

Improving the education outcomes
for all young Australians

acara AUSTRALIAN CURRICULUM,
ASSESSMENT AND
REPORTING AUTHORITY



National Assessment and Surveys Online Program

Tailored test design study 2013: Cognitive interviews





National Assessment and Surveys Online Program

Tailored test design study 2013: Cognitive interviews

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) and Stephen Phillip (numeracy).

Report author(s)

Educational Assessment Australia, working with ACARA staff (Dr Goran Lazendic, Dr Kelvin D. Gregory, Anna Cohen and Stephen Phillip), delivered the cognitive interviews in reading and numeracy and developed the report entitled *NAPLAN Online Tailored Test Design August 2013: Cognitive Interviews – Numeracy and Reading*.

National Assessment and Surveys Online Program

The National Assessment and Surveys Online Program, funded by the Australian Government Department of Education, 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 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

© Australian Curriculum, Assessment and Reporting Authority 2014

This work is copyright. You may download, display, print and reproduce this material in unaltered form only (retaining this notice) for your personal, non-commercial use or use within your organisation.

All other rights are reserved. Requests and inquiries concerning reproduction and rights should be addressed to:

ACARA Copyright Administration, ACARA
Level 10 255 Pitt Street
Sydney NSW 2000
Email: info@acara.edu.au

The appropriate citation for this report is:

Educational Assessment Australia (EAA), November 2013, *NAPLAN Online Tailored Test Design August 2013: Cognitive Interviews Numeracy and Reading*, UNSW Global Australia, University of New South Wales, Sydney. Prepared for the Australian Curriculum, Assessment and Reporting Authority.

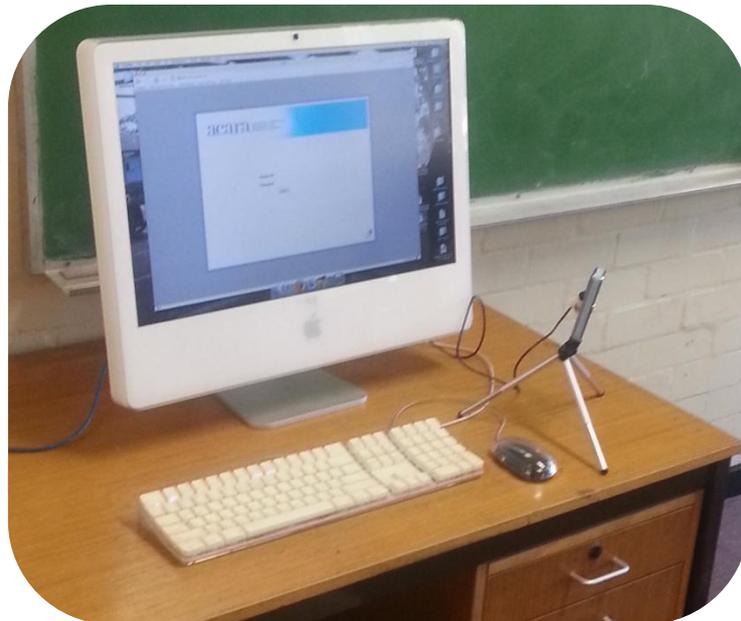


National Assessment and Surveys Online Program

NAPLAN Online Tailored Test Design August 2013

Cognitive Interviews – Numeracy and Reading

Part 1: Joint Summary Findings for Numeracy and Reading



Contents

Executive summary.....	4
Introduction.....	5
Background.....	5
The NASOP multistage model.....	5
Methodology.....	6
Protocol development.....	6
Sampling approach.....	7
Final sample.....	8
Findings: Overview.....	13
Theme: Branching.....	13
Theme: Computer-based test experience.....	17
Theme: Engagement.....	20
Findings: Domain-specific overviews.....	22
Numeracy-specific issues.....	22
Reading-specific issues.....	22
Conclusions and recommendations.....	24
Branching/Tailored test design.....	24
Computer-based test experience.....	24
Engagement.....	24
Numeracy.....	24
Reading.....	24
Further study.....	24
Appendix: Cognitive Interview Protocol.....	25

Tables

Table 1 Difficulty labels for pathways.....	5
Table 2 Difficulty ranges for testlets	5
Table 3 Targeting of items by method.....	7
Table 4 Proposed school sample by state and sector	7
Table 5 Proposed school sample by state and location.....	8
Table 6 Proposed school sample by area	8
Table 7 Final school sample by state and sector.....	9
Table 8 Final school sample by state and location	9
Table 9 Geographical distribution of schools in final sample	9
Table 10 Numbers of students interviewed by domain	10
Table 11 Students interviewed by location and sector.....	10
Table 12 Percentage of students in each domain by state and location	10
Table 13 Percentage of students by domain and gender.....	10
Table 14 Percentage of students who followed a given pathway	11
Table 15 Free and fixed pathways by domain	11
Table 16 Platforms observed by domain	11
Table 17 Computer use in school.....	12
Table 18 Computer use at home	12
Table 19 Touch screen phone use.....	12
Table 20 Comparison of first and second computer use questions.....	12
Table 21 Did the test differ from NAPLAN?	13
Table 22 Was the test easier or harder than NAPLAN?	13
Table 23 Comparative perceived test difficulty by pathway and domain (percentage of row).....	14
Table 24 Comparative perceived test difficulty by pathway (percentage of column).....	14
Table 25 Student perception of difficulty progression by domain	14
Table 26 Perceived difficulty progression by pathway (percentage of column)	15
Table 27 Perceived difficulty progression by domain and pathway (percentage of row)	15
Table 28 Students noticing change from first to second testlet by domain	15
Table 29 Students noticing change from second to third testlet by domain	15
Table 30 Comparison of students noticing changes between testlets	16
Table 31 Students noticing change from first to second testlet by pathway.....	16
Table 32 Students noticing change from second to third testlet by pathway.....	16
Table 33 Difference from NAPLAN: Responding to items	17
Table 34 Difference from NAPLAN: Assorted navigation/timing issues	17
Table 35 Presentation issues by domain	18
Table 36 Presentation issues by platform	18
Table 37 Reported engagement by domain	20
Table 38 Preferred testlet by domain	20
Table 39 Preferred testlet (percentage of column).....	20
Table 40 Effect of difficulty of third testlet	21
Table 41 Effect of difficulty of third testlet on preference by domain	21
Table 42 Effect of difficulty of third testlet on preference by pathway.....	21

[NOTE: Due to rounding, some sets of percentages do not appear to total exactly 100%]

Executive summary

The National Assessment and Surveys Online Program (NASOP) is a series of projects by the Australian Curriculum Assessment and Reporting Authority (ACARA) that investigates the online delivery of ACARA's assessment programs. As part of NASOP, a trial of online reading and numeracy tests took place in August 2013.

The online tests were designed around a partially adaptive multistage test model. This model consists of a series of mini-tests (known as testlets or modules) of varying difficulty. In this model students follow different paths through the testlets, depending on their score in each testlet. The purpose of this model is to enable students to spend more time in the test being assessed on knowledge and skills close to their level of ability. The 'branching' nature of the model was an important aspect of the study and is a significant innovation.

EAA was commissioned by ACARA to conduct cognitive interviews to investigate the impact of the multistage-branching design on students' test-taking behaviour and to gather information about students' interaction with the testlets at key branching points of the test.

EAA conducted two studies: one with the Numeracy test and one with the Reading test. Both studies used a common methodology that focused on three major themes – Branching, Computer-based test experience and Engagement. Within this methodology, specific questions were tailored around issues specific to either Numeracy or Reading.

A total of 16 schools in New South Wales and Queensland were visited over four weeks. These schools included schools in both metropolitan and regional areas, as well as schools from a variety of sectors. A total of 85 students were interviewed (44 for Numeracy and 41 in Reading) across Years 3, 5, 7 and 9. Interviews were conducted by EAA staff members with experience in cognitive interviewing as well as an understanding of the construct being tested. A combined video recording of the screen and audio recording of each student was made for all sessions.

Overall students responded to the test positively, with many students reporting that they preferred the computer-based test to the paper-based form of NAPLAN.

Interviewers found that students were not disturbed or unsettled by the unusual progression of the test. When questioned carefully, students could perceive the shifts in difficulty between testlets but otherwise these shifts in difficulty were largely ignored by students. There appeared to be no negative consequences to the branching structure of the test and in so far as students were aware of the branching they largely regarded it positively.

Interviewers also found that students responded positively to computer-based delivery of test items. Features such as the on-screen timer or the on-screen calculator (Numeracy) were regarded favourably by students. However the presentation of stimulus texts in the Reading tests produced unfavourable reactions due to limitations of screen size and resolution.

A majority of students found the test more engaging than the paper version of NAPLAN. The primary factor in improved engagement was the computer-based delivery of the test.

Methodologically, this was a fact-finding and hypothesis-generating study. All observations and conclusions are necessarily limited by the small and non-random sample. Because of the video recording component of the study, only students who had returned parental permission slips could be interviewed and consequently bias may have been introduced – in particular, it proved substantially harder to include boys in the study than girls.

Caution should be employed when considering the range of behaviours observed. Each student was observed throughout the test by a single observer concentrating on how they interacted with each item. This is clearly not typical of normal test conditions nor is it the expected manner in which proposed online tests would be delivered.

Introduction

Background

The National Assessment and Surveys Online Program (NASOP) is a series of projects by the Australian Curriculum Assessment and Reporting Authority (ACARA) that investigate the online delivery of ACARA’s assessment programs. As part of NASOP, a trial of online reading and numeracy tests took place in August 2013.

The online tests were designed around a partially adaptive multistage test model. This model consists of a series of mini-tests (known as testlets or modules) of varying difficulty. Students follow different paths through the testlets, depending on their score in each testlet.

EAA was commissioned by ACARA to conduct cognitive interviews to investigate the impact of the multistage-branching design on students' test-taking behaviour and to gather information about students' interaction with the testlets at key branching points of the test.

This document provides combined data from both the Numeracy and Reading domains. This allows for a larger data set and also for comparisons between the two domains. More detailed discussions of each domain can be found in Part 2 (Reading) and Part 3 (Numeracy).

The NASOP multistage model

The model for the test piloted in this study is a multistage partially adaptive branching model. A test at a given year level consists of six testlets of items of known difficulty. Students were presented with three of these testlets (plus a fourth testlet of additional items).

The following diagram shows the branching and facility range of the testlets.

Figure 1 Pathways



The model has four distinct pathways:

Table 1 Difficulty labels for pathways

Difficulty	Sequence of testlets
High	ADF
Middle	ADE
Middle	ABE
Low	ABC

Each testlet covered a set range of difficulty. This allowed students to be presented with more items that were appropriate to their level of ability. The relative difficulties of each testlet are given in this table:

Table 2 Difficulty ranges for testlets

Testlet	Facility (high)	Facility (low)
A	95%	45%
B	95%	30%
D	80%	20%
E	80%	10%
C	high	70%
F	40%	low

Testlets A, B and D are the branching testlets. The student’s score in a branching testlet determines which testlet they proceed to next. ACARA requested that key items in the two branching testlets through which a student passed should be discussed in detail in the cognitive interviews.

A key aspect of the multistage model is that the difficulty of items does not steadily increase as is typical of paper-based NAPLAN. ACARA also requested that student responses to items which show marked changes in difficulty from previous items be monitored carefully in the cognitive interviews.

The difficulty ranges of testlets C and F go beyond that of current NAPLAN tests. Consequently ACARA also requested that the behaviour of students who reach these testlets be examined by the cognitive interviews.

Fixed-path model

While the intended model of the test involves adaptive branching, for the purpose of this trial half of the students were allocated tests with a predetermined path. This fixed-path model ensured that sufficient numbers of students undertook all of the available pathways during the trial.

Methodology

Protocol development

The priority aspects investigated in this study as requested by ACARA were:

1. the impact of the multistage-branching design on the students' test-taking behaviour
2. students' interaction with the testlets at key branching points of the test:
 - a. investigate key items in the two branching testlets which students pass through
 - b. investigate student response to items which show marked changes in difficulty
3. student performance on testlets C and F.

EAA's methodology for protocol development proceeded from priorities set by ACARA. Protocols were devised around the three interrelated themes of branching, computer-based test experience and engagement.

- **Theme: Branching**

The major innovation of the test design for this trial is the branching process and it is this process that forms the primary theme of the cognitive interview study.

- **Theme: Computer-based test experience**

The term 'Computer-based test experience' (CBTE) relates to those aspects of the test experience that are particular to a computer-based test or which are significantly affected by computer-based delivery of the test. This theme was the primary focus of the 2012 NASOP study.

- **Theme: Engagement**

Student engagement with the test may be affected by both the mode of the delivery and the branching model (e.g. through improved targeting of items to student ability). Student engagement was a main focus of this study.

This was done using four methods:

- Concurrent observation (CO): observing student behaviour during the test
- Concurrent think-aloud (CTA): asking students to 'think aloud' while answering an item
- Concurrent interview (CI): asking students questions about the item or item set they have just answered
- Retrospective interview (RI): asking students questions about items in an interview that is held after the test is complete.

The protocols were developed collaboratively by seven EAA staff members who had experience in cognitive interviewing as well as an understanding of either the NAPLAN Numeracy or NAPLAN Reading constructs.

The methodology for identifying items to be targeted during the interviews was based on the study priorities listed above. Based on findings from the NASOP 2012 Cognitive Interview Study (EAA, 2012), it was determined that a total of 12-16 items was the optimal number to target per year level. It allowed EAA to adequately sample the range of items presented within a timeframe that was manageable for the students.

Table 3 Targeting of items by method

Method	Number of items	Theme	Items		
			Numeracy	Reading	
Concurrent	Observation	All	All	All	
	Think-aloud	4 (approx)	CBTE, Engagement	2: items A & B /D 2: items F/C/E + informally throughout test	4: one item set F/C/E + informally throughout test
	Interview	4 (approx)	Branching, Engagement	Y3 & 5 2: final items of A & B/D 1: initial items of B/D 1: initial items of F/C/E Y7 & 9 2: selected items A & B/D 1: selected items F/C/E	2: final items of A & B/D 1: initial items of B/D 1: initial items of F/C/E + supplementary question for each complete item set all testlets
Retrospective	Interview	Numeracy 8 (approx) Reading 4 (approx)	CBTE, Engagement	Y3 & 5 4: selected items A & B/D 4: selected items F/C/E Y7 & 9 2: final items of A & B/D 2: initial items of B/D & F/C/E 3: mid-testlet items F/C/E	Any sequencing, multiple-response or short response items not already targeted in approximately these proportions: 2: selected items A & B/D 2: selected items F/C/E

The protocols for each domain formed the basis of training materials for interviewers. All interviewers engaged in at least one day's training in the relevant protocol and all the interviewers had played some role in protocol development.

Sampling approach

ACARA provided comprehensive details on the full proposed sample of schools for the main NASOP study. These details included the sector the school was from (Government, Catholic or Independent) and whether the area they served was metropolitan (i.e. in the given states' main urban population centre) or regional. Additionally each school was classified as either "High", "Middle" or "Low" based on general performance in NAPLAN to help ensure an adequate range of abilities was sampled.

From this list EAA identified a range of schools that would form a sub-sample for the cognitive interviews. Schools were selected from New South Wales and from Queensland.

The proposed sample included many more government schools than schools from other sectors so as to include a broad range of students from a variety of educational and socio-economic backgrounds.

Table 4 Proposed school sample by state and sector

State	Sector			Total
	Cath	Gov	Ind	
NSW	1	21	4	26
QLD	2	8	1	11
Total	3	29	5	37

The proposed sample was split in an approximate 2:1 ratio of metropolitan and regional schools.

Table 5 Proposed school sample by state and location

State	Location		
	Metro	Regional	Total
NSW	17	9	26
QLD	8	3	11
Total	25	12	37

Schools were then further subdivided into various areas both for ease of allocating interviewers and for ensuring a variety of locations. For example, it was important that schools from the general category of “Metropolitan NSW” were drawn from a variety of Sydney areas including the outer western suburbs, the south-western suburbs as well as suburbs closer to the CBD.

Table 6 Proposed school sample by area

State	Location	Area	Total
NSW	Metro	Sydney	10
		Campbelltown	4
		Western Sydney	3
	Regional	Blue Mountains	2
		Central Coast	1
		Newcastle	3
		Orange	1
		Wollongong	2
QLD	Metro	Brisbane	4
		Gold Coast	4
	Regional	Rockhampton	3
Total			37

The total number of schools was selected on the basis that:

1. at each school approximately four students would be interviewed (2 Numeracy and 2 Reading)
2. approximately double the number of schools needed should be identified to ensure sufficient numbers of schools would be visited.

The schools were then approached to provide permission slips for students to take part in the cognitive interviews. Because the interviews were to be recorded, only students who returned permission slips could be included in the cognitive interview study.

Final sample

Schools

In total, sixteen schools participated in the cognitive interviews. Three factors contributed to a school’s not being able to participate.

1. Non-return of permission slips. Several schools did not return any permission slips and consequently could not be included in the cognitive interviews.
2. Scheduling issues. Cognitive interviews were intended to take place on the same day as the main study and in some cases it was not possible to send interviewers to the school on the day designated by the main study contractor.
3. Some schools in the proposed sample opted out of the main study completely.

To mitigate these issues, a number of the schools that had opted out completely were re-contacted to see if they would be willing to participate in the cognitive interviews study only. In addition, one school (an independent Christian school in Sydney) which had not been listed in the proposed sample was included in the cognitive interviews.

The proportion of schools from the proposed sample that actually took part in the cognitive interviews was not consistent across sectors. Only a third of the government schools in the proposed sample were able to take part in the cognitive interviews. The attrition rate amongst independent schools was much smaller. Consequently the relative proportion of independent schools in the final sample was larger.

Table 7 Final school sample by state and sector

State	Sector			Total
	Cath	Gov	Ind	
NSW		9	4	13
QLD	1	2		3
Total	1	11	4	16

The proportion of NSW and Queensland schools was also affected, with a greater proportion of NSW schools included in the interviews. However, the proportion of metropolitan and regional schools was largely unaffected.

Table 8 Final school sample by state and location

State	Location		Total
	Metro	Regional	
NSW	9	4	13
QLD	2	1	3
Total	11	5	16

Despite many schools not being able to participate in the study, the final sample still included a broad range of geographical areas.

