind what the object would look like when turned and followed the relative positions of dots and arrows (in the given design) to locate the new position of dots and arrows. They used body movements (gestures) to mimic the turning action, such as turning their head and/or using their fingers to coordinate a 1/4 turn. One student chose to re-draw the given design on paper and turn the whole paper around a quarter turn (see diagram below). Most probably he could not visualise the transformation.



Year 5

Questions 2, 6, 7 and 14 require spatial visualisation which essentially refers to the creation of a mental image from visual/spatial information and the manipulation of the mental image. Spatial visualisation is often accompanied by gestures (such as body movements) in manipulating an image. For instance, in Question 2, some of the students mentioned how they visualised Nathan’s shoe to determine the print that it made on sand: “When I saw the picture I just tried to work it out as if I put the original (shoe).” Other students worked by eliminating the options given in the multiple choice as they focused on the relative positions of the tails of the lizards. We could also observe students looking at their shoes and trying to imagine how the design would look like, some flipping their shoes.

In Question 6, which required folding the net of a heptagonal-based prism, some of the students could readily see the prism when the net is folded. For instance, one of the students mentioned: "The shape of it just came to my head." However, others had difficulty to visualise the resulting prism:

Students' verbatim comments

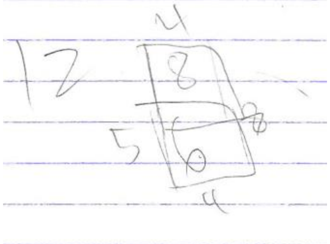
"I had a bit of struggle trying to fold the paper together in my head… I tried to (visualise) but it was not really coming in my head…My brain just could not do it."

"But it is hard to visualise. Three dimensional shape is one of my weaknesses."

A few students took the heptagon to be the more familiar hexagon and ended with the incorrect solution. Further, they took much time to visualise the action of folding the net. A few students attempted to draw the 3D object to count the number of edges as shown by the diagram below. Still other students did not differentiate clearly between edges and vertices. Those who got the answer 34 counted all the edges.

Solving such a problem on the computer (in contrast to pencil and paper) may require a different processing load. One of the students said that she always forgets where she starts to count the edges. This may not necessarily happen in paper-and-pencil mode since she can put a kind of mark once she counts it.

Similar to the previous visualisation item, Question 7 proved to be challenging for the students, firstly because it requires folding the net and secondly finding the greatest total. A few students could visualise the resulting cube with the numbers and add the opposite numbers. One student explained how he visualised folding the net with the numbers as follows: "First I visualised folding this part, 5, 6, 7, 8, and obviously it would fold…it would connect together." Others had to draw the cube and mark the numbers to be able to solve the problem (see diagram below).



The students took much time to determine/visualise which numbers were on the opposite sides. Question 7 is probably not within the scope of many Year 5 students who may be numerically agile but not necessarily have adequate visualisation skills. Those who could solve it apparently had done similar problems before.

The fourth visualisation item, Question 14, was the hardest item for the students. The cognitive interviews showed that it is not the complexity of the task that resulted in low performance. Rather, several students failed to realise that they had to count the base as well. Thus, they ended with the answer 16 rather than 14. Sometimes, students did not make the distinction between faces and edges. One student who was not sure about his answer mentioned: "Because you cannot see the whole thing and you have to visualise yourself." One student had to draw the net on paper and he could not visualise it: "I did try to visualise but it did not work well."

Students found this item different from school questions:

Students' verbatim comments

"They (at school) normally show the whole thing."

"No pyramid, shapes or nets question at school."

"At school, we just do cubes."

"At school, cube and cuboid only."

Year 7

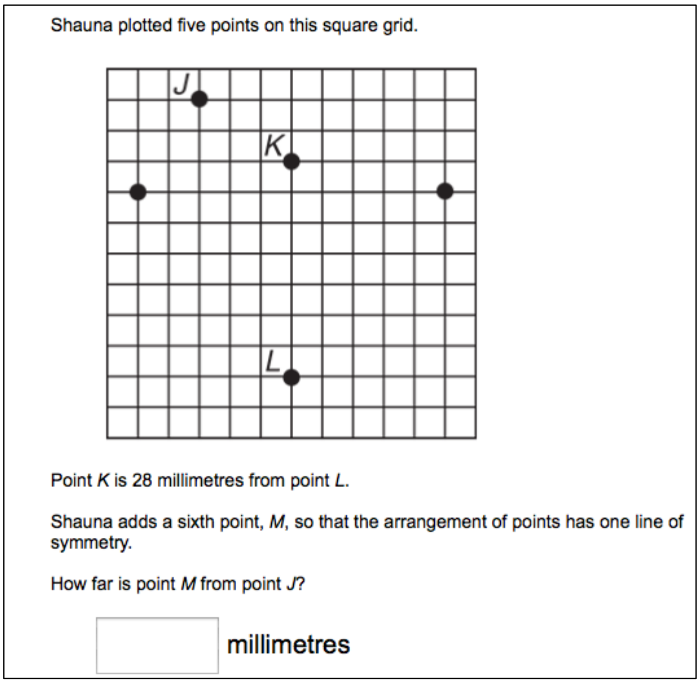
The common strategy to solve Question 3 was to visualise cutting the cross section of the cone: "Just imagined cutting a cone parallel to a base. So the cross section should be a circle." Those who missed it chose option B (triangle). They thought 'cross section' meant perpendicular to the base.