Table 9 Geographical distribution of schools in final sample

State	Geographical distribution		Total
	Location	Area	
NSW	Metro	Sydney	6
		Campbelltown	1
		Western Sydney	2
	Regional	Blue Mountains	1
		Newcastle	1
		Orange	1
Wollongong		1	
QLD	Metro	Brisbane	2
	Regional	Rockhampton	1
Total			16

The only areas not covered in the final sample that were listed in the proposed sample were the Gold Coast in Queensland and the Central Coast of NSW.

Within this sample of schools, three schools were single-sex schools. In all cases the single-sex schools were girls' schools.

Students

Across both domains, a total of 85 students were interviewed. Roughly equal numbers of students were interviewed in each of the four NAPLAN year levels (Year 3, Year 5, Year 7 and Year 9). Roughly equal numbers of students were interviewed for Numeracy and Reading. One student was interviewed for both Numeracy and Reading.

Table 10 Numbers of students interviewed by domain

Domain	Year 3	Year 5	Year 7	Year 9	Total
Numeracy	10	10	10	14	44
Reading	10	11	11	9	41
Total	20	21	21	23	85

Students were drawn primarily from government schools in metropolitan areas but significant numbers were also drawn from regional locations and from independent schools.

Table 11 Students interviewed by location and sector

Domain	Location	Sector			
		Cath	Gov	Ind	Total
Numeracy	Metro		29	6	35
	Regional		5	4	9
Reading	Metro		20	8	28
	Regional	2	7	4	13
Total		2	61	22	85

About 10% of students in both domains were interviewed in Queensland. A greater proportion of students in Reading were from regional schools than in Numeracy but overall a quarter of all students were from regional schools.

Table 12 Percentage of students in each domain by state and location

Domain	State	Location		
		Metro	Regional	Total
Numeracy	NSW	68%	20%	89%
	QLD	11%	0%	11%
Numeracy Total		80%	20%	100%
Reading	NSW	59%	27%	85%
	QLD	10%	5%	15%
Reading Total		68%	32%	100%
Total both domains		74%	26%	100%

In the 2012 NASOP Numeracy cognitive interviews study, approximately twice as many girls as boys were interviewed. In the 2013 study, more girls than boys were interviewed but the proportional difference was less. As in 2012, the main factors contributing to the gender imbalance were the return of permission notes, selection of students by the school, and the presence of girls' schools in the sample without a balancing presence of boys' schools.

Table 13 Percentage of students by domain and gender

Domain	Gender		
	Female	Male	Total
Numeracy	61%	39%	100%
Reading	66%	34%	100%
Total both domains	64%	36%	100%

Pathways

Because of the importance of the tailored nature of the test, a sustained effort was made to interview students with a range of abilities. This was done firstly by including a range of schools in the proposed sample and secondly by contacting schools and asking them to identify students within given ability ranges. However the final selection of students was dependent on the return of permission notes and the availability of students on a given day. As can be seen in this table, the net effect of these issues was a tendency for more able students to be interviewed. This issue is less pronounced in the Numeracy domain despite the common sampling approaches for Numeracy and Reading.

Table 14 Percentage of students who followed a given pathway

Path		Domain		
		Numeracy	Reading	Total
Low	ABC	32%	24%	28%
Middle	ABE	18%	10%	14%
	ADE	20%	15%	18%
High	ADF	30%	51%	40%
Total		100%	100%	100%

Pathways were also allocated as either “fixed” or “free”. With fixed pathways, students were allocated a predetermined sequence of testlets. With free pathways, the sequence of testlets was determined by the student’s score up to the branching point. It is notable that the ADF pathway was the most frequently followed pathway in Reading when a free pathway was allocated.

Table 15 Free and fixed pathways by domain

Path		Domain		
		Numeracy	Reading	Total
Fixed	ABC	7	8	15
	ABE	6	2	8
	ADE	2	2	4
	ADF	5	8	13
Fixed Total		20	20	40
Free	ABC	7	2	9
	ABE	2	2	4
	ADE	7	4	11
	ADF	8	13	21
Free Total		24	21	45
Total		44	41	85

Platforms and computer use

Students interviewed attempted the test on a range of different platforms. However, it should be noted that to accommodate the interviews schools often allocated computers that were available in quiet locations suitable for one-on-one observation. Consequently the platform used for the interview at a given school may not have been the same platform as was used by students in the main study at the same school.

Table 16 Platforms observed by domain

Platform		Domain		
		Numeracy	Reading	Total
Desktop	Mac	4	4	8
	PC	12	15	27
Big Notebook	Mac	4	6	10
	PC	16	12	28
Small Notebook	PC	6	2	8
Tablet	iOS	2	2	4
Total		44	41	85

It had been hoped that more students would have been observed using iPads, but at two schools technical issues unrelated to the platform prevented observation of students attempting the test on iPads.

At the start of the test students were asked three questions about their familiarity with computers. The following tables show the proportion of responses for each domain.

Table 17 Computer use in school

Q1. How often do you use computers or tablets at school?					
Domain	most days	some days	only for some classes	not often	Total
Numeracy	34%	23%	27%	16%	100%
Reading	39%	49%	12%	0%	100%
Total both domains	36%	35%	20%	8%	100%

Table 18 Computer use at home

Q2. How often do you use computers or tablets at home?					
Domain	every day	sometimes	not often	never	Total
Numeracy	48%	41%	9%	2%	100%
Reading	49%	41%	10%	0%	100%
Total both domains	48%	41%	9%	1%	100%

Table 19 Touch screen phone use

Q3. How often do you use a mobile phone with a touch screen?					
Domain	every day	sometimes	not often	never	Total
Numeracy	48%	32%	14%	7%	100%
Reading	44%	24%	17%	15%	100%
Total both domains	46%	28%	15%	11%	100%

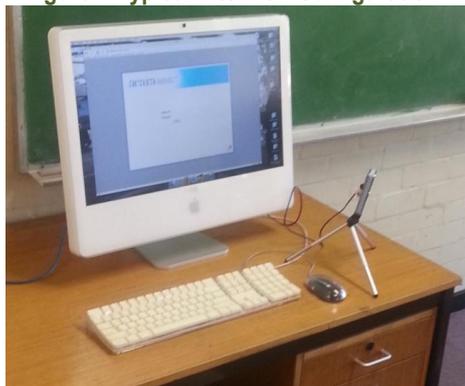
As can be seen from the tables the majority of the students were regular users of IT in school and at home. When both home and school use were considered together only one student was in the ‘not often’ category for both locations.

Table 20 Comparison of first and second computer use questions

		Q1. How often do you use computers or tablets at school?			
		most days	some days	only for some classes	not often
Q2. How often do you use computers or tablets at home?	every day	21	10	7	3
	sometimes	10	15	7	3
	not often	0	4	3	1
	never	0	1	0	0

Findings: Overview

Figure 2 Typical interview configuration



Theme: Branching

The major innovation of this trial was the branching process and this process formed the primary theme of the cognitive interview study.

To see if students were aware of the branching, the first question in the retrospective interview asked them for any differences they noticed between the test they had just completed and NAPLAN.

Q1: One difference between the test you took today and other tests you usually take (for example NAPLAN) is that you took it on the computer rather than on paper. Did the test you took today differ from the NAPLAN tests you have taken in the past in any other ways?

Most students responded that test was in some way different from NAPLAN.

Table 21 Did the test differ from NAPLAN?

Did it differ?	Domain		
	Numeracy	Reading	Total
Yes	35	29	64
No	9	11	20
Other	0	1	1
Total	44	41	85

Although this question overtly excluded difference due to the mode of delivery, several students gave responses focused on the difference between computer-based and paper-based testing. Those responses are covered in more detail below. Other responses pertained to the difficulty of the test. It is notable that no students identified unprompted that the order of difficulty of the items was in any way unusual – particularly given that the order of difficulty was one of the most significant differences to a normal NAPLAN test other than the mode of delivery.

Question 3 of the retrospective interview focused more specifically on the issue of test difficulty.

Q3: How difficult did you find the test? Was it more difficult than NAPLAN, less difficult or about the same?

Table 22 Was the test easier or harder than NAPLAN?

How difficult?	Domain		
	Numeracy	Reading	Total
Easier than NAPLAN	12	13	25
About the same	15	24	39
Harder than NAPLAN	12	2	14
Other	5	2	7
Total	44	41	85

In general, students found the test easier or about the same as NAPLAN. However a number of students in Numeracy found the test harder. Comparing these responses against the pathways students followed shows some relation with the difficulty of the pathway. For example, all of the students in Reading who said that the test was harder than NAPLAN had followed the hardest pathway (ADF).

Table 23 Comparative perceived test difficulty by pathway and domain (percentage of row)

Domain	How difficult?	Path				Total
		ABC	ABE	ADE	ADF	
Numeracy	Easier than NAPLAN	42%	17%	17%	25%	100%
	About the same	33%	13%	27%	27%	100%
	Harder than NAPLAN	8%	25%	17%	50%	100%
	Other	60%	20%	20%	0%	100%
Numeracy Total		32%	18%	20%	30%	100%
Reading	Easier than NAPLAN	46%	8%	23%	23%	100%
	About the same	17%	13%	8%	63%	100%
	Harder than NAPLAN	0%	0%	0%	100%	100%
	Other	0%	0%	50%	50%	100%
Reading Total		24%	10%	15%	51%	100%
Total		28%	14%	18%	40%	100%

Even so, many of the students in both domains who followed the ADF pathway did not report the test as being harder than NAPLAN. Similarly, while students in either domain who followed the ABC pathway were more likely to say it was easier than NAPLAN, many said it was about the same and one student reported that it was harder.

Table 24 Comparative perceived test difficulty by pathway (percentage of column)

How difficult?	Path				Total
	ABC	ABE	ADE	ADF	
Easier than NAPLAN	46%	25%	33%	18%	29%
About the same	38%	42%	40%	56%	46%
Harder than NAPLAN	4%	25%	13%	24%	16%
Other	13%	8%	13%	3%	8%
Total	100%	100%	100%	100%	100%

Question 4 probed the issue of the progression of difficulty further.

Q4: Overall did the test seem to get harder as you worked through it, or easier or was it more mixed?

In general students thought the difficulty progression was more mixed than any general progression from easy to hard or vice versa.

Table 25 Student perception of difficulty progression by domain

Difficulty change	Domain		
	Numeracy	Reading	Total
Got easier	3	4	7
Got harder	14	9	23
Mixed	27	24	51
Other		4	4
Total	44	41	85

Even students on the ADF pathway were about evenly split on whether the test got harder or was better described as being mixed.

Table 26 Perceived difficulty progression by pathway (percentage of column)

Difficulty change	Path				
	ABC	ABE	ADE	ADF	Total
Got easier	21%	8%	7%	0%	8%
Got harder	8%	8%	27%	47%	27%
Mixed	67%	75%	67%	47%	60%
Other	4%	8%	0%	6%	5%
Total	100%	100%	100%	100%	100%

Ignoring ‘other’ responses, there were some sharper distinctions in Reading, with ADF pathway students accounting for all of the “Got harder” responses in Reading but only 50% of the “Got harder” responses in Numeracy.

Table 27 Perceived difficulty progression by domain and pathway (percentage of row)

Domain	Difficulty change	Path				
		ABC	ABE	ADE	ADF	Total
Numeracy	Got easier	67%	33%	0%	0%	100%
	Got harder	14%	7%	29%	50%	100%
	Mixed	37%	22%	19%	22%	100%
Numeracy Total		32%	18%	20%	30%	100%
Reading	Got easier	75%	0%	25%	0%	100%
	Got harder	0%	0%	0%	100%	100%
	Mixed	25%	13%	21%	42%	100%
Reading Total		24%	8%	16%	51%	100%
Total both domains		28%	14%	19%	40%	100%

The sequence of questioning was designed so that students would receive an increasing amount of prompting on the issue of branching. By question 5, interviewers had explained to the students that the test was intended to be in three parts of varying difficulty. Students were asked whether they had noticed the changes.

Q5: Did you actually notice the test changing in any way when you moved from part 1 to part 2? Did you notice the test changing in any way when you went from part 2 to part 3?

Not all students responded to this question but the two tables summarise the responses of the students who did.

Table 28 Students noticing change from first to second testlet by domain

Notice change from 1 to 2?	Domain		
	Numeracy	Reading	Total
Ambivalent	18	3	21
No	12	18	30
Yes	12	18	30
Total	42	39	81

Table 29 Students noticing change from second to third testlet by domain

Notice change from 2 to 3?	Domain		
	Numeracy	Reading	Total
Ambivalent	16	3	19
No	12	11	23
Yes	14	25	39
Total	42	39	81

Many students gave ambivalent responses, although this was less pronounced in Reading. In general their responses were consistent between both parts of the question, although some students who didn't notice the part 1 to 2 change did notice the part 2 to 3 change.

Table 30 Comparison of students noticing changes between testlets

		Notice change from 2 to 3?			
		Ambivalent	No	Yes	Total
Notice change from 1 to 2?	Ambivalent	17	1	3	21
	No	1	16	13	30
	Yes	1	6	23	30
	Total	19	23	39	81

When looked at by pathway, the B/C change and the D/F change appeared to be more noticeable to students than the A/B or A/D change.

Table 31 Students noticing change from first to second testlet by pathway

Notice change from 1 to 2?	Path				
	ABC	ABE	ADE	ADF	Total
Ambivalent	6	3	5	7	21
No	10	6	4	10	30
Yes	7	3	5	15	30
Total	23	12	14	32	81

Table 32 Students noticing change from second to third testlet by pathway

Notice change from 2 to 3?	Path				
	ABC	ABE	ADE	ADF	Total
Ambivalent	5	3	4	7	19
No	7	4	5	7	23
Yes	11	5	5	18	39
Total	23	12	14	32	81

Branching: summary

Many students reported they noticed the different ways in which item and stimulus difficulty progressed through different pathways. However, interviewers felt that students only noticed this in retrospect. When students were asked initially about any differences they may have noticed between NAPLAN and the online tailored tests, no students identified difficulty progression as a point of difference.

Student responses to direct questions regarding the difficulty of items around branching points suggest that they were aware of significant shifts in difficulty of items and stimulus texts around branching points. It is notable that students generally did not seem to regard this as a significant issue or something that distracted them from their test-taking experience.

Theme: Computer-based test experience

The term ‘Computer-based test experience’ (CBTE) relates to those aspects of the test experience that are particular to a computer-based test or which are significantly affected by computer-based delivery of the test. This theme was explored in a number of ways during the concurrent and retrospective phases of the interview. In general, unsolicited comments from students were more likely to pertain to technical issues or general computer-based experience issues than to the other two themes in the study.

When asked about the differences between the test they had just sat and NAPLAN, 21 students gave responses pertaining to the ease or difficulty of entering their responses. These responses could be roughly divided into the four categories below.

Table 33 Difference from NAPLAN: Responding to items

Entering data	Domain		
	Numeracy	Reading	Total
Easier to correct	2	0	2
Easier to respond	11	6	17
Entry mistakes more likely	1	0	1
Harder to type	0	1	1
Total	14	7	21

Other CBTE issues students raised when asked about the differences between the test they had just sat and NAPLAN related to some of the navigation and timing features of the test.

Table 34 Difference from NAPLAN: Assorted navigation/timing issues

Navigation	Domain		
	Numeracy	Reading	Total
Can't go back/Liked back button	2	3	5
Isolated items (positive)	1	6	7
Isolated items (negative)	0	2	2
Timer	2	2	4
Total	5	13	18

The two students in Reading who made a point of saying that they liked the back button in the Reading test were both in Year 9 and had sat the Numeracy test the day before. In Year 7 and Year 9 Numeracy, no back button was available because calculator and non-calculator items were interspersed. The two students in Numeracy who were unhappy that they couldn't ‘go back’ were also in Year 9. Aside from those students the restricted navigation was not a major issue among students.

‘Isolated items’ relates to responses from students (primarily in Reading) raising the issue that in the computer-based tests each item is presented in isolation from the other items. This was largely regarded as a positive point by those who raised it, except for two students who preferred to see all the items in an item set together.

Those students who mentioned the timer generally regarded it as positive feature. As well as the four students who raised it in their response to the first question of the retrospective phase, other students raised it in unsolicited comments at other stages of the interview.

A key issue raised by interviewers and students was the variability of presentation of the test. With students having to attempt the test on computers with different-sized screens and of varying quality, the test-taking experience could be analogous to a paper-based test that was presented to students on different-sized paper and with varying quality of print. Students raised presentation issues in various stages of the interview. The following table summarises the numbers of students who raised issues at varying points of the interview.

Table 35 Presentation issues by domain

Presentation issues	Domain		
	Numeracy	Reading	Total
Easier to read	13	3	16
Mixed		2	2
Harder to read	1	8	9
Total	14	13	27

There was a marked difference between Numeracy and Reading on this issue. Numeracy comments often related to larger graphics or diagrams or the occasional use of colour. Reading comments typically related to the display of stimulus texts. Surprisingly the frequency of comments does not seem to relate to the size of screen used.

Table 36 Presentation issues by platform

Presentation issues	Platform				Total
	Desktop	Big Notebook	Small Notebook	Tablet (iPad)	
Easier to read	3	9	3	1	16
Mixed		2			2
Harder to read	4	4	1		9
Total	7	15	4	1	27

However, it also needs to be noted that not all students who encountered presentation issues raised them when interviewed. In one case a student attempting the Numeracy test on a computer with unusual screen resolution settings (which couldn't be changed due to permission restriction) identified a division symbol (\div) as an addition symbol (+). Interviewers in Reading reported a variety of issues with students delayed by having to zoom text or by reading more slowly. Interviewers in Reading raised most issues with the test when presented on smaller notebooks. The presentation on the small number of iPads observed was regarded positively, in part due to screen quality and to the ease with which text could be magnified when needed.

Technical issues

This study was not intended to be an evaluation of the test-delivery platform, but when considering the computer-based test experience, technical problems have to be considered.

The primary technical issues encountered by interviewers and students related to logging into the system. In some cases this was just an issue of finding the right way to turn off pop-up blockers. In other cases logging in was delayed by network issues either at the school or at the server side of the system. In general students reacted to such issues stoically, presumably because technical issues are a not-uncommon feature of using computers in a school environment.