Similarly, in Question 9, most of the students mentioned that they visualised/imagined turning clockwise through a half turn: "Visualised the dancer's point of view and turned 1/4 way and an extra 1/4 turn." One student used the compass point as reference: "The audience is in the North direction. If he (the dancer) turns half way, he will be facing the backstage." Another student imagined the turning motion as in a clock. A few used hand gestures to mimic the turning action. Two students took 'half turn' to mean a 90-degree rotation.

Year 9

Three items from Year 9 (Questions 1, 12 and 16) explicitly required spatial reasoning ability. The common strategy to solve Question 1 was to add all the edges on top (7) and times the answer by 2 to get 14 and to finally add all the edges on the sides (7) to get a total of 21. One student clearly described how he visualised counting the edges: "Visualised the skeleton of the shape; took away the solid from the shape so that you could see the individual lines of the shape." Other students visualised twisting the arrow or dissembled the arrow into its constituent shapes.

In Question 12, mainly the identification of symmetry made this problem inaccessible for some of them. The students who missed the problem could not identify the correct location of the letter M (as they could not apply the required symmetry on the given grid). However, they could deduce that the length of 1 square = 4 mm.


Shauna plotted five points on this square grid.

Point *K* is 28 millimetres from point *L*.

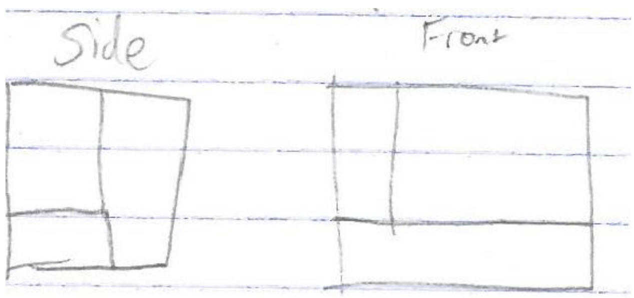
Shauna adds a sixth point, *M*, so that the arrangement of points has one line of symmetry.

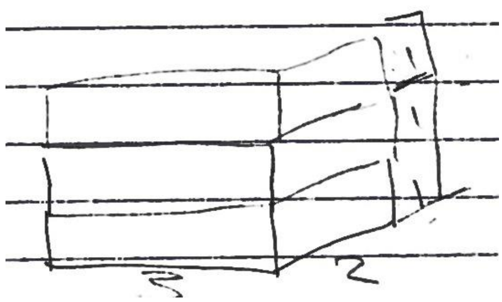
How far is point *M* from point *J*?

millimetres

Question 12 from the Year 9 items

In Question 16, those who answered it correctly deduced that the base needed to be a square. They used the fact that a common multiple of 2 and 3 is 6 to decide that the base should have 6 squares. Thus, the volume was calculated as 6 × 6 × 6 = 216. Others tried to make different diagrammatic configurations to stack the blocks to be able to visualise if the arrangements made a cube (see diagram below). Three of the students missed it because of computational mistakes such as 36 × 6 = 198, 36 × 6 = 316, or 6 × 6 × 6 = 252. One student was not clear about the meaning of ‘face to face’.





Algebraic reasoning

Only a few of the items in Testlet F required algebraic (or pre-algebraic) reasoning. At Year 5, the Fibonacci sequence item gave a glimpse of students' capacity for pre-algebraic reasoning in Question 3. One of the students referred to the first unknown term in the Fibonacci sequence by a place holder (x) in his solution procedure: "Make the question mark (given in the problem) x; so $x + 16 = 23$."

At the Year 9 level, the linear relationship item in Question 2 showed that a few students could express the relationship between the two quantities symbolically as $y = x \times 3 + 5$. In Question 8, one student inferred the rule $y = (x \div 2) - 0.5$ from the given table. In Question 10, a few students multiplied each side of the given equation $64 \div n = n \times n$ in solving it algebraically.

Exploiting the problem solver's intuitive knowledge

Estimation and systematic guess-and-check

Year 3

In the absence of a deterministic procedure to solve the problem situations, the students fell back on using estimation. For example, in Question 7, unable to infer that the length of one subdivision can be obtained by division, several students started with a guessed length such as 10, 20, 30, 40 and 50 and counted along the number line to see if they added up to 1000.

Similarly, in Question 12, those students who could not relate the strategy to multiplication (or repeated addition) and the corresponding division situation, visually compared a fork and a spoon. The length of the spoon was estimated to be more than half of the fork (i.e., greater than 7.5) and 8 was given as an answer. Those who obtained the correct answer first added $15 + 15 + 15$ or multiplied 15 by 3. However, not all of them could immediately deduce that they had to perform division to find the number of spoons. Thus, they attempted to add particular numbers (three times) such as 7, 7.5, 8 and eventually 9 to deduce the length of the spoon. In Question 9, one of the Year 3 students used a systematic guess-and-check. He started with 10 to observe that $10 + 16 = 26$. To adjust the sum to 23, he changed his initial guess to $10 - 3 = 7$.

Year 5

Students resorted to using estimation apparently when they did not know which mathematical procedure to use. As in the case of Year 3, a few Year 5 students used the more intuitive strategy to visually compare the lengths of the spoon to that of the fork in Question 9. For example, a few students observed that the spoon was about half of the fork and gave the answer 1/2 of 15. Others went one step further and argued that it is a bit more than 7.5 and gave answers such as 8 cm and 9 cm.