On occasions when students had to close their browser because of technical problems, the system typically returned students to the last item they had attempted when they logged back in. This feature helped mitigate the effect of technical glitches on the student's experience of the test.

More generally, variations in hardware and software meant that student experiences of the test were highly variable. In the case of the cognitive interviews, this issue was exacerbated by students being allocated computers that were 'spare' machines – often older machines with less than up-to-date software.

The impact of these issues on the Numeracy test was less than on Reading. On a per-item basis, most computers could more than adequately display the content of a test question reliably. However in Reading, the demands of displaying a whole stimulus text consistently were a greater challenge.

During the interview, students also raised the question of technical problems more generally. Two students raised the possibility of technical problems in response to the first question of the retrospective phase.

CBTE: Summary

Students found many positive things in taking the test on computer. Large numbers of students reported preferring the computer-based test on the simple grounds that it was more interesting because it was on a computer. Features such as the timer and the on-screen calculator received favourable responses. Similarly, graphics were often regarded as being clearer and more colourful.

The issue of the stimulus text in Reading is a major one for the computer-based test experience. Interviewers were concerned that students' experience of the test varied significantly depending on the size and quality of their screen. Interviewers themselves found the variation in readability of texts frustrating at times. Thought needs to be given to ways in which students can receive a more consistent experience with the stimulus texts in Reading.

Some lesser presentation issues occurred in Numeracy but these were due to incorrect settings on specific machines.

Theme: Engagement

Both the mode of delivery and the improved targeting of students may improve student engagement with the test. Interviewers monitored engagement during the concurrent observational phase and also asked students questions about their level of engagement during the retrospective interview phase.

Question 2 in the retrospective interview phase asked students to compare their level of interest between the NASOP test and NAPLAN.

Q2: How interesting or engaging did you find the test? Was it more engaging than NAPLAN, less engaging or about the same?

Table 37 Reported engagement by domain

How engaging?	Domain		
	Numeracy	Reading	Total
Less than NAPLAN	2	2	4
More than NAPLAN	24	25	49
About the same	18	13	31
Other		1	1
Total	44	41	85

Of those who said that it was more interesting than NAPLAN, 24 students (9 in Numeracy and 15 in Reading) cited reasons related to simple computer delivery as being the primary reason why it was more engaging. Other students cited content-specific reasons such as a specific item or a more interesting text. Others cited reasons covered in the CBTE section above, such as the items being presented one at a time.

No student cited issues that were clearly related to either the targeting or the branching aspects of the test. However this does not mean that the improved targeting of the tailored test did not generate an overall improvement in the level of engagement.

Students were asked in Question 7 which of the three parts of the test (i.e. testlets) they enjoyed the most.

Q7: Which of the three parts did you enjoy the most?

Many students had no particular preference and final testlets were not particularly more popular.

Table 38 Preferred testlet by domain

Part enjoyed most	Domain		
	Numeracy	Reading	Total
1. A	13	7	20
2. B/D	16	6	22
3. F/E/C	9	9	18
No preference	6	19	25
Total	44	41	85

Examining the same data but by specific testlet reveals no particular pattern. Differences in preference are explained primarily by varying numbers of students who took a given testlet (e.g. all students in a given year level took testlet A and more students in Reading took testlet F than in Numeracy).

Table 39 Preferred testlet (percentage of column)

Preferred testlet	Domain		
	Numeracy	Reading	Total
A	34%	32%	33%
B	24%	9%	18%
D	18%	18%	18%
C	13%	9%	12%
E	5%	9%	7%
F	5%	23%	12%

Total	100%	100%	100%
--------------	-------------	-------------	-------------

Of particular interest was whether testlets C and F had any effect on engagement. Consequently students on ABC and ADF pathways were asked an additional question.

Q8: For testlet C: Did the easiness of the last part make a difference to how you liked the test/this part of the test?

Q8: For testlet F: Did the difficulty of the last part make a difference to how you liked the test/this part of the test?

When asked this specific question students gave some indication that the targeting of the final testlet may have had some effect.

Table 40 Effect of difficulty of third testlet

Make a difference?	Domain		
	Numeracy	Reading	Total
Ambivalent	2	2	4
No	9	4	13
Yes	14	24	38
Total	25	30	55

Of those students who responded that the difficulty of the final testlet had made a difference, about two-thirds saw it as having a positive effect.

Table 41 Effect of difficulty of third testlet on preference by domain

Positive or Negative	Domain		
	Numeracy	Reading	Total
Negative	5	5	10
Positive	9	19	28
Total	14	24	38

This positive effect was consistent across pathways.

Table 42 Effect of difficulty of third testlet on preference by pathway

Positive or Negative	Path		
	ABC	ADF	Total
Negative	4	6	10
Positive	13	15	28
Total	17	21	38

Engagement: Summary

The primary improvement in engagement appears to derive simply from the test being delivered on computer. Students were less aware, in general, of branching and targeting effects. In so far as they were aware, the effect on engagement appeared to be positive.

Findings: Domain-specific overviews

This section covers general findings for each domain. Parts 2 and 3 of the report cover each domain in depth.

Numeracy-specific issues

Two issues were identified in the 2012 NASOP Cognitive Interview study as being of particular interest. Both issues fall under the general theme of Computer-based test experience.

Interaction with graphics

“Interaction with graphics” refers to how students respond to graphics and diagrams presented on-screen.

Interaction with graphics was observed in all year levels, but only for questions involving map, direction, measure of distance and certain graphs. Some students moved the cursor over the text as they read. However most of the time, they just read the screen and picked the answer without overt interaction with graphical elements.

In general the range of responses was similar to the previous study. The presentation of diagrams on screen did not present a significant obstacle to students.

The scrap paper problem

The “scrap paper problem” relates to students not using paper to help work out problems when a question is presented on-screen.

Scrap paper was used to solve various types of questions. Student preference using scrap paper was evenly distributed among students of all ability levels. Some high-ability students (especially Year 3 and 5) were able to work out many questions mentally, whereas others (usually medium- to low-ability students) would quickly pick an answer by skimming questions superficially without further investigation. Lower ability students tended to use scrap paper to work out simple calculations. Students in Year 3 and 5 used less scrap paper than students in Year 7 and 9.

The issue of using paper for rough working when using a computer remains unclear. As this study did not compare students attempting tests in both modes of delivery (computer-based and paper-based), it is unclear whether computer delivery is reducing students’ willingness to do rough working on paper.

Reading-specific issues

Several issues that apply primarily to Reading were identified by interviewers during the study.

Presentation of items

Several students commented on liking the presentation of one question at a time. They felt it allowed them to focus on the question without the distraction of glancing at other questions and being seduced into thinking about them. One Year 7 student commented that this test probably took them longer to complete as they actually considered each question in turn.

One issue of concern is the presentation of stimulus texts on smaller or poorer quality screens. The type for the texts was frequently fuzzy and in some cases very small. The fuzzy type suffered in contrast with the sharp, easily-read type used for the questions and options. Some students found it hard to separate appearance from content in gauging difficulty of a text. One text (*Behind the waterfall*) did not display the final two lines which were covered by an illustration.

Platform and stimulus interaction

Reading interviewers were concerned that variations in platforms appeared to be affecting test performance. This issue follows on from the issues with presentation of stimulus texts mentioned above.

Hardware varied greatly, from large-screen (55 cm) fixed Macs with a mouse, to tiny 25 cm laptops with a touch pad for navigation, to iPads. How the student encountered and interacted with the test very much depended on the hardware available. Students using iPads were able to enlarge the screen quickly and easily by spreading and pinching, while others had to use less elegant keyboard controls.

If the computer used had a small screen it sometimes discouraged students from looking back at the text to find an answer, as it was too time-consuming to repeatedly enlarge and shrink the screen. When the view was enlarged the navigation button was not always visible so the screen had to be decreased in order to proceed to the next item.

Each new item took the text back to the starting position. Students sometimes leaned in and squinted closely at the screen in order to read the stimulus, rather than fuss with enlarging the screen. It appeared to affect their willingness to go back to the text to locate quotations or find or check answers.

Variety of item types

Some able students liked the variety of different question formats; less able students tended to like the support of the multiple choice options ('I have a one-in-four chance of getting it right'; 'I don't have to think of an answer; you get ideas for the answer'). The longer constructed-response questions were more likely to cause confusion and uncertainty.

Few students seemed to give quality responses to constructed-response questions. They often quoted directly from the text, sometimes even replicating wording from the question stem as their answer. The prevailing idea seemed to be to write as little as possible.

One student recommended that for a sequencing question it would make more sense to have the numbers available to drag and drop into position rather than typing them in. That student took the test on an iPad where the keyboard had to be brought up separately for each number.

Timing

It was noticeable that the Reading interviews took significantly longer than the Numeracy interviews. While there was some difference in the two protocols which would lead to Reading interviews taking longer, there was a general feeling among interviewers that the Reading test was simply harder to complete in the given time.

Interaction with the stimulus texts may have exacerbated this issue. Four students timed out while taking the test. Being a slow reader seemed to be a problem that was compounded by the difficulty of reading the texts on-screen and the difficulty of finding the place in the text to locate answers. Poor readers may have been disadvantaged by the online presentation of the test.

Conclusions and recommendations

Branching/Tailored test design

No cognitive or engagement problems with the tailored test design were identified in this study. Students largely ignored shifts in item difficulty and treated the test as just another test.

When the branching and targeting was drawn to students' attention, they largely regarded it positively.

Computer-based test experience

Students responded positively to computer-based delivery. Features such as the timer, on-screen calculator and presentation of a single item at a time were regarded favourably.

Restriction on navigating back to previous items was largely not regarded as an issue by students. However teachers may regard it less favourably and students may need some preparation to adjust to tests where they cannot review their answers at the end.

Engagement

New item types, clearer and colourful stimulus all appear to aid engagement. Exploiting the advantages of the technology is likely to be appreciated by students.

Numeracy

Students are increasingly solving mathematical problems in on-line environments. Students need to be aware that cognitive aids such as working-out paper remain important regardless of the mode of delivery.

Reading

The delivery of items via computer was unproblematic in this study but the presentation of stimulus texts produced unfavourable reactions due to limitations of screen size and resolution. The testing software could be set up to allow students to "full-screen" the stimulus text with a click, and then to quickly swap back to the split-screen view displaying both the text and the item. It is also advisable that a minimum text resolution and inter-line spacing be maintained for all texts.

Further study

Cognitive interviews not directly tied to a main trial could provide a better variety of students in a study.

There would be great value in conducting whole-class observations during a main trial. Student behaviour in a whole-class context may differ significantly from the student behaviour that was observed in this study. Some students raised concerns about other students cheating and it would be of some value to see how students interact with a test in a more natural environment.

Appendix: Cognitive Interview Protocol

Generic Structure

The two domain specific protocols are based on a single generic structure. This structure organises the cognitive interviews in a series of phases. This table gives an overview of the phases.

Phase 1: Set-up	Set-up phase		Outline the process to the student. Log them into the system. Prepare recording.
	[Numeracy only] Demonstrate 'Think Aloud' technique		Use the Think Aloud sample item (hard copy) to demonstrate the think aloud technique.
Phase 2: Test - 1 hour (approx)	Brief students on ongoing comments		Explain to the student that we would like them to make regular comments on the items and stimulus texts (reading) as they encounter them.
	Prompt commenting with first group of items		For the first item set (reading) or first four items (numeracy) prompt the student to give feedback on the issues of: <ul style="list-style-type: none"> • difficulty • interest level • any items they find notable.
	During the test	Observe	As students take the test observe relevant and interesting behaviour. In particular note the target behaviours for the domain.
		Monitor comments	Note down any comments students make about items as they proceed through the test.
		Concurrent think aloud [Reading: testlet F/C/E]	When students reach items flagged as CTA stop them before they attempt the item and take them through the think-aloud protocol. Make notes on what they say.
		Concurrent interview	When students reach items flagged as CI stop them once they have finished the item. Take the student through the concurrent interview protocol for the item.
		[Reading only] Item set	After the student has completed the final item of an item set ask them about the overall difficulty and interest level of the item set.
		[Reading only] End of testlet B/D; Demonstrate 'Think Aloud' technique	Use the Think Aloud sample item (hard copy) to demonstrate the think aloud technique.
Finishing the test		Ensure students have finalised the test properly and finish up. Allow them to take a short break.	
Phase 3: Post Test - 35 minutes (approx)	Log back in		Log students back into the system so that they can review the items
	Retrospective interview:	Student identified items	During phase 2 the student should have identified some items of particular note. Identify 2 of these and take the student to them. Also take students to targeted items. These will include: <ul style="list-style-type: none"> • items of particular note for online delivery not already targeted • items before and after branching points in the test not already targeted in Phase 2 (Y7 & Y9 Numeracy).
		Targeted Items	
		Whole test	Debrief student about the experience with the test and how it compared with typical pen-and-paper NAPLAN test. Was the test more or less engaging? Explain the branching process to the student. Ask them if they noticed the shifts in difficulty.
Thank the student for their time and contribution			

Phase 1: Setup

Initial set up

Liaise with the Trial Invigilator and a representative of the school regarding the students you will be interviewing and the times. Ensure you have a quiet area to conduct Session 1 and Session 2 that has a reliable computer with internet access.

Clarify any housekeeping and timing issues regarding break times, toilet visits etc.

Ask a knowledgeable teacher if there are any special steps when logging in for students. Confirm also what the preferred web-browser is – there is likely to be one that is more regularly updated.

Ensure you have with you:

1. This document
2. Student log-on guide
3. The printed think aloud example item
4. The Phase 2 year level specific observation sheet for the correct domain
5. The Phase 3 Questionnaire
6. The Phase 3 year level specific interview sheet for the correct domain
7. Video camera, tripod and cables

Once you are in the room set up the camera and arrange the room so that it is suitable for the interview.

Student set up

Once the student has arrived at the room where the interviews will take place introduce yourself to the student.

Hello, I am _____.

I am visiting your school to see how well a [numeracy]/ [reading] test works. I need your help to find out how we can improve this test. I am not here to see how well you do in the test but we do need you to try your best.

Show the student the video camera and say:

I am going to use this video camera to record what appears on the screen. I will place it so that it will not video your face.

Put the camera in a position where it has a good view of the screen but not the student. Switch on the camera so the student can see what it will be recording.

The camera also has a microphone that will be recording our conversation. The purpose of this is to capture in detail what we say; it would be difficult for me to write down everything and listen carefully to what you say at the same time...

We will only use the recording to help us research the test you are doing. The recording will not be used in public. Your name won't be on the recording.

I will tell you when the camera has started recording.

Explain about the sessions.

There are two parts to this interview. The first part is the test and that will take about an hour.

When we have finished the test you can have a short break. After that I will ask you some questions about the test. There are no right or wrong answers to these questions because we want to hear your own opinion. This will take about 30 to 40 minutes.

Ensure the student has a pen or pencil and some writing paper. Ask them to write their name on the paper.

Log in process

On your observation session sheet note the student's name, gender, log-in code and password. Also note:

- FIXED PATH MODEL: Note the path the student has been assigned to
- BRANCHING PATH MODEL: Note that the student is on the branching model.

Give the student their log-in and password. Ensure the computer is showing the Assessment Master log-in screen.



Get the student to log themselves in.

Demographic details screen

The students will be first presented with a screen asking for their demographic details. Ask them to fill this in and give assistance if they need it.

Questions about computer use are also included. As the student answers them also record their answers on your observation session sheet.

Computer usage						
1. How often do you use computers or tablets at school?						
most days		some days		only for some classes		not often
2. How often do you use computers or tablets at home?						
every day		sometimes		not often		never
3. How often do you use a mobile phone with a touch screen?						
every day		sometimes		not often		never

Phase 2: Reading Tests

Introduce the test

Explain how you will proceed.

Firstly explain that you would like them to comment throughout the test:

In a moment I would like you to start the test. The camera will record what happens on the screen. I will be making notes about how the test is working for you.

As you attempt each question please feel free to say anything about the question you are working on. First, we are really interested in when you notice a change in difficulty in the texts or questions, so when that happens let us know.

Just as much, we're interested in when you get bored, or when you find something in the test you really like, so be sure to tell us when that happens.

Finally, if there's anything that you have trouble with in the process of taking the test that makes it hard to focus on just reading the texts and doing the questions, that's important to us too.

Then explain that you will also be stopping them from time to time to ask them questions:

From time to time I will also ask you to stop for a moment. I may then ask you some questions about what you are doing or thinking. This will never be because I think you are doing something wrong. I will only ask questions because I want to learn something from you. Anything you can tell me will be helpful.

Unfortunately I cannot help you with any of the questions or explain what you have to do. If you are stuck, please say so and if you really cannot answer a question move on to the next one.

Ask the student if they have any questions about the process and answer them. Then say:

I am now starting the camera recording.

Start the camera recording and tell the student they may begin the test.

Start the test

For the first item set prompt the student to tell you about:

- the difficulty of the stimulus text: Was that text easy to understand?
- the interest level of the stimulus text: Was that text interesting?
- the difficulty of the question: Did you find that question easy or difficult?
- the interest level of the question: Did you find that question interesting?

Note down their answers and any other notable comments. Encourage the students to make similar comments at any point during the test.

Continue the test

Please refer to the year level specific observation sheet for guidance on specific items.

The sheet indicates items that:

- you will be observing only (Concurrent Observation CO)
- you will be asking the students to describe what they are doing as they attempt the item (Concurrent Think Aloud CTA) [Third testlet only]
- you will be asking students to explain what they just did directly after they answered the item (Concurrent Interview CI).

All Testlets

CO Concurrent Observation items: On the response sheet please make a note of any relevant issues you see. Some items have been flagged as being of particular interest. The following table provides a guide for areas of observation.

Both domains	
Item/text difficulty	Student comments on the difficulty of texts or items.
Engagement	Student comments or behaviours indicating interest in or particular engagement with a text or an item.
Disengagement	Student comments or behaviours indicate a lack engagement with a text, item or the test as a whole.
Branching points	Items before and after branching points have been highlighted on the observation sheet. Note down behaviours/comments for these items in particular.
Testlet navigation	Movement between items is restricted to within testlets. Once students leave a testlet they can't return to those items. Note down any issues that arise as a consequence.
Technical issues	This study is not a review of the test delivery platform but issues students may have with the platform should be noted as these may impact engagement or performance.
Reading specific issues	
Stimulus text interaction	Note down these behaviours when they occur: Student returning to the text to answer items. Student starting to answer items before reading the text. Student using the mouse or a finger to track lines or words in the text.
Stimulus sizing	Note down these behaviours when they occur: Student using the Assessment Master 'expand screen' function for the stimulus text Student using the scroll bars to move the text. Student NOT scrolling a text even though some of the text is hidden. Student using web browser's zoom function (e.g. ctrl+/ctrl- in Firefox) to resize stimulus text.
New item types	Note down behaviours around new item types: Multiple-response: do they notice these are different than usual MC items? Sequencing: how does the student go about answering these? Short response: How much text do they enter? Are they unsure of how much to write?

CI Concurrent Interview items: Once the student has given a response, interrupt them before they begin the next item. Ask them to explain how they attempted the item.

Please stop working for a moment [wait for the student to pay attention to you. If they have already clicked onto the next item ask them to click back to the relevant item.]

For this question that you just answered, I would like you to explain to me how you chose your answer.

Make short notes on what they say. Then ask them:

Can you give me a number from 1 to 5 describing how difficult you found this question? 1 means very easy and 5 means very hard.

CI Concurrent Interview supplementary (whole item set): Once a student has answered the last item in an item set and before they start the next item set interrupt them before they start reading the next text.

Please stop working for a moment [wait for the student to pay attention to you. If they have already clicked onto the next item set ask them to click back to the relevant item.]

Thinking of all the questions for this text [say the title of the text] *can you give me a number from 1 to 5 describing how difficult the whole set of questions were? 1 means very easy and 5 means very hard.*

Again thinking of all the questions for this text [say the title of the text] *can you give me a number from 1 to 5 describing how interesting the whole set of questions were? 1 means very boring and 5 means very interesting.*

Overall would you like to see more sets of questions like that in NAPLAN or fewer sets like that?

Testlet F/C/E Only

When students reach the start of the third testlet stop them. During this testlet you will be asking them to attempt some items using the ‘think aloud’ method. Explain this process.

For some questions, I will ask you to think aloud as you are attempting to answer them. What I mean with think aloud is that I want you to say out loud everything you say to yourself silently as you answer the question. Let’s look at an example.

Show the student the printed example item for their year level. Then model think aloud by answering the question while speaking your thoughts about the steps you are taking.

Once the student is ready get them to proceed onto the third testlet. Follow the same process for CO and CI as in the previous testlets. For CTA items follow the steps below.

CTA Concurrent Think Aloud items: Before the student starts the item set interrupt them. Explain that for the questions in this item set you would like them to speak their thoughts out loud as they work out their answer.

Please stop working for a moment [wait for the student to pay attention to you]

For the next few questions I would like you to talk clearly while you answer the questions. Please say what you are thinking as you work.

When they have finished their response, thank them.

Finishing the test

Follow the procedure for finalising their answer [to be confirmed].

Stop the video camera recording and switch it off and explain to the student that the camera is now off.

Ensure they have completed all the necessary steps to finalise their test. Thank them for their responses and their effort. Make sure they have written their name on any scrap paper they have used and collect it off them.

Ensure they have a chance for a short rest and a toilet break.

Phase 2: Numeracy Tests

Introduce the test

Explain how you will proceed.

Firstly explain that you would like them to comment throughout the test:

In a moment I would like you to start the test. The camera will record what happens on the screen. I will be making notes about how the test is working for you.

As you attempt each question please feel free to say anything about the question you are working on.

First, we are really interested in when you notice a change in difficulty in the texts or questions, so when that happens let us know.

Just as much, we’re interested in when you get bored, or when you find something in the test you really like, so be sure to tell us when that happens.

Finally, if there’s anything that you have trouble with in the process of taking the test that makes it hard to focus on just doing the questions, that’s important to us too.

Then explain that you will also be stopping them from time to time to ask them questions:

From time to time I will also ask you to stop for a moment. I may then ask you some questions about what you are doing or thinking. This will never be because I think you are doing something wrong. I will only ask questions because I want to learn something from you. Anything you can tell me will be helpful.

Then explain the ‘Think Aloud’ process:

For some questions, I will ask you to think aloud as you are solving them. What I mean with think aloud is that I want you to say out loud everything you say to yourself silently as you solve the problem. Let’s look at an example.

Show the student the printed example item for their year level. Then model think aloud by solving the problem while speaking your thoughts about the steps you are taking.

Unfortunately I cannot help you with any of the questions or explain what you have to do. If you are stuck, please say so and if you really cannot answer a question move on to the next one.

Ask the student if they have any questions about the process and answer them. Then say:

I am now starting the camera recording.

Start the camera recording and tell the student they may begin the test.

Start the test

For the first three items prompt the student to tell you about:

- The difficulty of the question: Did you find that question easy or difficult?
- The interest level of the question: Did you find that question interesting?

Note down their answers and any other notable comments. Encourage the students to make similar comments at any point during the test.

Continue the test

Please refer to the year level specific observation sheet for guidance on specific items.

The sheet indicates items that:

- you will be observing only (Concurrent Observation CO)
- you will be asking the students to describe what they are doing as they attempt the item (Concurrent Think Aloud CTA)
- you will be asking students to explain what they just did directly after they answered the item (Concurrent Interview CI).

All Testlets

CO Concurrent Observation items: On the response sheet please make a note of any relevant issues you see. Some items have been flagged as being of particular interest. The following table provides a guide for areas of observation.

Both domains	
Item/text difficulty	Student comments on the difficulty of texts or items.
Engagement	Student comments or behaviours indicating interest in or particular engagement with a text or an item.
Disengagement	Student comments or behaviours indicate a lack engagement with a text, item or the test as a whole.
Branching points	Items before and after branching points have been highlighted on the observation sheet. Note down behaviours/comments for these items in particular.
Testlet navigation	Y3 and Y5 Numeracy: Movement between items is restricted to within testlets. Once students leave a testlet they can't return to those items. Note down any issues that arise as a consequence. Y7 and Y9 Numeracy: Movement is restricted to advancing to the next item. Students cannot return to a previous item. Note down any issues that arise as a consequence.
Technical issues	This study is not a review of the test delivery platform but issues students may have with the platform should be noted as these may impact engagement or performance.
Numeracy specific issues	
Interaction with graphics	For example: Students using the mouse or a finger to help engage with a diagram. Use of physical aids such as paper or a straight edge to compare parts of a diagram.
Scrap paper use	Note down any use of paper by students to help solve problems.
Online calculator use	Note down behaviours around calculator use.

CI Concurrent Interview items: Once the student has given a response interrupt them before they begin the next item. Ask them to explain how they attempted the item.

Please stop working for a moment. [Wait for the student to pay attention to you. If they have already clicked onto the next item ask them to click back to the relevant item.]

For this question that you just answered, I would like you to explain to me how you chose your answer.

Make short notes on what they say. Then ask them:

Can you give me a number from 1 to 5 describing how interesting this question was? 1 means very boring and 5 means very interesting.

Can you give me a number from 1 to 5 describing how difficult you found this question? 1 means very easy and 5 means very hard.

CTA Concurrent Think Aloud items: Before the student starts the item interrupt them. Explain that for this question you would like them to speak their thoughts out loud as they work out their answer.

Please stop working for a moment. [Wait for the student to pay attention to you]

For this question I would like you to talk clearly while you solve the problem. Please say what you are thinking as you work.

When they have finished their response, thank them.

Finishing the test

Follow the procedure for finalising their answer [to be confirmed].

Stop the video camera recording and switch it off and explain to the student that the camera is now off.

Ensure they have completed all the necessary steps to finalise their test. Thank them for their responses and their effort. Make sure they have written their name on any scrap paper they have used and collect it off them.

Ensure they have a chance for a short rest and a toilet break.

Phase 3: Post test – Both domains

Log back in

Thank the student for their responses so far.

Thank you for all your answers so far. What you have told me will be very helpful. We have another set of questions to ask you.

The first questions are about specific parts of the test. I'll need you to sit back at the computer so we can look at some of them together.

Log the student back into the system using a log-in that corresponds with the same initial pathway they took in Phase 2.

Retrospective interview: Items

Take students to particular items in the order they appear in the test.

Targeted Items

Please refer to the year level specific question sheet for guidance on specific items. Targeted items will include: any items around key branching points that have not yet been targeted (Y7 & Y9 Numeracy), alternate item types (multiple-response, sequencing, short response, drag & drop) used in the first 3 testlets or other items of particular interest that have not already been targeted.

Take the student to the item. Ask them to show you how they worked out the item. Ask any additional questions for that item from your year level specific question sheet. Make short notes on what they say.

For alternate item types, ask:

If you had to answer the same question in a normal paper based test would you find it harder or easier or about the same difficulty?

Probe as needed: *Can you explain what made it harder/ easier/ about the same difficulty?*

Would you attempt the question differently if it was presented on paper?

Probe as needed: *How would you go about answering the question if it was on paper?*

For short responses ask:

For this question how did you decide how much to write?

Retrospective interview: Whole test

General impressions

In this section you will be asking the student about their experience with the whole test.

One difference between the test you took today and other tests you usually take (for example NAPLAN) is that you took it on the computer rather than on paper. Did the test you took today differ from the NAPLAN tests you have taken in the past in any other ways?

Probe as needed: *Are there any other ways the tests differ? Can you explain what you mean? /Can you give an example?*

Ask the student about their overall engagement with the test.

How interesting or engaging did you find the test? Was it more engaging than NAPLAN, less engaging or about the same?

Probe as needed: *What was it about the test you found more/less engaging? What makes a test more/less engaging for you?*

Ask the student about their overall perception of the test difficulty.

How difficult did you find the test? Was it more difficult than NAPLAN, less difficult or about the same?

Probe as needed: *What was it about the test you found more/less difficult? What makes a test more or less difficult for you? Did the number of texts included in the test make a difference? (For Reading only)*

Ask the student about how the overall difficulty changed.

Overall did the test seem to get harder as you worked through it, or easier or was it more mixed?

Probe as needed: *Can you explain a bit more? Give an example?*

Branching process

Explain to the student the branching process:

This test was actually in three parts of different levels of difficulty.

- For students on a C pathway: *The third part of the test you did was meant to be easier than the first two parts.*
- For students on an E pathway: *The third part of the test you did had a few questions at the end that were meant to be harder than the rest of the test.*
- For students on an F pathway: *The third part of the test you did was meant to be harder than the first two parts.*

Show the student the last testlet they did. Explain this was the third part of the test.

Ask the student:

Did you actually notice the test changing in any way when you moved from part 1 to part 2? Did you notice the test changing in any way when you went from part 2 to part 3?

Then ask the student:

Which of the three parts did you enjoy the most?

Which part did you find easiest? Which part did you find hardest?

For testlet C: *Did the easiness of the last part make a difference to how you liked the test/this part of the test?*

For testlet F: *Did the difficulty of the last part make a difference to how you liked the test/this part of the test?*

Ask the student if there is anything else they would like to talk about regarding the test. Note down any comments they make.

Finishing up

Stop the video recording and turn off the camera. Show the student that the camera is now off. Thank the student again for all their hard work. Check with them that they know where they need to go next and send them on their way.

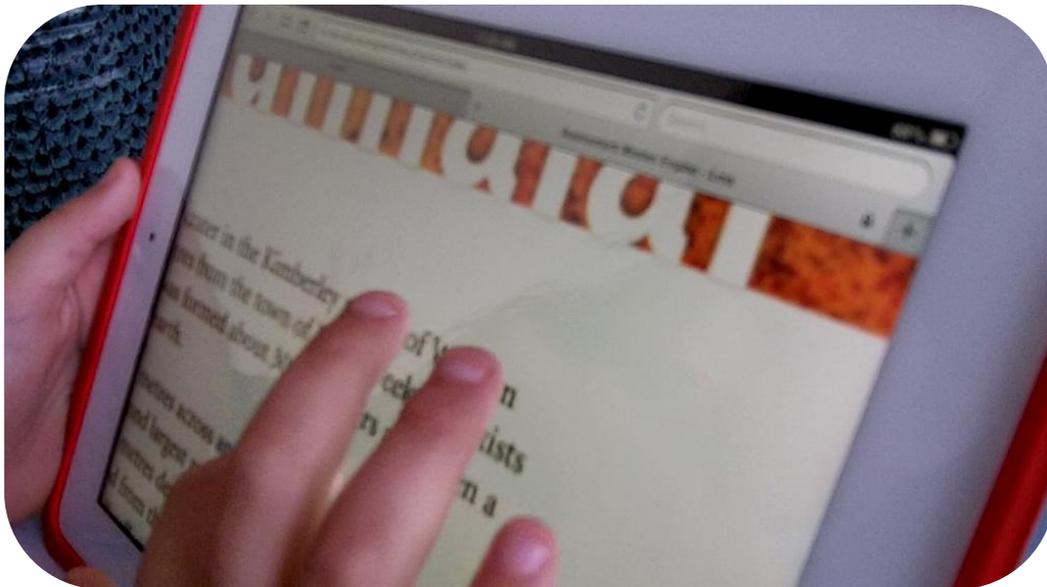


National Assessment and Surveys Online Program

NAPLAN Online Tailored Test Design August 2013

Cognitive Interviews – Numeracy and Reading

Part 2: Findings for Reading



Contents

Introduction.....	4
Background.....	4
The NASOP multistage model.....	4
Methodology.....	5
Protocol development.....	5
Sampling approach.....	6
Final sample.....	6
Findings.....	10
Theme: Branching.....	10
Theme: Computer-based test experience.....	15
Theme: Engagement.....	20
Conclusions and recommendations.....	27
Branching/Tailored test design.....	27
Computer-based test experience.....	27
Engagement.....	27

Tables

Table 1 Difficulty labels for pathways.....	4
Table 2 Targeting of items by method of study	6
Table 3 Final school sample by state and sector	6
Table 4 Proposed school sample by state and location.....	7
Table 5 Geographical distribution of schools in final sample	7
Table 6 Students interviewed by location and sector.....	7
Table 7 Percentage of students by gender.....	7
Table 8 Percentage of students who followed a given pathway	8
Table 9 Free and fixed pathways	8
Table 10 Platforms observed	8
Table 11 Computer use (percentages of total)	9
Table 12 Did the test differ from NAPLAN?	13
Table 13 Was the test easier or harder than NAPLAN?	13
Table 14 Comparative perceived difficulty by pathway (percentage of row)	13
Table 15 Student perception of difficulty progression	14
Table 16 Perceived difficulty progression by pathway (percentage of row)	14
Table 17 Students noticing change between testlets.....	14
Table 18 Assorted navigation/timing issues	15
Table 19 Reported engagement	22
Table 20 Preferred testlet	22
Table 21 Preferred testlet (percentage of column).....	23
Table 22 Effect of difficulty of third testlet	23
Table 23 Effect of difficulty of third testlet on preference	23

[NOTE: Due to rounding, some sets of percentages do not appear to total exactly 100%]

Figures

Figure 1 Pathways.....	4
Figure 2 Average difficulty estimate by pathway	10
Figure 3 Average difficulty estimate ADF testlets	11
Figure 4 Average difficulty estimate ABC testlets	12
Figure 5 Average difficulty estimate ABE and ADE testlets	12
Figure 6 Pinching text on an iPad (R0013) and stretching text on an iPad (R0012).....	16
Figure 7 Stimulus display on a notebook with a small screen.....	16
Figure 8 Short stimulus text from testlet C displayed adequately on a small screen.....	16
Figure 9 Stimulus text displayed on a small notebook, large desktop and an iPad	17
Figure 10 (R0012) Student logging back in.....	18
Figure 11 R0004 impatiently waiting for next question to load	18
Figure 12 Double scrolling for student R0007	19
Figure 13 Screen which confused Student R0006	19
Figure 14 R0024 tracking options with cursor	20
Figure 15 R0028 highlighting text.....	20
Figure 16 Scatterplot of averages of student ratings for difficulty and engagement	21
Figure 17 Average engagement ratings by year level and testlet	21

Introduction

Background

The National Assessment and Surveys Online Program (NASOP) is a series of projects by the Australian Curriculum Assessment and Reporting Authority (ACARA) that investigate the online delivery of ACARA's assessment programs. As part of NASOP, a trial of online reading and numeracy tests took place in August 2013.

The online tests were designed around a partially adaptive multistage test model. This model consists of a series of mini-tests (known as testlets or modules) of varying difficulty. Students follow different paths through the testlets, depending on their score in each testlet.

EAA was commissioned by ACARA to conduct cognitive interviews to investigate the impact of the multistage-branching design on students' test-taking behaviour and to gather information about students' interaction with the testlets at key branching points of the test.

This document is Part 2 of a set of three documents discussing findings from the cognitive interviews, and focuses on data from the Reading domain. More general information about the sampling process and comparative data between Reading and Numeracy can be found in Part 1. To enable this document to be read in isolation, some information from Part 1 is repeated in this document.

The NASOP multistage model

The model for the test piloted in this study is a multistage partially adaptive branching model. A test at a given year level consists of six testlets of items of known difficulty. Students were presented with three of these testlets.

The following diagram shows the branching and facility range of the testlets.

Figure 1 Pathways



The model has four distinct pathways:

Table 1 Difficulty labels for pathways

Difficulty	Sequence of testlets
High	ADF
Middle	ADE
Middle	ABE
Low	ABC

Testlets A, B and D are the branching testlets. The student's score in a branching testlet determines which testlet they proceed to next. ACARA requested that key items in the two branching testlets through which a student passed be discussed in detail in the cognitive interviews.

Fixed-path model

While the intended model of the test involves adaptive branching, for the purpose of this trial half of the students were allocated tests with a predetermined path. This fixed-path model ensured that sufficient numbers of students undertook all of the available pathways during the trial.

Methodology

Protocol development

The priority aspects investigated in this study as requested by ACARA were:

1. the impact of the multistage-branching design on the students' test-taking behaviour
2. students' interaction with the testlets at key branching points of the test:
 - a. investigate key items in the two branching testlets which students pass through
 - b. investigate student response to items which show marked changes in difficulty
3. student performance on testlets C and F.

EAA's methodology for protocol development proceeded from priorities set by ACARA. Protocols were devised around the three interrelated themes of branching, computer-based test experience and engagement.