Similarly, the guess-and-check strategy seemed to be prompted when students did not know a deterministic procedure to solve a problem. In Question 4, some of the students proceeded by guess-and-check and systematic counting (e.g., observing that 500 is half way to 1000).

Year 7

In Question 8, only two students realised the equivalence of the given length (9 cm) and the three unknown lengths. The remaining students estimated the lengths of the three sides where measurements were not given. However, these estimations were not random. They used the measures of the known sides as benchmarks in their estimation. They tried to allocate values to the three bottom unknown horizontal lengths to sum them to 9. A few said that the problem did not include enough information to be able to solve it.

Year 9

In Question 10, a common strategy was to plug in trial values in $64 \div n = n \times n$ and see if that number satisfied both sides of the equation. The substitution of the values was not random. It was common to start with the value $n = 4$ as it is a factor of 64.

Factors That Enabled and Constrained Access to Testlet F

Research question 3.2: What understandings and strategy use most influence performance?

The performance of the students on the Testlet F items was influenced by the task features, the digital environment, as well as their knowledge and skills. We have commented on the task features and the influence of the digital media in earlier sections. Now we explain what understandings and strategy use influenced students' performance on the items, both in terms of enabling and constraining access to the problems. We comment on three main aspects, namely the foundational knowledge of the students, their analytical skills and their reasoning abilities.

Foundational knowledge of the students

Year 3

The group of Year 3 students who were at the expected level showed much flexibility in performing arithmetic operations on numbers (as in Questions 1 and 8). However, not all of them were agile in operating on a number involving a thousand as in Question 7. They could manipulate familiar fractions involving halving and taking a quarter of a quantity as in Question 10. However, the variation in the item did generate an obstacle for some of them. They mentioned that the phrase 'is the same as half of' was not clear. It is more likely the processing load of the item that led to such responses.

> *Students' verbatim comments*
>
> "I understand 'a quarter of 120' but I don't get it that it says 'is the same as a half of'."
> "The word was a bit confusing."
> "I don't really understand this fully…. Probably I am having trouble with the question…a quarter of 120 is the same as half of."

Questions 6, 9, 11 and 12 were hard for the students. Question 6 shows that linear measurement is one of the areas of weakness. Only 50% of the students could answer this item correctly. They experienced difficulties in deducing the magnitude of one partition and in counting along the incomplete height chart (without starting and end point) involving metres and centimetres. Question 11 also drew our attention as to whether the students had the requisite foundational knowledge. Some of the students were not conversant with the concept of area. It appears that they had not been in touch with this concept in recent times. They used their common sense to count the squares. Those who missed it did not count the central square or counted all the given squares ($5 \times 5 = 25$). It appears the difficulty came from students not being familiar with 'counting' area on a grid. To summarise, in general, the Year 3 students had most of the foundational knowledge to attempt the Testlet F items. It was more the demand of the problems in terms of their formulation that pitched them at a high-order level that made them challenging.

Year 5

The group of Year 5 students who were at the expected level demonstrated that they had good numerical flexibility as could be observed from Questions 1, 3, 4, 5 and 8. However, they were not always confident in using decimals as in Question 11. Similarly, they showed weaknesses in doing division by a two-digit divisor as in Question 13. The students clearly understood that this question required the division of 390 by 13. However, the division was challenging for a few students. One student mentioned: "But I am not good at division. So I can't get what $390 \div 13$ is." He attempted to find which number multiplied by 13 gives 390 and tried 18×13, 23×13 and so forth without success. Thus, in the absence of the mathematical procedure, they resorted to the guess-and-check procedure. For example, in the foregoing problem, one student started with an estimation of 30 cm for one unit and counted as follows: 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 330, 360, 390. Questions 6, 10, 12 and 14 were hard for the students. Questions 6 and 14 required spatial reasoning ability, a mathematical behaviour where students do not always show flexibility. Question 10 also involved elements of spatial reasoning besides the notion of volume. Question 12 was a novel situation for the students and only those who could apparently think in terms of an unknown could access the problem.

Year 7

The one-to-one interviews and the retrospective analyses that we carried out allow us to infer that, in general, this cohort of Year 7 students was conversant with the concepts, facts, definitions, procedures and algorithms typical of this year level. From the first two questions, we could observe their fluency with numbers and Question 6 gave us an apercu of their ability to work with fractions. Questions 7, 12, 14 and 16 were hard for the students. Some of the students were not fluent in the division of decimals as could be inferred by the difficulties they experienced in Question 14. After adding $1.82 + 2.73 = 4.55$, a few students attempted to use multiplication to determine which number times 182 gives 455. Two students crossed out 1.82 from the numerator and denominator and either gave the answer 2.73 or $1 + 2.73 = 3.73$. Questions 7 and 16 revealed their lack of fluency in reasoning inductively, while Question 12 indicated their lack of facility to reason proportionally when decimals are involved. Moreover, the responses collected from Question 15 revealed that some of the students did not have a strong foundation of ratios. Some of them seemed to have used guesses. They split 1350 into 2 and tried to decrease the result (675) to see if it satisfied the conditions of the problem.