- **Theme: Branching**

The major innovation of the test design for this trial is the branching process and it is this process that forms the primary theme of the cognitive interview study.

- **Theme: Computer-based test experience**

The term 'Computer-based test experience' (CBTE) relates to those aspects of the test experience that are particular to a computer-based test or which are significantly affected by computer-based delivery of the test. This theme was the primary focus of the 2012 NASOP study.

- **Theme: Engagement**

Student engagement with the test may be affected by both the mode of the delivery and the branching model (e.g. through improved targeting of items to student ability). Student engagement was a main focus of this study.

This was done using four methods:

- Concurrent observation (CO): observing student behaviour during the test
- Concurrent think-aloud (CTA): asking students to 'think aloud' while answering an item
- Concurrent interview (CI): asking students questions about the item or item set they have just answered
- Retrospective interview (RI): asking students questions about items in an interview that is held after the test is complete.

The protocols were developed collaboratively by EAA staff members who had experience in cognitive interviewing and an understanding of either the NAPLAN Numeracy or NAPLAN Reading constructs. The Reading experts had experience with literacy test development as well as NAPLAN Reading item development.

This table outlines the quantities of items for particular study:

Table 2 Targeting of items by method of study

Method	Number of items	Theme	Reading Items
Concurrent	Observation	All	All
	Think-aloud	4 (approx)	CBTE, Engagement 4: one item set F/C/E & informally throughout test
	Interview	4 (approx)	Branching, Engagement 2: final items of A & B/D 1: initial items of B/D 1: initial items of F/C/E + supplementary question for each complete item set all testlets
Retrospective	Interview	4 (approx)	CBTE, Engagement Any sequencing, multiple-response or short response items not already targeted in approximately these proportions: 2: selected items A & B/D 2: selected items F/C/E

Sampling approach

Full details of the sampling approach are given in the document *Part 1: Joint Summary Findings for Numeracy and Reading*.

Final sample

Schools

In total, sixteen schools participated in the cognitive interviews, approximately 40% of the schools in the proposed sample. Three factors contributed to a school's not being able to participate.

1. Non-return of permission slips. Several schools did not return any permission slips and consequently could not be included in the cognitive interviews.
2. Scheduling issues. Cognitive interviews were intended to take place on the same day as the main study and in some cases it was not possible to send interviewers to the school on the day designated by the main study contractor.
3. Several schools in the proposed sample opted out of the main study completely.

To mitigate these issues a number of the schools that had opted out completely were re-contacted to see if they would be willing to participate in the cognitive interviews only. In addition, one school (an independent Christian school in Sydney) which had not been listed in the proposed sample was included in the cognitive interviews.

Table 3 Final school sample by state and sector

State	Sector			Total
	Cath	Gov	Ind	
NSW	0	9	4	13
QLD	1	2	0	3
Total	1	11	4	16

The proportion of NSW and Queensland schools was also affected, with a greater proportion of NSW schools included in the interviews. However, the proportion of metropolitan and regional schools was largely unaffected.

Table 4 Proposed school sample by state and location

State	Location		
	Metro	Regional	Total
NSW	9	4	13
QLD	2	1	3
Total	11	5	16

Despite many schools not being able to participate in the study, the final sample included a broad range of geographical areas.

Table 5 Geographical distribution of schools in final sample

State	Geographical distribution		
	Location	Area	Total
NSW	Metro	Sydney	6
		Campbelltown	1
		Western Sydney	2
	Regional	Blue Mountains	1
		Newcastle	1
		Orange	1
		Wollongong	1
QLD	Metro	Brisbane	2
	Regional	Rockhampton	1
Total			16

Within this sample of schools, three schools were single-sex schools. In all cases the single-sex schools were girls' schools.

Students

For Reading, a total of 41 students were interviewed – ten in Year 3, eleven each in Years 5 and 7 and nine in Year 9. Students were drawn primarily from government schools in metropolitan areas, but significant numbers were also drawn from regional locations and from independent schools.

Table 6 Students interviewed by location and sector

Location	Sector			Total
	Cath	Gov	Ind	
Metro	0	20	8	28
Regional	2	7	4	13
Total	2	27	12	41

More girls than boys were interviewed. The main factors in the gender imbalance were issues with the return of permission notes, selection of students by the school and by the presence of girls' schools in the sample without a balancing presence of boys' schools.

Table 7 Percentage of students by gender

Gender		
Female	Male	Total
66%	34%	100%

Pathways

Because of the importance of the tailored nature of the test, a sustained effort was made to interview students with a range of abilities. This was done firstly by including a range of schools in the proposed sample and secondly by contacting schools and asking them to identify students within given ability ranges. However, the final selection of students was dependent on the return of permission notes and the availability of students on a given day. As can be seen in this table, the net effect of these issues was a tendency for more able students to be interviewed.

Table 8 Percentage of students who followed a given pathway

Path		Reading
Low	ABC	24%
Middle	ABE	10%
	ADE	15%
High	ADF	51%
Total		100%

Pathways were also allocated as either “fixed” or “free”. With fixed pathways, students were allocated a predetermined sequence of testlets. With free pathways, the sequence of testlets was determined by the student’s score up to the branching point. It is notable that the ADF pathway was the most frequently followed pathway in Reading when a free pathway was allocated.

Table 9 Free and fixed pathways

Path		Reading
Fixed	ABC	8
	ABE	2
	ADE	2
	ADF	8
Fixed Total		20
Free	ABC	2
	ABE	2
	ADE	4
	ADF	13
Free Total		21
Total		41

Platforms and computer use

Students interviewed attempted the test on a range of different platforms. However, it should be noted that to accommodate the interviews schools often allocated computers that were available in quiet locations suitable for one-on-one observation. Consequently the platform used for the interview at a given school may not have been the same platform as was used by students in the main study at the same school.

Table 10 Platforms observed

Platform	OS	Reading
Desktop	Mac	4
	PC	15
Big Notebook	Mac	6
	PC	12
Small Notebook	PC	2
Tablet	iOS	2
Total		41

It had been hoped that more students would have been observed using iPads, but at two schools technical issues unrelated to the platform prevented observation of students attempting the test on iPads.

At the start of the test students were asked three questions about their familiarity with computers. The following tables show the proportion of responses for Reading.

Table 11 Computer use (percentages of total)

Q1. How often do you use computers or tablets at school?			
most days	some days	only for some classes	not often
39%	49%	12%	0%
Q2. How often do you use computers or tablets at home?			
every day	sometimes	not often	never
49%	41%	10%	0%
Q3. How often do you use a mobile phone with a touch screen?			
every day	sometimes	not often	never
44%	24%	17%	15%

As can be seen from the tables the majority of the students were regular users of IT in school and at home.

Findings

Theme: Branching

The major innovation of this trial was the branching process and this process formed the primary theme of the cognitive interview study.

Students' responses to branching were studied in a number of ways. During the concurrent phase of the interviews students were asked to rate item sets on both difficulty and engagement/interest. Additionally, during the retrospective phase students were asked several questions designed to evaluate the extent to which they had been aware of the branching process.

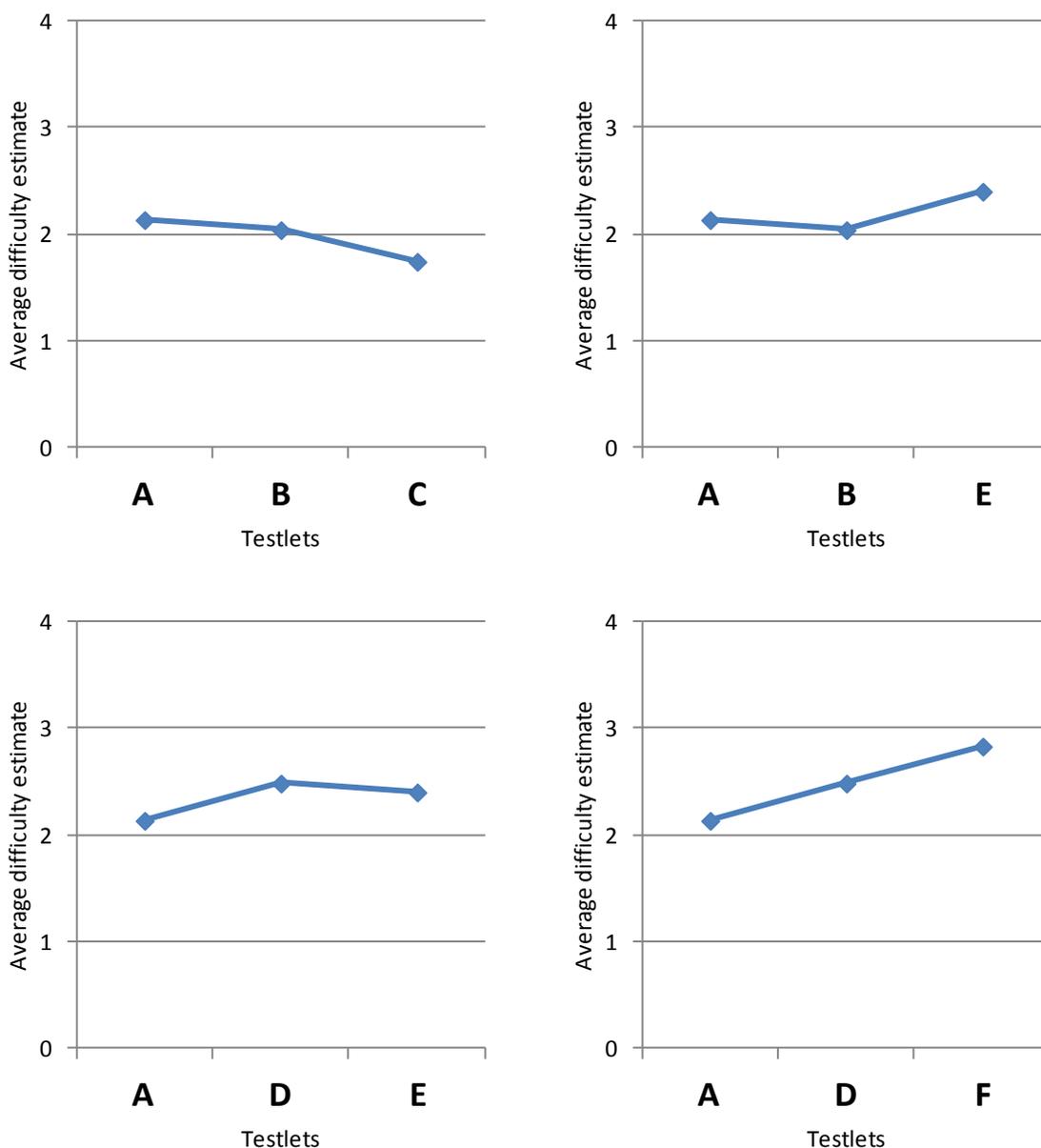
Concurrent data

At the end of each item set students were asked to rate the item set's difficulty on a 1 to 4 scale (4 being the hardest end of the scale).

In general, students perceived the structure of the test as mixed. Most students identified a mix of difficulty in each testlet. The move between testlets at branch points usually resulted in a harder text being followed by a (relatively) easier text and this shift was noticed by students.

The following graphs show average estimate of difficulty across all year levels by each kind of pathway.

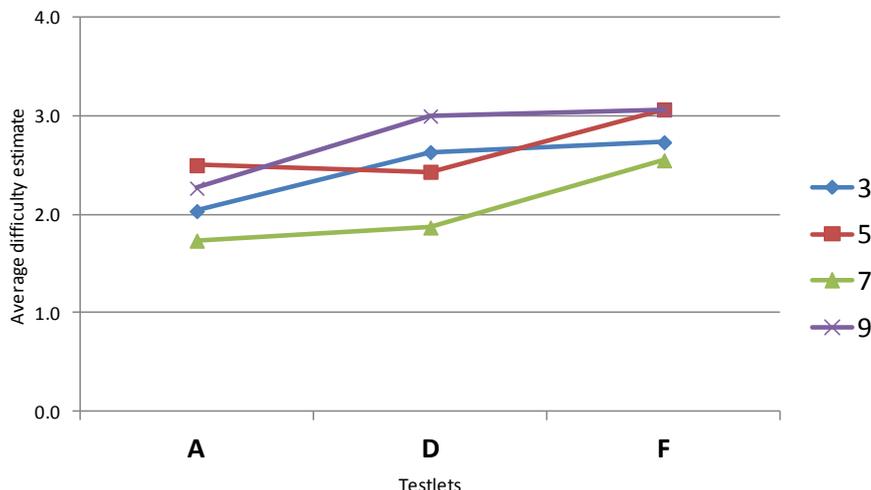
Figure 2 Average difficulty estimate by pathway



Averaged across several item sets and year levels, the perceived difficulty progressions of the pathways do reflect the intended test design. This suggests that the actual objective difficulty of the items sets had at least some effect on student perceptions of the test pathways.

However, these patterns are only obvious by averaging many responses across year levels. When the data is more disaggregated, student experiences can be seen to be more varied. In addition, interviewers reported that students on the ADF pathway perceived a change going from A to D, but felt the progression from D to F was smoother. The perceived comparative difficulty of the F testlet was not consistent across years. In Year 7 the average shift from D to F is most pronounced but for some students D was perceived as being more difficult than F. However, interviewers observed that error rate increased in F, even if the student was not reporting increased difficulty.

Figure 3 Average difficulty estimate ADF testlets



Students who took the ADF pathway were sometimes aware of the increased difficulty coincident with movement to testlet F, though many were not perceptive about this increased difficulty even when their performance on the test worsened. The difficulty seemed more pronounced in some years than others.

(R0014 Y9 Fixed ADF) Observer: So what do you think of this text and these questions [first text of testlet F] compared to the last ones that you did...?

Student: I love it, it's really interesting, but it takes a while for everything to be comprehended. It's like reading Jane Austen or something where you have to read it twice.

Observer: So you have to read it more times, or more thoroughly, to understand it?

Student: Yeah, you've just got to think about it more.

(R0012 Y3 Free ADF) Hiding in plain sight is a 4, it's harder than Down by the river, it has more hard words and the writing is fuzzy. There are quite a few words that people don't understand.

(R0020 Y9 Fixed ADF) This is harder, the options for the questions are harder, some I don't understand, and I had to read the text over and over again.

(R0019 Y9 Fixed ADF) Layout and length made it hard to know what's going on – really hard... I just eliminate some answers for each question and then I guess. This is a much harder text.

(R0026 Yr7 Free ADF) At the beginning it was quite easy and you could understand everything and then it got progressively harder and I started to not understand some stuff and the questions were more difficult.

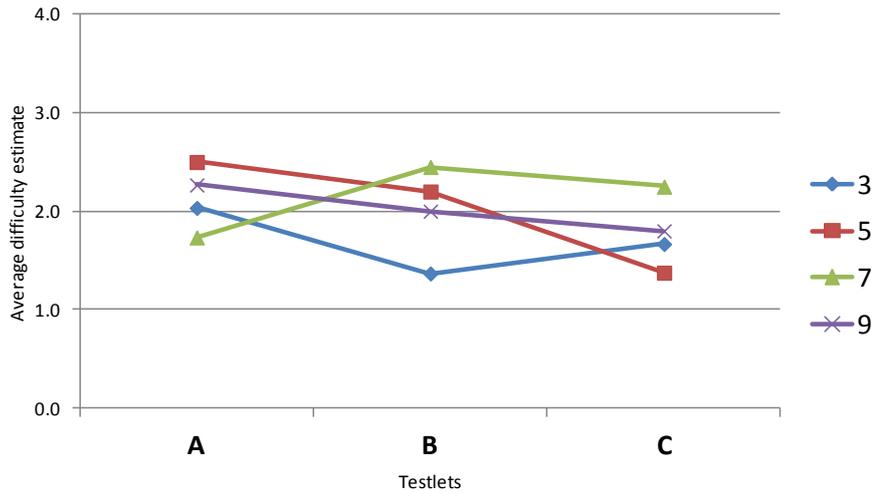
Most students were aware of testlet C being easier. Most students who took the ABC pathway recognised a significant decrease in difficulty coincident with movement to testlet C. Many students commented on the difference between the last item set of testlet B and the first item set of testlet C.

(R0032 Y9 Fixed ABC)(Retrospective): The GPS one [terminal text in testlet B] was really confusing, and I didn't really get it, and then when we went to the email one [initial text testlet C] it was heaps more easy.

(R0006 Y5 Fixed ABC) judged Space junk a 4 for difficulty That's the hardest one I've done, I didn't get some of the words and I didn't get what the answers were. Then The littlest bird (first text in testlet C) was rated 1 – it's pretty easy, much easier than the last one. The littlest bird also scored 4 for interest.

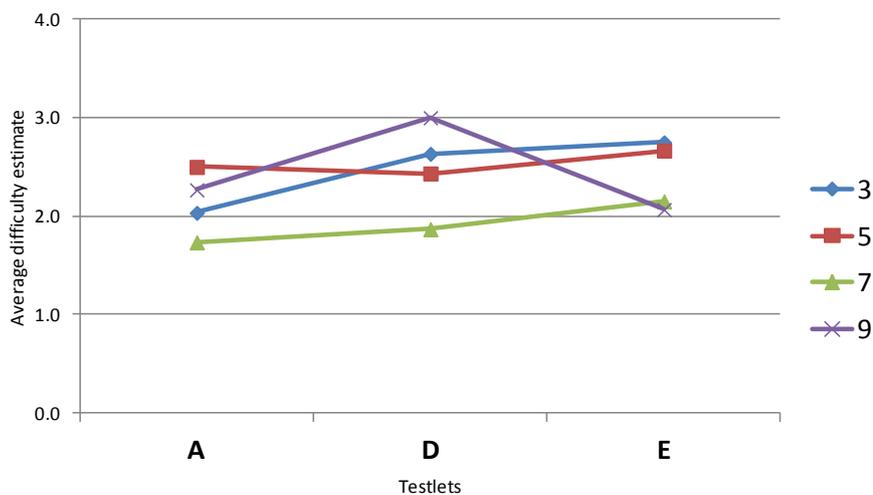
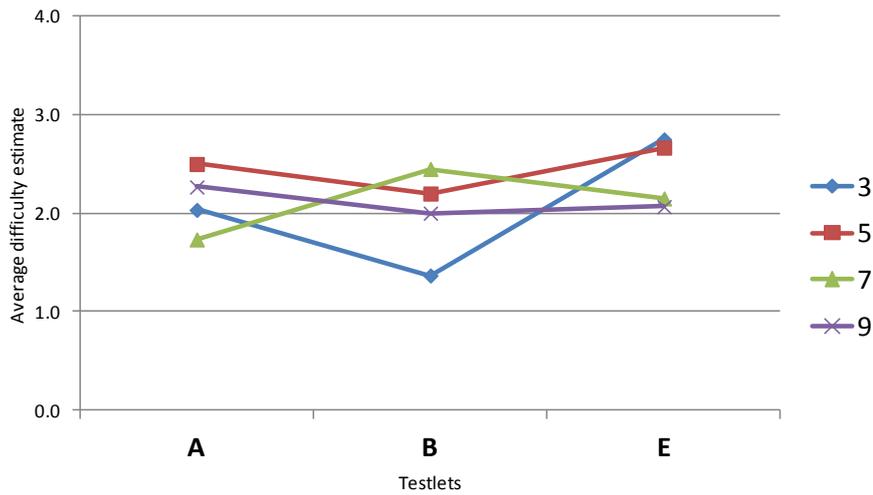
Even so student difficulty estimates for the ABC testlets still paint a somewhat mixed picture.