Year 9

Being in the top 30% of their class cohort, we expected these students to be fluent in working with whole numbers (including negative numbers), fractions, decimals, percentages, ratios and proportions. Similarly, we expected them to be comfortable with algebraic manipulations and geometrical concepts. While a majority of them did show such competencies, a few of them had not yet fully developed these mathematical skills. We got a sense of their ability to work with decimals in Question 6, where they had to estimate the result of the multiplication of two decimals: 0.263×0.767. The majority of them could readily deduce the answer: "When you times something by a decimal less than 1, it goes down; and 0.263 is between 0 and 1, so the answer has to be less than 0.263." or "When you multiply a number by a number in decimals, it gets smaller." However, when they were asked to divide two decimals in Question 11, we could see that several of them lacked confidence. The students changed the division problem to a multiplication problem as follows: $0.12 \times$ (required number) $= 12.24$. For example, one student deduced that $0.12 \times 100 = 12$ and $0.12 \times 2 = 0.24$. This strategy allowed him to obtain the multiplier as 102. Clearly, the students lacked confidence to work with division of decimals. For example, one of the students mentioned that there is a way to turn a decimal division into a decimal multiplication but he could not remember. He acknowledged: "I am not very good in decimals." A few students seemed to have a faulty concept of decimal division. For example, one student divided 12 (from 12.24) by 12 to get 1. Then she divided 0.24 by 12 to get 0.02. Thus, the final answer was incorrectly given as 1.02. We could also observe that they tended to make computational mistakes.

As we proceeded further with the interview, we could see their versatility to work with a multiplicative relationship in Question 7. The common strategy to solve this problem was to interpret the multiplicative relationship as a ratio or fraction. Using a ratio approach, they proceeded as follows: Pear:Apple = 1:9 = 4:36; $36 \div 9 = 4$; $36 - 4 = 32$. Using a fraction approach they proceeded as follows: 1 in 9 chance is the same as 4 in 36 chance ($1/9 = 4/36$); $36 - 4 = 32$. Question 13 gave them the opportunity to show their flexibly to work with percentages.

Questions 3, 5, 8, 11 and 16 were hard for the Year 9 students. From Question 3, we could deduce that not all of them were conversant in working with large numbers. Those who got this item wrong divided 23 million by 10 million rather than the other way. Sometimes they made computational and estimation mistakes due to the large numbers. From Question 4, we could observe that some of them did not have a good mastery of the concept of speed. A few students obtained the answer using estimation rather than the required method. Some students could perform the calculation for the first part of the journey—i.e., Indra took $10 \div 6 = 1\ 2/3$ hrs or 1 hr 40 mins—but were not sure how to proceed with the second part or made computational errors.

Similarly their concept of geometry was not always consistent. For example, in Question 5, 7 out of the 10 students answered incorrectly. The students were unsure about the meaning of 'diagonals that cross at right angles'. A few just focused on the right angles and chose option A. Some just guessed the answer. Similarly, in Question 1, one student mentioned that he was not very sure what 'edges' meant. Like the Year 7 students, the Year 9s were not always capable of reasoning inductively as elicited by Question 8. Moreover, Question 16 indicated that the students were yet to fully develop their spatial reasoning ability in novel situations.

Analytical skills

A range of analytical skills could be observed from the data gathered from the students, such as making a diagram or using estimation and systematic guess-and-check. They applied these skills with varied levels of flexibility. In this section, we comment specifically of their graphic decoding skills as they were critical in the solution of some of the problems.

Decoding graphics

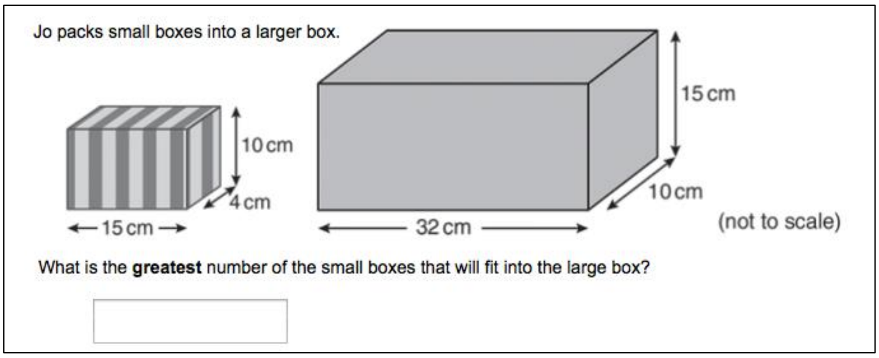
Year 3

Some of the Year 3 students experienced difficulties in interpreting the scale in the graphics given in Questions 6 and 7. Question 6 was apparently a novel situation for most of the students, although a few of them had seen a similar question in their home mathematics textbooks. The small size of the subdivision imposed much demand on the students. One of the students got 108 cm as the answer, apparently because of the small size of the partition on the grid. Students who answered this question correctly coordinated their movements along the subdivisions and the size of each partition. Some of the students could not identify the size of one subdivision, which they assumed to be 10 cm to get an answer of 170.

Another item that requires graphic decoding skills is Question 7. The common strategy was to start with an estimation such as 10, 20, 30, 40 and 50 and count along the number line. Some of the students initially (incorrectly) took the end number as 100 instead of 1000. Further, counting the small and 'lengthy' (20) subdivisions required much focus on the screen. Often students had to move close to the screen and restart counting the small subdivisions. It is to be noted that 10 out of the 12 Year 3 questions involve a graphic, either in the form of a diagram, picture, net, table, design, ruler, number line or grid.

Year 5

Question 10 provides an insightful example where a graphic may have led the students to use an incorrect procedure. In fact, this item was incorrectly answered by a majority of the students, although they thought that they answered it correctly. The common strategy was to compare the corresponding lengths based on the given orientation of the two boxes (i.e., the number of times 15 cm goes in 32 cm; the number of times 4 cm goes in 10 cm; and the number of times 10 cm goes in 15 cm). For example, one of the students obtained his answer as 4 and gave the following explanation: "15 can go to 32 twice. 4 can go to 10 twice and there are 4 boxes at the bottom. (He visualised how the boxes were packed.) There will be a bit of space left." They did not pay attention to the fact that two of the given lengths were equal in the two boxes.

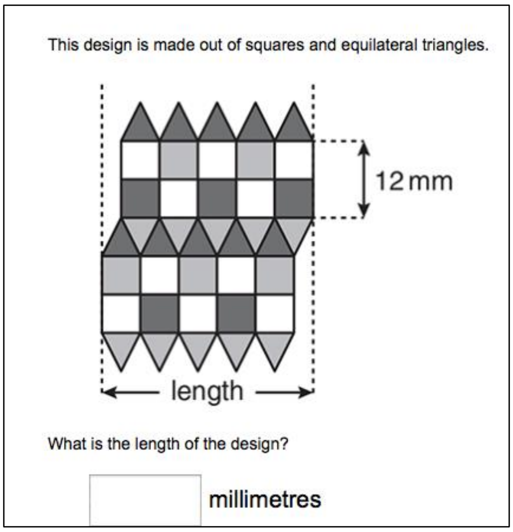


Question 10 of the Year 5 items

Question 13 is another item where the graphic imposed much demand. Some students could not figure out what the question was asking for. A grid involves both horizontal and vertical measurement and this prompted a few students to count the number of squares covering the car as 13 × 4 = 52. Others merely counted the number of subdivisions as 13 cm (or miscounted them as 14 cm) and did not find the necessity to use the 390.

Year 7

In Question 13, most of the students calculated the length of one square (6 mm) to find the total length as 33 cm. Some students gave 30 mm as their first answer and when asked to explain their answer, they realised that they did not count the little space before the dotted line (see diagram below). One of the students calculated the height of the design instead. Another student could not find the length of the end part of the design (3 cm) and approximated it as 2 cm.

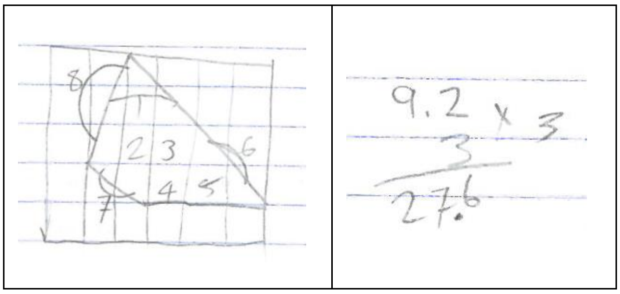


Question 13 of the Year 7 items

In general, the graphics imposed much less demand on the Year 7 and 9 students than on the Year 3 and Year 5 students.

Year 9

The challenge in Question 15 was in finding the area of the irregular shape. It was quite demanding for some students to keep track of the mental subdivision of the area on the screen. One student redrew the grid on paper (see student working below). Sometimes they just did not know the procedure to find the area of irregular shapes. Generally, they answered this question incorrectly because they did not calculate the area accurately. They approximated the area to 9 rather than finding the precise area as 9.5 units squared.



Reasoning abilities

Not all the students could reason inductively as evidenced by the pattern items. Similarly, lack of flexibility in proportional reasoning in situations involving decimals also influenced students' access to the problems. Another more critical resource necessary for attempting Testlet F items is spatial reasoning. During the interviews, we could see the efforts that the students were putting in to attend to the spatial tasks. Some of the students were comfortable with this form of reasoning involving the manipulation of images while others were not always at ease. Finally, only a couple of students were capable of reasoning algebraically at the Year 9 level, although only a few items assessed this form of mathematical behaviour.

Students' Perceptions of Testlet F

Research question 3.3: What perceptions do the students have of Testlet F items? How do these identify with the intended assessment outcomes?

Perceptions of questions

After the students had solved each task, the interviewer asked the students to rate the level of difficulty of the questions (along a scale of 'Easy', 'Just OK', 'Hard'). Table 14 summarises students' perception of the questions by grade levels. At times it was difficult to classify the items into the three categories based on the students' responses as equal numbers were in each category and the sample size was relatively small. We have looked at the general tendency. Table 14 highlights that across all four year levels, the majority of questions were perceived as easy and OK. Only a few questions were perceived as being hard. At times, the students did rate a question as easy when in fact their solution was incorrect. For example, the Year 3 students rated Question 12 as 'Easy' but more than half of them answered incorrectly.

A look at the nature of the questions gives an indication for the reasons why they were rated as 'Hard'. For instance, the Year 3 students rated Questions 5 and 6 as being hard. Question 5 involves spatial reasoning while Question 6 requires students to read a measurement on a partial ruler. These concepts are traditionally known to be demanding. A similar pattern can be seen for the Year 5s where Questions 6 and 7 involve spatial items. The Year 5 students equally rated Questions 8 and 12 as being hard probably because these require working with unknown quantities. As for the Year 7 students, only item 15 was rated as hard. This question involves working with a ratio. Similarly, the Year 9 students rated only one item as being hard, namely Question 4 involving a two-stage journey and requiring the calculation of speed.