Figure 4 Average difficulty estimate ABC testlets



Students taking the middle ABE and ADE pathways (and sometimes ADF) typically perceived the overall difficulty of the test as ‘mixed’ or ‘up and down’, especially when both inter- and intra-testlet perceptions of variations in difficulty are considered.

Figure 5 Average difficulty estimate ABE and ADE testlets



While students noticed and could discuss this mixed difficulty pattern, it did not seem to have any significant impact on their test experience.

(R0008 Y5 Free ABE) It was more mixed – some easy then hard, then maybe easy again then medium.

(R0027 Y7 Free ADF) This was mostly easier to hardest but some seemed to be taken from the hard end and placed near the easy. Definitely at the end they seemed mixed around.

(R0028 Y9 Free ADE) The difficulty was mixed. The easiest text (Urban Greenie) was halfway through, in amongst harder texts. The difficulty felt random or alternating hard and easy.

Only rarely was the zigzag difficulty perceived as an ‘issue’ for students. Two reported being unsettled by the unpredictability of the test, while two thought that unpredictability was a positive outcome of the design.

Retrospective data

To see if students were aware of the branching, the first question in the retrospective interview asked them for any differences they noticed between the test they had just completed and NAPLAN.

Q1: One difference between the test you took today and other tests you usually take (for example NAPLAN) is that you took it on the computer rather than on paper. Did the test you took today differ from the NAPLAN tests you have taken in the past in any other ways?

Most students responded that test was in some way different from NAPLAN.

Table 12 Did the test differ from NAPLAN?

Did it differ?	Reading
Yes	29
No	11
Other	1
Total	41

None of the students who responded affirmatively identified the order of difficulty of the items as being a difference. Typically affirmative responses related to the test delivery mode. Apart from the mode of delivery, students did not perceive the test experience as being qualitatively different from their most recent paper NAPLAN test.

Question 3 of the retrospective interview focused more specifically on the issue of test difficulty.

Q3: How difficult did you find the test? Was it more difficult than NAPLAN, less difficult or about the same?

Table 13 Was the test easier or harder than NAPLAN?

How difficult?	Reading
Easier than NAPLAN	13
About the same	24
Harder than NAPLAN	2
Other	2
Total	41

In general, students found the test easier or about the same difficulty as NAPLAN. Comparing these responses against the pathways students followed shows some relation with the difficulty of the pathway. All of the students who reported that the test was harder than NAPLAN had followed the ADF pathway.

Table 14 Comparative perceived difficulty by pathway (percentage of row)

How difficult?	Path				Total
	ABC	ABE	ADE	ADF	
Easier than NAPLAN	46%	8%	23%	23%	100%
About the same	17%	13%	8%	63%	100%
Harder than NAPLAN	0%	0%	0%	100%	100%
Other	0%	0%	50%	50%	100%
Total	24%	10%	15%	51%	100%

Question 4 probed the issue of the progression of difficulty further.

Q4: Overall did the test seem to get harder as you worked through it, or easier or was it more mixed?

In general, students thought the difficulty progression was more mixed than any general progression from easy to hard or vice versa, possibly due to the “reset” effect between testlets. This is where difficulty declines between the terminal text of one testlet and the initial text of the following one. Additionally, irregular shifts of difficulty within testlets (where an easy text is followed by a hard text, which is followed by an easy one) may have contributed to an overall impression of a mixed difficulty progression.

Table 15 Student perception of difficulty progression

Difficulty change	Reading
Got easier	4
Got harder	9
Mixed	24
Other	4
Total	41

“Mixed” responses included responses like these:

(R0042 Y9 Free ADE) Normally tests get harder, in this they were all over the place. One was hard, and then one was easy. Glowsticks was easy compared to Salinity. (The changes in difficulty) did throw me a bit, not in a good way; it wasn't like a usual test. It didn't get more difficult smoothly.

(R0012 Y3 Free ADF) On paper it's harder when you get further into it but here the texts felt more the same.

(R0001 Y7 Fixed ADE) NAPLAN went easier to harder; this was mixed, easy then hard, not medium.

For one student this unpredictability was a positive feature of the design:

(R0039 Y5 Free ADF) Usually we expect easy to hard, when it's mixed up its good, you don't know what to expect.

For two others it was unsettling: (R0042, quoted above) and (R0029 Y7 Free ADF) who said:

The paper test just gets harder, this didn't seem like it was in any particular order. It uneases me, thinking the next one's going to be harder.

Ignoring ‘other’ responses, students on the ADF pathway accounted for all of the “Got harder” responses in Reading. It should also be noted that while some of the students on the ADF pathway reported that the difficulty change was ‘mixed’, none reported that it got easier.

Table 16 Perceived difficulty progression by pathway (percentage of row)

Difficulty change	Path				Total
	ABC	ABE	ADE	ADF	
Got easier	75%	0%	25%	0%	100%
Got harder	0%	0%	0%	100%	100%
Mixed	25%	13%	21%	42%	100%
Total	24%	8%	16%	51%	100%

The sequence of questioning was designed so that students would receive an increasing amount of prompting on the issue of branching. By question 5, interviewers had explained to the students that the test was intended to be in three parts of varying difficulty. Students were asked whether they had noticed the changes.

Q5: Did you actually notice the test changing in any way when you moved from part 1 to part 2? Did you notice the test changing in any way when you went from part 2 to part 3?

Not all students responded to this question but this table summarises the responses of the students who did.

Table 17 Students noticing change between testlets

	Notice change from 1 to 2?	Notice change from 2 to 3?
Ambivalent	3	3
No	18	11
Yes	18	25
Total	39	39

Branching: Summary

Students could perceive the different ways in which item and stimulus difficulty progressed through different pathways. However, this perception was variable and it is unclear whether students would have paid any attention to it without prompting from interviewers.

Student responses to direct questions regarding the difficulty of items around branching points suggest that they were aware of significant shifts in the difficulty of item sets around branching points. It is notable that most students did not regard the shifts in difficulty as a significant issue or something that distracted them from their test-taking experience.

Theme: Computer-based test experience

The term ‘Computer-based test experience’ (CBTE) relates to those aspects of the test experience that are particular to a computer-based test or which are significantly affected by computer-based delivery of the test. This theme was explored in a number of ways during the concurrent and retrospective phases of the interview. In general, unsolicited comments from students were more likely to pertain to technical issues or general computer-based experience issues than to the other two themes in the study.

On the whole, students were very accepting of the computer-based test experience. However, observers watched students struggle with resizing texts and navigation.

CBTE issues raised in the first question of the Retrospective phase largely related to navigation and timing features of the test.

Table 18 Assorted navigation/timing issues

Navigation	Reading
Navigation between items	3
Isolated items (positive)	6
Isolated items (negative)	2
Timer	2
Total	13

‘Isolated items’ relates to responses from students raising the issue that in the computer-based tests each item is presented in isolation from the other items. Several students commented on liking the presentation of one question at a time. They felt it allowed them to focus on the question without the distraction of glancing at other questions and being seduced into thinking about them. One Year 7 student commented that this test probably took them longer to complete as they actually considered each question in turn.

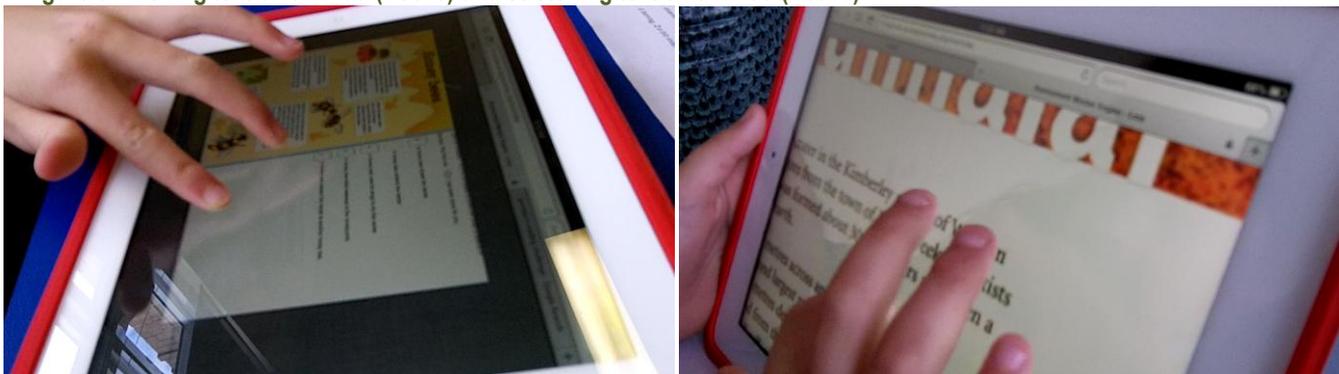
Those students who mentioned the timer generally regarded it as positive feature.

A key issue raised by interviewers and students was the variability of presentation of the test. With students having to attempt the test on computers with different-sized screens and of varying quality, the test-taking experience could be analogous to a paper-based test that was presented to students on different-sized paper and with varying quality of print. Students raised presentation issues in various stages of the interview.

For students taking the test on very small screens (less than 30cm), texts were often not readable without zooming in, and when zoomed in, the question was not fully visible and the student was unable to navigate to the next question.

iPads, despite being small, are well suited to resizing text and images and proved surprisingly adaptable to presenting the stimulus texts.

Figure 6 Pinching text on an iPad (R0013) and stretching text on an iPad (R0012)



Screen size and resolution had a significant impact on the student test-taking experience, regardless of student ability or pathway.

(R0034 Y7 Free ADF) When I zoom I lose the bottom, its irritating.

(R0028 Y9 Free ADE) Zooming was painful; sometimes I'd just squint rather than enlarging.

(R0004 Y5 Fixed ABC) Observer: So, when you answer this one, are you going back and rereading, or are you just remembering what you read before?

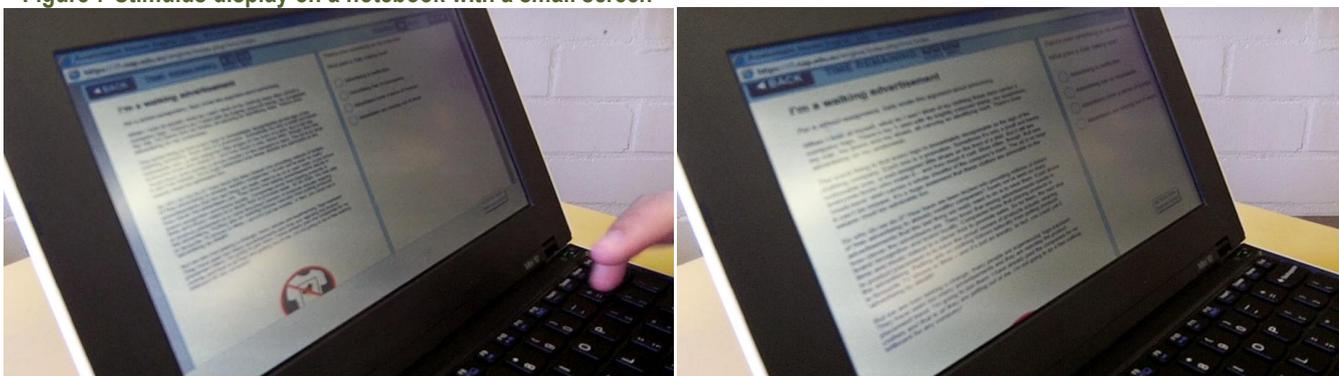
Student: Yeah, I just remember what I read before....

Observer: So are you doing it because it's easy to remember... or because it's annoying to keep zooming back and forth and rereading...?

Student: <laughs> Yes.

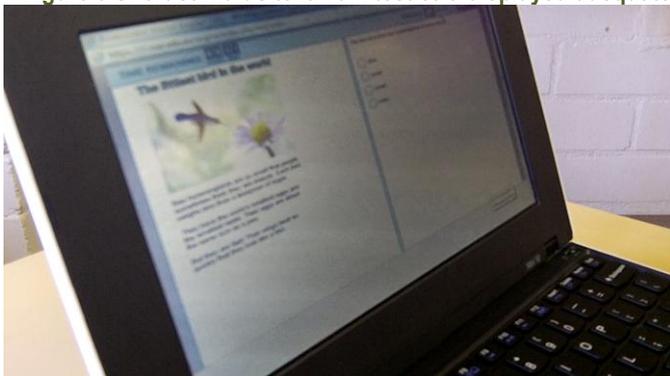
Student R0004 could not see this whole text, and they could not read the text at the size it loaded.

Figure 7 Stimulus display on a notebook with a small screen



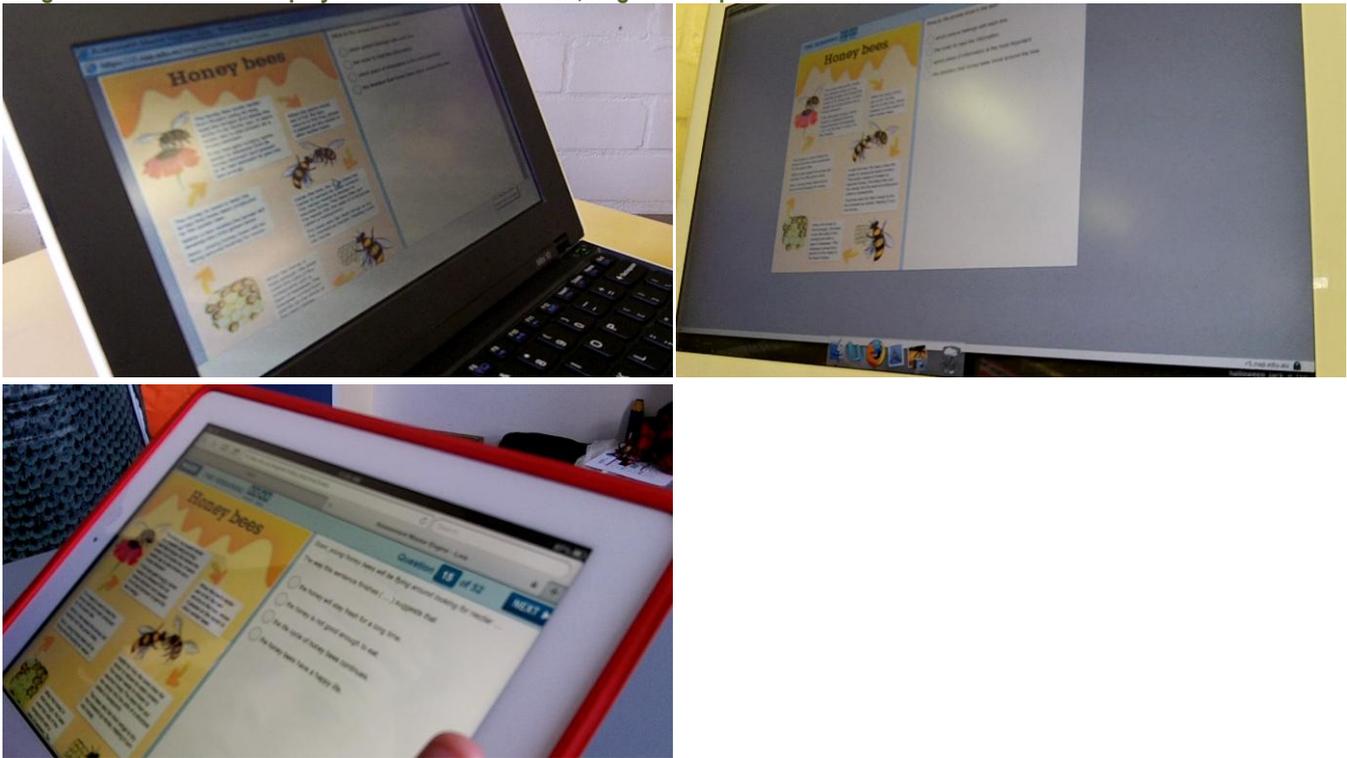
Zooming to read meant loss of some question text and could also lead to the navigation buttons becoming hidden. However in testlet C, the very small text and question did fit on R0004's tiny screen.

Figure 8 Short stimulus text from testlet C displayed adequately on a small screen



Screen size and screen resolution was varied. Small notebooks provided the least good experience while desktops with large screens had more space for texts to be enlarged.

Figure 9 Stimulus text displayed on a small notebook, large desktop and an iPad



One issue of concern was the presentation of stimulus texts on smaller or poorer quality screens. The type for the texts was frequently fuzzy and in some cases very small. The fuzzy type suffered in contrast with the sharp, easily-read type used for the questions and options. Some students found it hard to separate appearance from content in gauging difficulty of a text. One text (*Behind the waterfall*) did not display the final two lines which were covered by an illustration.

Regardless of screen size, some texts at their native size and resolution were regarded by a majority of students as difficult to read easily. This includes longer texts that were unusually small, low-resolution ('blurry'/'not sharp') or for which inter-line spacing made lines too close.

(R0019 Y9 Fixed ADF) The text is hard to read, it's grey and small.

(R0012 Y3 Free ADF) The writing is fuzzy, even when it's big. I think you should change the fuzzy writing, the font is too small.

(R0035 Y3 Free ADF) The more clearer the writing is the better you can read it. Some of them were not as clear and were harder to read. The platypus was the hardest to read. On a paper test they are all clear. When it's not clear it's less interesting.

(R0039 Y5 Free ADF taking the test on a large screen desktop computer.) The size could be a bit bigger.

Network speed and connection problems

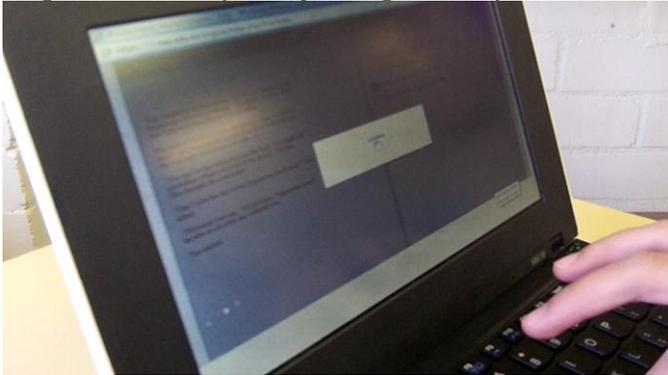
The speed and reliability of a school's internet network, whether wired or wireless, could have a significant impact on a student's test-taking experience, for example when loading times for texts or items are slow, or technical issues such as 'crashing' force a student to log out and log back in.

Figure 10 (R0012) Student logging back in



In some cases, these connection and loading problems resulted in issues as serious as forced abandonment of the test itself. In other cases the student could log back in and return to where they had left off.

Figure 11 R0004 impatiently waiting for next question to load



(R0004 Y5 Fixed ABC waiting for a screen to load from 05:25 to 07:52) Does it usually take this long to load?

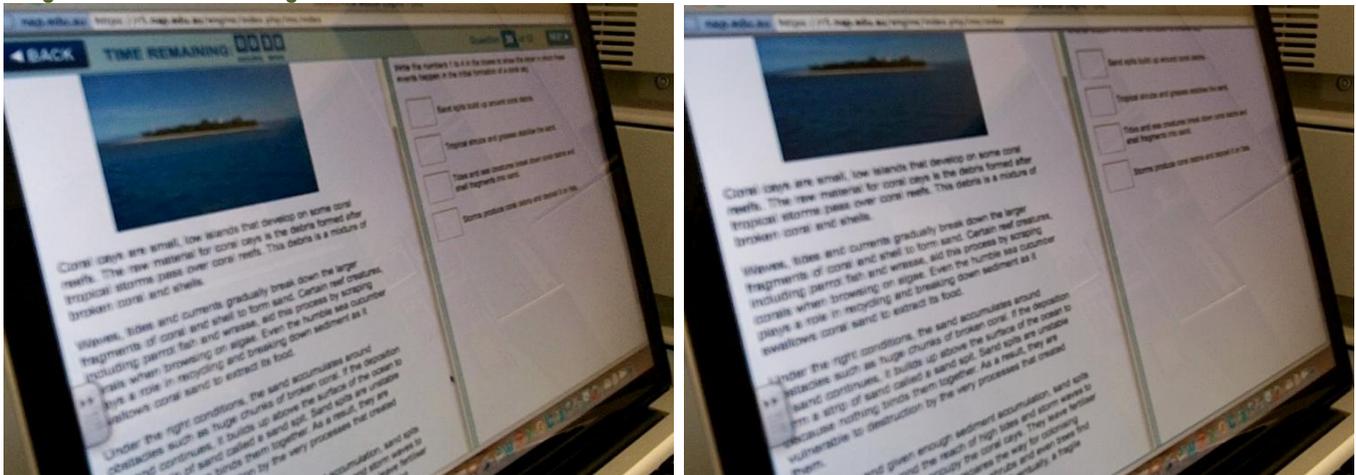
(R0002 Y7 Fixed ABE) The slow loading times made me distracted, I stopped concentrating.

(R0012 Y3 Free ADF had to shut down and log back in several times) I'd recommend doing the test on paper. There are too many technical issues.

Double scrolling

Even on larger screens, a ‘double-scrolling’ issue was observed which frustrated some students. Double-scrolling occurred when the student had to scroll vertically in *both* the text window (central scroll bar) and browser window (right-hand scroll bar) in order to read the bottom of a text. Some students attempted questions without realising that the section of the text referred to was not visible to them without ‘double-scrolling’.

Figure 12 Double scrolling for student R0007



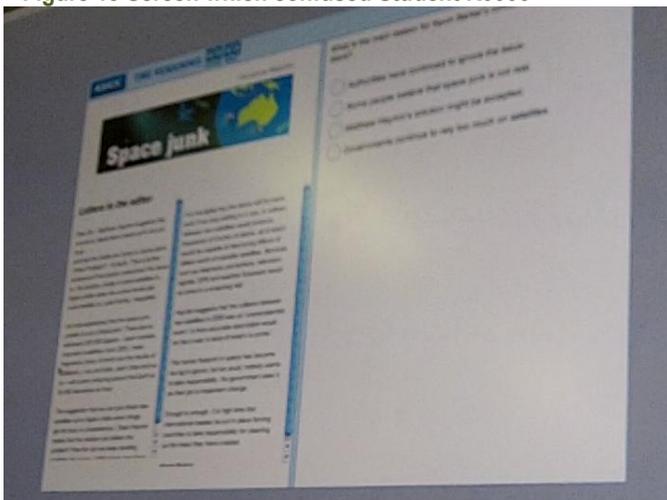
In the first image, part of the last paragraph is not visible on screen even though the student has used the central scroll bar to move the text up as far as they are able. With help from the interviewer they managed to display the last part of the text by using the right hand scroll bar – as shown in the second image.

While instances of double-scrolling were relatively few, it constituted a frustration for students and a hindrance to the testing process, and ideally should be avoided.

(R0034 Y7 Free ADF) This one was more inconvenient because of size and scrolling; two scroll bars are really annoying.

Some texts laid out in columns appeared with their own distinct scroll bars. Student R0006 (Free ABC, doing the test on a huge screen desktop Macintosh) made an unprompted comment while reading *Space junk* (Year 5 testlet B) that he was confused by each column of the text having a separate scroll bar. This led him to interpret the text as two separate letters on the topic of space junk. When the observer suggested it was probably laid out that way to look like a newspaper he still found it confusing and said that if it was one letter then he would have found it much easier to read in one column across the page.

Figure 13 Screen which confused Student R0006

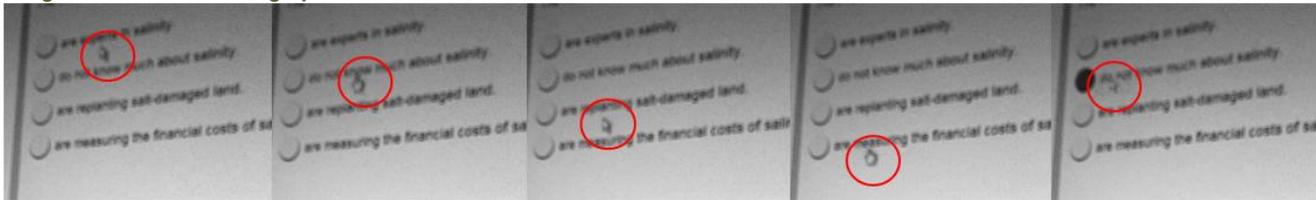


Using the cursor to aid reading

Many students used the cursor to help them find their way through the text or to focus on options as they considered an answer. Those who had a mouse found this easier to do than those using a touch pad.

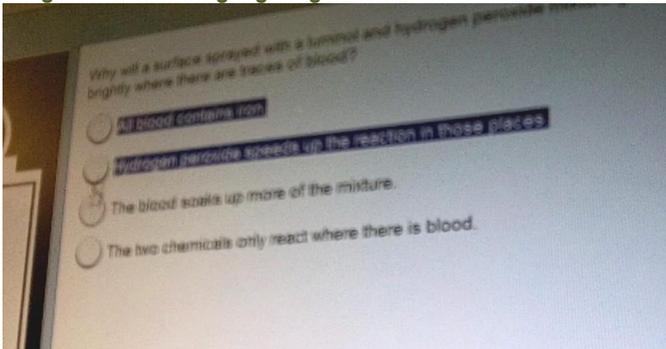
Student R0024 (Y9 Free ADE) was a very slow reader and used the cursor to track difficult parts of the stimulus text. He also tracked along each option as he considered it before selecting an answer.

Figure 14 R0024 tracking options with cursor



One student also highlighted sections of the text by mistake as he used the cursor to track over them.

Figure 15 R0028 highlighting text



CBTE: Summary

Students found many positive things in taking the test on the computer. Large numbers of students reported preferring the computer-based test on the simple grounds that it was more interesting because it was on a computer. Features such as the timer received favourable responses. Similarly, graphics were often regarded as being clearer and more colourful.

The issue of the stimulus text in Reading is a major one for the computer-based test experience. Interviewers were concerned that the students' experience of the test varied significantly depending on the size and quality of their screen. Interviewers themselves found the variation in readability of texts frustrating at times. Thought needs to be given to ways in which students can receive a more consistent experience with the stimulus texts in Reading.

Theme: Engagement

Both the mode of delivery and the improved targeting of items may improve student engagement with the test. Student engagement was studied in a number of ways. During the concurrent phase of the interviews students were asked to rate items sets on the engagement/interest level. Additionally, during the retrospective phase students were asked several questions designed to evaluate the extent to which they had been aware of the branching process.

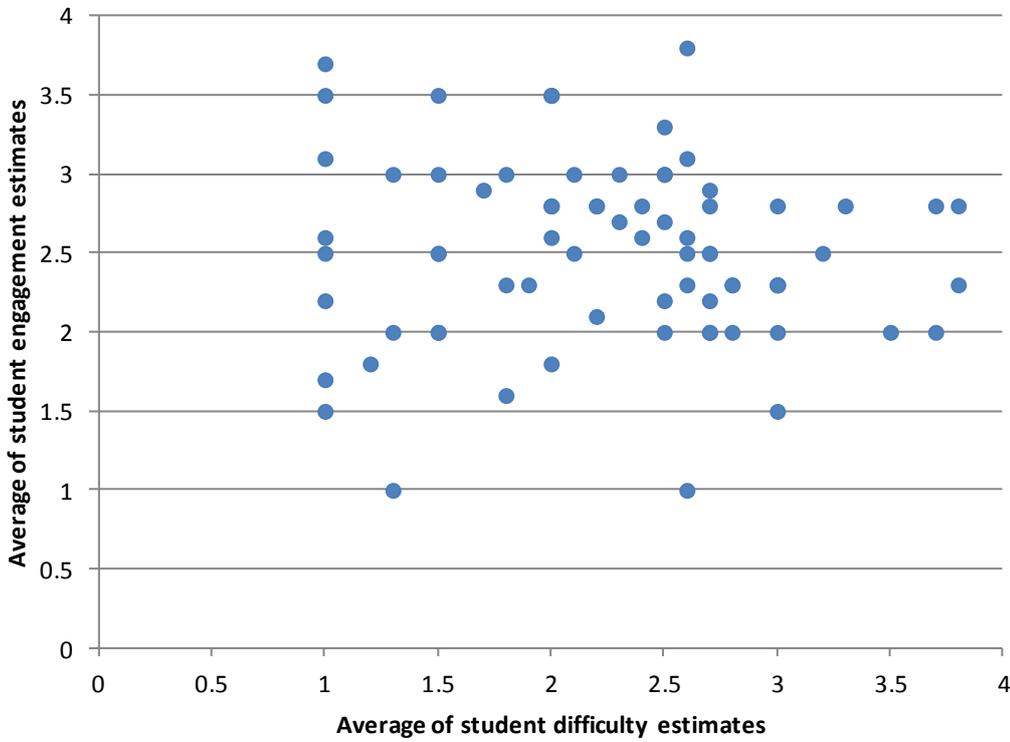
For many students the novelty of taking a test online contributed to their engagement and perception of the NASOP test as being 'more interesting than (ordinary) NAPLAN'. Several spoke of being more relaxed taking the test in this way and suggested this test had a greater variety of texts. Other positive aspects of the online experience included being able to change answers without rubbing out, being able to type more easily and neatly than handwriting, their hand not getting sore from colouring bubbles. Several commented on the benefit of focussing on one question at a time rather than being distracted by being able to see all the questions at once.

Concurrent data

At the end of each item set students were asked to rate the whole set on a scale of 1 to 4 (with 4 being the very interesting end of the scale).

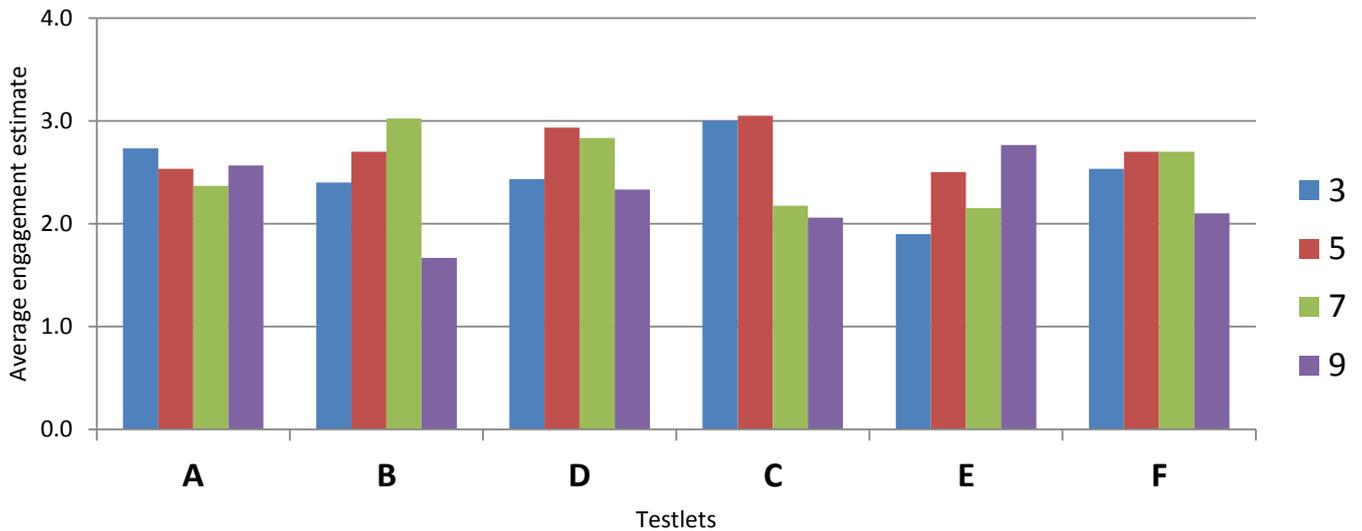
Overall there was no particular relation between the average ratings given by student for difficulty and the ratings for engagement.

Figure 16 Scatterplot of averages of student ratings for difficulty and engagement



Similarly there was no clear relation between testlets and student engagement ratings.

Figure 17 Average engagement ratings by year level and testlet



No clear picture emerges from this numerical data but questioning by interviewers is suggestive of some link between the testlet structure and engagement. Some students reported enjoying being ‘challenged’ by testlet F. It is hard to say if this is a genuine response, or one that they think puts them in a positive light or that the interviewer would like to hear. These students had a fairly clear sense of being able. However a Year 9 student who encountered F on a fixed pathway felt out of her depth and completely disengaged by the difficulty of the texts.

Students who naturally fell on the ABC pathway reported enjoying being able to answer questions at the end of the test.

Retrospective data

Question 2 in the Retrospective Interview phase asked students to compare their level of interest between the NASOP test and NAPLAN.

Q2: How interesting or engaging did you find the test? Was it more engaging than NAPLAN, less engaging or about the same?

Table 19 Reported engagement

How engaging?	Reading
Less than NAPLAN	2
More than NAPLAN	25
About the same	13
Other	1
Total	41

Many comments from students suggested that the test was engaging simply because it was on a computer. This halo-effect of computers appears to have had a broad positive effect on engagement.

(R0006 Y5) This had more interesting texts and answers, more interesting than on paper.

(R0005 Y3) This is more interesting than NAPLAN, with this one I liked to read on computer. You can change your answers easier.

(R0040 Y5) This was quite a bit more interesting. I haven't done a test on computer before. I wanted to see how it worked. It's interesting to see it can be done on computer as well as on paper.

(R0039 Y5) Most tests are on paper – this is different and fun.

(R0039 Y5) Doing it on computer brought a new level, it's very interesting, it's different, it helps the environment because it saves paper.

(R0024 Y9) It's more interesting than paper NAPLAN, it's more focussed, it's different.

No student cited issues that were clearly related to either the targeting or the branching aspects of the test. However, this does not mean that the improved targeting of the tailored test did not generate an overall improvement in the level of engagement.

Students were asked in Question 7 which of the three parts (i.e. testlets) they enjoyed the most.

Q7: Which of the three parts did you enjoy the most?

Many students had no particular preference and final testlets were not particularly more popular.

Table 20 Preferred testlet

Part enjoyed most	Reading
1. A	7
2. B/D	6
3. F/E/C	9
No preference	19
Total	41

Examining the same data but by specific testlet reveals no particular pattern.

Table 21 Preferred testlet (percentage of column)

Preferred testlet	Reading
A	32%
B	9%
D	18%
C	9%
E	9%
F	23%
Total	100%

Of particular interest was whether testlets C and F had an effect on engagement. Consequently students on ABC and ADF pathways were asked an additional question.

Q8: For testlet C: Did the easiness of the last part make a difference to how you liked the test/this part of the test?

Q8: For testlet F: Did the difficulty of the last part make a difference to how you liked the test/this part of the test?

When asked this specific question students gave some indication that the targeting of the final testlet may have had some effect.

Table 22 Effect of difficulty of third testlet

Make a difference?	Reading
Ambivalent	2
No	4
Yes	24
Total	30

Of those students who responded that the difficulty of the final testlet had made a difference, about two-thirds saw it as having a positive effect.

Table 23 Effect of difficulty of third testlet on preference

Positive or Negative	Reading
Negative	5
Positive	19
Total	24

Testlet F and engagement

Several students made comments relating their level of engagement during testlet F.