Table 14. Students' Evaluation of the Questions by Year Level

Level of difficulty	Year 3	Year 5	Year 7	Year 9
Easy	1,2,4,10, 11,12	1,3,5,9, 10,11,13,14	1,3,4,5,6, 9,11	1,3,6,7,8,9, 10,12,13,14
OK	3,7,8,9	2,4	2,7,8,10, 12,13,14,16	2,5,11,15,16
Hard	5,6	6,7,8,12	15	4

In the following sections, we present some verbatim comments to show the qualitative responses given by the students when they were asked to describe in what ways were the items different from the problems they encounter at school.

Students' verbatim comments on items rated as 'OK'

Year 3 – Question 3

It seems that the net problems done at school are much easier and are often provided in prototypical representations.

> *Students' verbatim comments*
>
> "The shape of the net is unfamiliar."
>
> "Not really done this type before."
>
> "We don't really work with nets at school."

Year 5 – Question 2

Several students mentioned that this type of problem is not familiar to them.

> *Students' verbatim comments*
>
> "We haven't seen any of these at school."
>
> "Not done in school."
>
> "I rarely get any picture question at school."
>
> "We don't usually do this type of questions."

Students' verbatim comments on items rated as 'Hard'

Year 3 – Question 5

This question was seen as much different and more challenging than the ones done at school.

> *Students' verbatim comments*
>
> "Harder than normal."
>
> "The type of question is not done in class."
>
> "We haven't done turns."
>
> "Not done it before."
>
> "We don't do too much turning of shapes with the missing (elements)."

Year 5 – Question 6

This question was rated as being hard primarily because of the complexity of the net. A number of students were not confident about their answers unlike the other numerical problems. According to some of the students this kind of problem is not common at school.

> *Students' verbatim comments*
>
> "The ones at school are easier."
>
> "We do nets and edges with cubes only."
>
> "We don't normally do this kind of problems. At school, we use cut outs from paper."

Year 5 – Question 8

This item was regarded as hard. The students understood the problem but did not have a deterministic method to approach it. A few students acknowledged that they could not solve it: "I can't figure out"; "no clue".

> *Students' verbatim comments*
>
> "I have done this before. I did it in NAPLAN but they put the number for the other people."
>
> "Never been taught before."
>
> "At the school, we don't have just words but numbers and pictures."
>
> "We don't use the same person twice."
>
> "Because of the missing number, I cannot figure out the question."
>
> "Never have done problems with no starting number."

Year 5 – Question 12

This item was challenging for the students and several of them were unsure about their answers.

> *Students' verbatim comments*
>
> "We don't do these type of tricky questions."
>
> "No such word real-life situation in school."
>
> "Have not seen it (the question) before."

Perceptions of problem solving in a computer-based environment

The interviewers asked the students to describe if they thought they had to solve questions on the computer differently than on pencil-and-paper via the following question: In what way is solving mathematics problems different on the computer as compared to solving it on paper? Some of their statements are outlined below.

Year 3 and Year 5

At Year 3 and Year 5, students seem to have a preference for computer-based tests rather than paper. They say that it is more fun to work on the computer (like *Mathletics*). It is easy to write and make changes faster using the backspace key. It is less straining for the fingers. Some of them feel that information is more clearly presented on computer than on paper. Further, on paper-based tests, one has to colour circles properly on the answer sheet which is apparently inconvenient. Students' comments on the computer-based test are given below:

> *Students' verbatim comments*
>
> "It is easier to write things. My hand writing is bad."
>
> "On paper your hands get sweaty; hand writing gets mess up; the pencil gets bold."
>
> "On computer, it is easy and fast."
>
> "Faster on computer."
>
> "Takes longer to write (on paper)."
>
> "In a computer people can't cheat. People cheat on paper."
>
> "I prefer computer; more modern and high tech."
>
> "Writing takes time (on paper)."
>
> "Because you don't have to turn pages. Hands don't get sore."

The few students who prefer paper-based tests feel that they have more control over the test on paper. It is more concretely accessible. One can turn the page and rework previous questions. One student mentioned that he does not trust himself on computer in contrast to paper.

> *Students' verbatim comments*
>
> "I feel more confident on paper. If it has a diagram, I can look at it more closely. I can point to it. On paper I can put a dot. On paper you have more control. You can look at the paper directly. You can point directly rather than (using) your eyes."
>
> "Feel more comfortable on paper."
>
> "Easier to think when using paper."
>
> "On computer you have to move mouse around and type."
>
> "You can draw on paper."

Year 7 and 9

There were mixed feelings; some preferred paper-and-pencil, others preferred computer and, still for some, the test medium did not matter. However, there was more tendency towards paper-and-pencil at both Years 7 and 9.

Those who preferred the paper-and-pencil mode find it easier to work on paper since they can write on the questionnaire, rub off and make a diagram. A few students feel that working mathematics problems on the computer does not give you much flexibility to draw and write around the question. They feel that it is more time consuming. They can concentrate more on paper than on the computer. They also prefer paper because they can go back and change an answer as they want. Further, a few students mentioned that reading the problem on the computer takes more time to process the information. They were more at ease with paper.

> *Students' verbatim comments*
>
> "Paper definitely; because I normally draw around it; clicking the calculator was slower; the working out process was slow."
>
> "On paper, because I can use working out, draw diagrams easier."
>
> "On paper. You can work out a lot easier to visualise and draw it whereas on the computer you can't do that."
>
> "You can't write down what you are actually going through on a computer."
>
> "On the computer, you don't feel serious, looking forth and back."
>
> "It is easier to think and focus on paper; more distracted on the computer."
>
> "On computer, you can't show your working."
>
> "I prefer paper because I am more used to it."
>
> "On paper, one is more concentrated."
>
> "Paper—more chance to go back to change (answer)."

Those who prefer to do the test on computer think that it is easier to type and write or to change their answer. It is faster to key in the answer unlike the current answer page on paper.

Students' verbatim comments

"On computer. It makes it easier. If you make a mistake you can easily get rid of it."

"Prefer computer; generally easier to type and write."

"Quicker to do test on computer, instead of writing."

Some students said they enjoyed the challenge in the Testlet as some of the problems were novel for them. Others said that it was not necessarily challenging but they had forgotten some of the concepts. A few students also commented that they prefer to work mathematics mentally independent of whether the medium is paper or computer.

Key Findings (KF) and Recommendations (R)

Priority 1: Establish the extent to which the proposed challenging items in Testlet F provide adequate testing context for highly capable students.

KF1.1 These items constructed for Testlet F were designed to evoke high-order thinking. This is achieved by creating items that require the application and/or analysis of conventional mathematical concepts and procedures. In general, the language used to formulate the questions was clear. The main characteristic of poor design was associated with inconsistencies in graphic clarity and font size.

The items in Testlet F cover a broad range of areas across the mathematics curriculum. Further, Mayer's framework (2002) showed that for each of the four year levels, the majority of items require conceptual knowledge in contrast to purely mechanical procedural knowledge. In terms of cognitive processing, most of the items require higher-order skills such as analysis and application. Our detailed analysis of the structure of the problems showed that some of the tasks were however more procedural than conceptual.

R1.1 The majority of items across the four year levels tend to be concentrated on Number and Algebra and Measurement and Geometry. A more balanced Testlet F should include more items from the strand Statistics and Probability.

KF1.2 Some inadvertent effects were observed emanating from the visual weight of the graphics in the items in a digital environment. For instance, at Year 3 level, some of the graphics imposed additional demands for the students. Counting the small subdivisions on the screen was quite time consuming both for the Year 3 and Year 5 students. Some of the students counted the subdivisions by pointing their pencils to the screen. Similarly, finding the area of an irregular polygon by counting squares on the computer screen may be demanding. One has to mentally or visually split the region and keep track of the different areas as in the last item in Year 9. Further, the fonts used in some of the items were not always homogeneous, prompting students to focus on particular aspects of the problem and ignoring other important ones.

R1.2 It is recommended that the display of information in item design be tightly scrutinised to ensure that inadvertent effects do not bring additional load in the item. Counting subdivisions on the computer screen may be time consuming and it suggested that some form of online markers be made available to the students to help them benchmark their counting (on the computer screen).

KF1.3 Many students did not notice that a calculator was available for some of the items. The interviewers often had to tell the participants that a calculator was included (whenever the calculator first appeared in the Testlet). Another technical issue is that when the calculator opens, it covers the question or part of the question. We had to tell the students that the calculator could be moved. Further, the keys on the calculator as well as its functioning did create some issues.

R1.3 The calculator should be made more visible on the screen and its keys and operational capacities improved. As a more general recommendation, we would suggest that a tutorial be included (in the form of a demonstration) at the start of the test to remind students about the availability of resources in the online environment. For instance, the demonstration may illustrate how to move the calculator. A more direct suggestion is to allow the calculator to pop up on the side of the screen for questions where it is allowed to be used. The tutorial may serve other functions such as informing students whether they can move backwards to check their answers. In some of the problems, the answer space accepts only decimals and not fractions and the tutorial can be very useful here.

Priority 2: Examine the performance of such students on Testlet F of the tailored test.

KF2.1 These capable students demonstrated both flexibility and fluency in their problem solving. These skills were necessary to engage with these tasks since most items required students to use application and analysis processes. It was evident that some of the Year 3 students lacked the capacity to think flexibly when compared to the other three cohorts. This is unsurprising, given the students maturity.

R2.1 We commend the fact that the task designers were able to construct items that required higher-order cognitive processes. We encourage ACARA to maintain this design "stance" and not simply increase task complexity via increased content difficulty.

Priority 3: Monitor the students' knowledge, thinking skills and strategy use, and how these relate to the intended assessment outcomes envisaged by the test developers.

KF3.1 The students employed a range of problem-solving strategies to solve the items. The students used diagrams, sketches and tables, looked for patterns and other non-algorithmic approaches to make their way through the problems. Such analytical skills are crucial in dealing with novel situations, characteristic of several Testlet F items. The students particularly exploited their knowledge of numbers in either written or mental forms to open solution paths. The Testlet F items prompted them to coordinate multiple pieces of information in the various multi-stage problems. Often, the same problem was solved in different ways by the students. Four types of higher-level mathematical reasoning were particularly apparent from the students' responses, namely inductive/deductive reasoning, proportional reasoning, spatial reasoning and to a lesser extent, algebraic reasoning.

R3.1 Test designers may not always look at the mode of reasoning in designing items. They may be more inclined to focus on content and the structural nuances that elicit higher-order thinking. We would suggest that Testlet F include more items associated with proportional reasoning and pre-algebraic/algebraic reasoning. These forms of reasoning are known to be demanding for elementary and middle school students.

KF3.2 Typically these students possessed sound generic problem-solving skills but may have lacked foundational knowledge. Relatively elementary concepts such as decimals, diagonals or performing a division at times prevented the students from getting the correct answer. Consequently, errors occurred when the students lacked flexibility to use higher-order reasoning such as inductive, proportional and spatial reasoning. We appreciate the tension between introducing more difficult content knowledge and exposing students to high-order thinking.

R3.2 From a research perspective, it would be prudent of ACARA to explicitly assess foundations skills in easier testlets. The application of these concepts could then be assessed via higher-order problem solving in Testlet F. Such a design would provide both diagnostic and research-based opportunities for a range of stakeholders.

KF3.3 Our observations indicated that many students encountered challenges in solving the spatial tasks. This is an important finding especially since NAPLAN items seem to give much consideration to this form of reasoning.

R3.3 ACARA be encouraged to ensure that the numeracy assessment frameworks point to the importance of visual and spatial reasoning in solving graphics-based NAPLAN tasks. This is increasingly important given the move to online assessment.

KF3.4 Estimation and systematic guess-and-check strategies were observed across all four year levels. On the one hand, the use of these more primitive strategies suggests that the students were exposed to challenging tasks. On the other, these fall back and non-deterministic strategies also suggest that these students are yet to develop more advanced mathematical abilities.

R3.4 Findings suggest that these students are able to produce correct answers despite not drawing on the higher-order problem-solving skills the tasks were designed to elicit. Consequently, reporting may need to be based on understanding of concepts and possible strategies used by students to solve the tasks, including the more primitive strategies that these students often fall back to.

KF3.5 During the cognitive interviews, we could observe that the students took much time to work out some of the items and to complete the Testlet as a whole. For some of the questions they took over five minutes.

R3.5 It should be recognised that students will attempt these higher-order questions after completing two testlets. Thus the number of items to be included in Testlet F needs to be judiciously thought through, as well as the time allocated. We recommend that less is best.

KF3.6 We had to remind some of the students that paper-and-pencil was available as they tried to solve the problems mentally. Some of the students tend not to use paper-and-pencil. On the other hand, we encountered students who would write their stepwise methods for each problem on paper. Moreover, at times, some students tend to rush through the problems and do not spend time verifying the soundness of their answers. In the cognitive interviews, often students changed their incorrect answers when we asked them to explain their thinking.

R3.6 Solving mathematics tasks in an online environment requires sophisticated levels of digital literacy. Students need to be exposed to these skills in the classroom otherwise, assessment results could be problematic. It is necessary for students to develop fluency within the digital assessment environment.

KF3.7 In general, the students regarded the items to be ‘easy’ or ‘moderately difficult’. Not surprisingly, the students had a positive affect towards mathematics. Interestingly, this was the case even though their answers were not always correct. Perception is closely associated to success. It is encouraging to see that students rated the majority of the items within their reach.

KF3.8 At Year 3 and 5, students tend to prefer doing the test on computer while at Year 7 and 9 there was a preference for paper-and-pencil mode. The students felt that they have more flexibility in working around the problem on paper than on computer. They can concentrate more on paper-and-pencil, and interpreting the problem on the computer requires more effort. Those who preferred a computer-based test did so because they think that it is easier to key-in and change the answer on the computer than on paper.

R3.8 It is important to be mindful of students' preferences for test mode since it may inadvertently influence performance. A growing body of research literature (including our own work) shows that test mode does influence performance (e.g., Bennett et al., 2008; Lowrie & Logan, 2015). We suggest that ACARA continue research in this area, especially differences between student performance in pencil-and-paper and digital forms.

References

Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., & Yan, F. (2008). Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *The Journal of Technology, Learning and Assessment*, *6*(9). Retrieved March 4, 2014 from http://www.jtla.org.

Diezmann C. M., & Lowrie, T. (2012). Learning to think spatially: What do students 'see' in numeracy test items. *International Journal of Science and Mathematics Education, 10,* 1469–1490.

Lamon, S. J. (2007). Rational numbers and proportional reasoning: Towards a theoretical framework for research. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 629–667). Charlotte, NC: Information Age Publishing.

Lowrie, T., & Logan, T. (2015). The role of test-mode effect: Implications for assessment practices and item design. In C. Vistro-Yu (Ed.), *In pursuit of quality mathematics for all* (Proceedings of the 7th ICMI-East Asia Regional Conference on Mathematics Education, pp. 649–656). Cebu, Philippines: Philippine Council of Mathematics Teacher Educators (MATHTED), Inc.

Mayer, R. E. (2002). A taxonomy for computer-based assessment of problem solving. *Computers in Human Behavior*, *18*, 623–632.

Peressini, D., & Webb, N. (1999). Analyzing mathematical reasoning in students' responses across multiple performance assessment tasks. In L. V. Stiff & F. R. Curcio (Eds.), *Developing mathematical reasoning in grades K-12 (1999 Yearbook)* (pp. 157-174). Reston, VA: The National Council of Teachers of Mathematics, Inc.

Resnick, L.B. (1987). *Education and learning to think*. Washington, D.C.: National Academy Press.

Smith, M. S., & Stein, M. K. (1998). Selecting and creating mathematical tasks: From research to practice. *Mathematics Teaching in the Middle School*, *3*(5), 344–350.