(R0040 Y5 Fixed ADF) I thought it was good to have it (F) hard. If it was all really easy I would not be as interested in doing it. NAPLAN was not as interesting as this, it was mainly easy. This one had quite a lot of hard questions, that's what made it most interesting for me.

One very reflective student commented that what engaged her is:

(R0027 Y7 Free ADF) When the text is accessible and I can understand what is going on and the questions are hard and thought-provoking.

An able Year 3 student did single out difficulty as contributing to her engagement:

(R0012 Free ADF) From A to D the writing was more engaging; the texts were longer and had more difficult words. From D to F was a huge difference, the words were quite a bit harder (but not too challenging).

Other notable quotes included:

(R0018 Y7 Fixed ADF) Yes I like things being difficult, it's a challenge. It's a good way to learn, it makes me think.

(R0039 Y5 Free ADF) I don't like things being easier, I like having a challenge, life is challenging and that section was challenging.

(R0029 Y7 Free ADF) I liked doing F texts because they were a bit harder for me. I like challenging myself. If it's too easy it feels like using up time where I could be doing something else. These texts were more like the ones towards the back of the paper, there were more of those sorts of texts.

(R0013 Y3 Free ADF) F was a bit different; you don't want to do the same thing all the time. F was a bit more challenging, that made me enjoy it more.

(R0025 Y7 Free ADF) Hard is interesting in my view. I don't agree that there were more hard texts in this than in the paper NAPLAN test. I felt quite comfortable with F; I'm used to doing hard stuff. [This despite a high error rate in F]

(R0007 Y5 Free ADF) In the last few, the texts were getting larger and some of the questions were a bit more complex than the ones at the start... I liked the third part because it was more challenging for me.

(R0026 Y7 Free ADF) Even though (F is) hardest it was still OK so it didn't make any difference to how I liked the test.

However students who reached testlet F on a fixed pathway did not always react positively:

(R0016 Y3 Fixed ADF) Observer:... how do you feel about that? Is it good, bad, middle?

Student: Um, bad.

Observer: Why is it bad?

Student: Because I want to be able to understand it.

Testlet C and engagement

Several students who reached the lowest-difficulty terminal testlet C reported increased engagement. Interestingly, some very weak students who were disengaged from the very beginning of the test found testlet C sufficiently easy that they could not resist engaging with its easiest text and items.

(R0004 Y5 Fixed ABC) Observer: [on first item of testlet C] How much easier than the last one is it?

Student: A LOT easier.

Observer: So you noticed that it was much easier than the last one. Is that a good thing, a bad thing, does it not matter?

Student: <laughs> A good thing... Because it's easier we can read it and take our time...

Observer: So because it's easier you don't feel rushed?

Student: <laughs> Yeah, we feel comfortable.

(R0032 Y9 Fixed ABC) Observer: So which of these three parts did you like the best?

Student: I'd say I enjoyed... the email one, the final test... It was really simple and I was able to easily read the text... and when I got to the questions, I knew where to go straightaway.

(R0031 Y7 Fixed ABC) See how I went from the poem [final text testlet A] to the surfer one [middle text Testlet B] and then I went to the Earth one [initial text testlet C]? I was like surprised it [Earth] was so easy, and it made it more interesting I guess.

(R0031 Y9 Free ABC) Observer: So what did you think about that [going from hard text at the end of testlet B to very easy text at beginning of testlet C]? Was it good, or bad, or weird, or what...?

Student: I thought it was good, because after the GPS one [terminal text testlet B], I was bored, and it got me interested again.

(R0006 Y5 Fixed ABC) I liked that at the end got easier. I felt better getting to C. (I liked C best) because most of the texts were easy and I knew most of the answers. It wasn't like the other sections with words that I didn't get some that I didn't know.

(R0011 Y5 Free ABC) I liked it [Testlet C] better. You could memorise the answers.

Even amongst students reporting increased engagement coincident with the decreased difficulty of testlet C, two students found the texts and item sets in that testlet ‘too easy’. This may be due in part to them being assigned a fixed ABC pathway that was below their actual ability. Though some very weak students reported that certain texts and item sets were still too easy, even for them, observed engagement did not decrease markedly as a result.

*(R0010 Y9 Fixed ABC) Observer: [speaking of a text in testlet C] So why was it a 2? Why was it less interesting?
Student: It's just really basic. It was good that you didn't have to read much, but... It's probably way too easy.*

Disengagement

The difficulty of reading the texts was frustrating for some students and could easily lead to disengagement. The artificial circumstances in which they were being observed led most to persevere.

Another major contributor to disengagement from texts and questions was unfamiliar vocabulary. It was surprising to see what words proved to be stumbling blocks:

- ‘conceal’ in Year 3 *Hiding in plain sight*
- ‘steps’ in Year 3 *Apple in a blanket*
- ‘lofty’ Year 5 *Lofty ambitions*.
- ‘course of action’ in Year 7 *When is a song original?*

What makes a test engaging?

Although some students may have spoken of enjoying the comfort of C or the challenge of F, the main determiner of student interest and engagement seems to be the subject matter and text-type of the individual texts themselves.

Subject matter that a student liked was most likely to have been rated as interesting, regardless of absolute difficulty and difficulty relative to the student’s ability level.

Though individual differences were observed in what texts students found ‘interesting’, it was text type and subject matter, rather than computer-based-testing-experience considerations, or ramifications of the ‘branching’ structure, that most determined student engagement with the test.

Reasons for liking a text were various – familiarity, learning about something new and interesting, preferring narrative/preferring factual, enjoying expositions.

(R0014 Y9 Fixed ADF) Observer: So how interesting or engaging did you find the test?

Student: I thought that parts of it were really interesting and I found other bits of it boring.

Observer: So what did you find interesting and what did you find boring and why?

Student: I loved [names texts in all three testlets]...

Observer: So you liked those just because the topic of what it was talking about was interesting?

Student: Yeah, it was something that was new; a little bit different... it wasn't like your stock comprehension piece....

When something's different to what I've been reading then its engaging.

(R0029 Y7 Free ADF) In response to the question what makes a test interesting for you?: It depends if it's on a subject that I like, then I kind of enjoy it as much as you can enjoy a test.

Several students reported that texts by which they were engaged or interested were ‘easier’ for them, because it made them want to read and do the questions. Some students also reported that texts that they found too difficult were unengaging. Often, students reported as more engaging those texts in which they were personally interested, regardless of their performance on the associated items.

(R0032 Y9 Fixed ABC) Student: If I actually start reading and it's really interesting, I read the whole story and then start answering the questions.

Observer: So what is it that makes a text easy or difficult for you?

Student: The stories in NAPLAN, most of them were boring, basically, so like I didn't really get into the stories and stuff, and that probably affected the questions as well.

Observer: So for this, being able to get into a story and be interested in it makes it a little bit easier?

Student: Yeah.

(R0031 Y7 Fixed ABC) Observer: What do you think it is that for you makes a test more difficult?

Student: If you're interested in it or not.

Observer: So if you're interested in the test it makes it less difficult?

Student: Yes.

Some students reported that the presence or absence of an appropriate graphic element in the text made it more or less engaging for them. They reported that they used the graphic element as a way or interpreting what a text was about, and that the absence of an appropriate graphic element could make a text significantly more difficult to interpret, and so less engaging.

Students mentioned liking texts in which the graphic was integral to the information, for example the scientific diagrams accompanying *Salinity*. Decorative presentation was appreciated as long as it did not distract from the text or mislead.

(R0031 Y7 Fixed ABC) Observer: So what could you put in the test that would make it more interesting for you?

Student: Like, pictures, and different interesting layouts, I guess. Cos I learn with my eyes, I don't learn by reading long... pages.

(R0023 Y5 Fixed ADE) Observer: So did you notice the lack of a picture for that one?

Student: Yeah.

Observer: What do you think about it?

Student: It was a little bit harder, and it had more text, so it was a bit more confusing.

Observer: So why was it harder without the picture?

Student: With a picture you could see what they, like, mean about it.

Observer: So with a picture, it's easier to see what the story's about?

Student: Yeah.

Variety of item types

Some able students liked the variety of different question formats; less able students tended to like the support of the multiple-choice options ('I have a one-in-four chance of getting it right'; 'I don't have to think of an answer; you get ideas for the answer'). The longer constructed-response questions were more likely to cause confusion and uncertainty.

Few students seemed to give quality responses to constructed-response questions. They often quoted directly from the text, sometimes even replicating wording from the question stem as their answer. The prevailing idea seemed to be to write as little as possible.

One student recommended that for a sequencing question it would make more sense to have the numbers available to drag and drop into position rather than typing them in. That student took the test on an iPad where the keyboard had to be brought up separately for each number.

Timing

It was noticeable that the Reading interviews took significantly longer than the Numeracy interviews. While there was some difference in the two protocols which would lead to Reading interviews taking longer, there was a general feeling among interviewers that the Reading test was simply harder to complete in the given time.

Interaction with the stimulus texts may have exacerbated this issue. Four students timed out while taking the test. Being a slow reader seemed to be a problem that was compounded by the difficulty of reading the texts on-screen and the difficulty of finding one's place in the text to locate answers. Poor readers may have been disadvantaged by the online presentation of the test.

Engagement: Summary

The primary improvement in engagement appears to derive simply from the test being delivered on computer. Students were less aware, in general, of branching and targeting effects. In so far as they were aware, the effect on engagement appeared to be positive. With the exception of the mostly positive "halo effect" granted to the test experience by the computer-based mode, the single largest determiner of student interest was not text difficulty, but the subject and type of the texts themselves. Topics and writing that students liked were engaging, and those that students did not like were not. This was largely independent of any other factor, though students reported that an interesting graphic or layout for texts could also add to their engagement (or detract from it if absent).

Conclusions and recommendations

Branching/Tailored test design

No cognitive or engagement problems with the tailored test design were identified in this study. Students largely ignored shifts in difficulty and treated the test as just another test.

When the branching and targeting was drawn to students' attention, they largely regarded it positively.

Computer-based test experience

Students responded positively to computer-based delivery. Features such as the timer and presentation of a single item at a time were regarded favourably.

Restriction on navigating back to previous items was largely not regarded as an issue by students. However, teachers may regard it less favourably and students may need some preparation to adjust to tests where they cannot review their answers at the end.

The delivery of items via computer was unproblematic in this study but the presentation of stimulus texts produced unfavourable reactions due to limitations of screen size and resolution. The testing software could be set up to allow students to "full-screen" the stimulus text with a click, and then to quickly swap back to the split-screen view displaying both the text and the item. It is also advisable that a minimum text resolution and inter-line spacing be maintained for all texts.

Engagement

New item types, clearer and colourful stimulus all appear to aid engagement. Exploiting the advantages of the technology is likely to be appreciated by students.

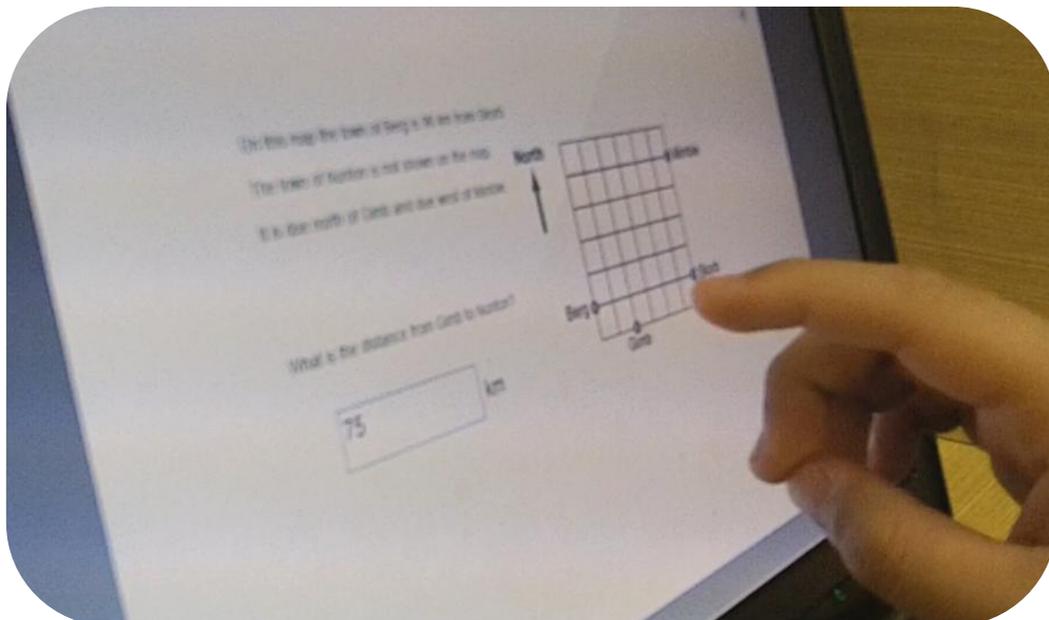


National Assessment and Surveys Online Program

NAPLAN Online Tailored Test Design August 2013

Cognitive Interviews – Numeracy and Reading

Part 3: Findings for Numeracy



Contents

Introduction.....	4
Background.....	4
The NASOP multistage model.....	4
Methodology.....	5
Protocol development.....	5
Sampling approach.....	6
Final sample.....	6
Findings.....	10
Theme: Branching.....	10
Theme: Computer-based test experience.....	14
Theme: Engagement.....	20
Conclusions and recommendations.....	22
Branching/Tailored test design.....	22
Computer-based test experience.....	22
Engagement.....	22

Tables

Table 1 Difficulty labels for pathways.....	4
Table 2 Targeting of items by method of study	6
Table 3 Actual sample by state and sector.....	6
Table 4 Proposed sample by state and location	7
Table 5 Geographical distribution of schools in final sample	7
Table 6 Students interviewed by location and sector.....	7
Table 7 Percentage of students by gender.....	8
Table 8 Percentage of students who followed a given pathway	8
Table 9 Free and fixed pathways	8
Table 10 Platforms observed	9
Table 11 Computer use	9
Table 12 Did the test differ from NAPLAN?	10
Table 13 Was the test easier or harder than NAPLAN?	10
Table 14 Easier/Harder by pathway (percentage of row)	11
Table 15 Student perception of difficulty progression by domain	11
Table 16 Difficulty progression by pathway (percentage of row).....	11
Table 17 Students noticing change between testlets.....	12
Table 18 Difference from NAPLAN: responding to items.....	14
Table 19 Assorted navigation/timing issues	14
Table 20 Mean IWG observations per student by year level and pathway	15
Table 21 Engagement	20
Table 22 Preferred testlet	20
Table 23 Effect of difficulty of third testlet	21
Table 24 Effect of difficulty of third testlet on preference	21

[NOTE: Due to rounding, some sets of percentages do not appear to total exactly 100%]

Figures

Figure 1 Pathways.....	4
Figure 2 Student moves calculator N1030 Y7 Fixed ABE.....	15
Figure 3 Graph showing the number of students by frequency of IWG observed	15
Figure 4 Mean IWG observation by year level and gender	16
Figure 5 Student using a finger on a grid while checking an answer N1030 Y7 Fixed ABE	16
Figure 6 Graph showing the number of students by frequency of scrap-paper use observed	17
Figure 7 Mean scrap-paper observations by year level and gender	17
Figure 8 Working – student N1042 Y3 Free ADE	17
Figure 9 Working – student Y5 Free ABC N1039	18
Figure 10 Working and related screen shot of space item – student N1009 Y5 Fixed ADF.....	18
Figure 11 Working – student N1034 Free ABC	18
Figure 12 Item displaying incorrectly then correcting itself N1032 Y9 Free ABC.....	19
Figure 13 Graphical options scrambled N1030 Y7 Fixed ABE	19

Introduction

Background

The National Assessment and Surveys Online Program (NASOP) is a series of projects by the Australian Curriculum Assessment and Reporting Authority (ACARA) that investigate the online delivery of ACARA's assessment programs. As part of NASOP, a trial of online reading and numeracy tests took place in August 2013.

The online tests were designed around a partially adaptive multistage test model. This model consists of a series of mini-tests (known as testlets or modules) of varying difficulty. Students follow different paths through the testlets, depending on their score in each testlet.

EAA was commissioned by ACARA to conduct cognitive interviews to investigate the impact of the multistage-branching design on students' test-taking behaviour and to gather information about students' interaction with the testlets at key branching points of the test.

This document is Part 3 of a set of three documents discussing findings from the cognitive interviews, and focuses on data from the Numeracy domain. More general information about the sampling process and comparative data between Reading and Numeracy can be found in Part 1. To enable this document to be read in isolation, some information from Part 1 is repeated in this document.

The NASOP multistage model

The model for the test piloted in this study is a multistage partially adaptive branching model. A test at a given year level consists of six testlets of items of known difficulty. Students were presented with three of these testlets (plus a fourth testlet of additional items).

The following diagram shows the branching and facility range of the testlets.

Figure 1 Pathways



The model has four distinct pathways:

Table 1 Difficulty labels for pathways

Difficulty	Sequence of testlets
High	ADF
Middle	ADE
Middle	ABE
Low	ABC

Testlets A, B and D are the branching testlets. The student's score in a branching testlet determines which testlet they proceed to next. ACARA requested that key items in the two branching testlets through which a student passed should be discussed in detail in the cognitive interviews.

Fixed-path model

While the intended model of the test involves adaptive branching, for the purpose of this trial half of the students were allocated tests with a predetermined path. This fixed-path model ensured that sufficient numbers of students undertook all of the available pathways during the trial.

Methodology

Protocol development

The priority aspects investigated in this study as requested by ACARA were:

1. the impact of the multistage-branching design on the students' test-taking behaviour
2. students' interaction with the testlets at key branching points of the test:
 - a. investigate key items in the two branching testlets which students pass through
 - b. investigate student response to items which show marked changes in difficulty
3. student performance on testlets C and F.

EAA's methodology for protocol development proceeded from priorities set by ACARA. Protocols were devised around the three interrelated themes of branching, computer-based test experience and engagement.

- **Theme: Branching**

The major innovation of the test design for this trial is the branching process and it is this process that forms the primary theme of the cognitive interview study.

- **Theme: Computer-based test experience**

The term 'Computer-based test experience' (CBTE) relates to those aspects of the test experience that are particular to a computer-based test or which are significantly affected by computer-based delivery of the test. This theme was the primary focus of the 2012 NASOP study.

- **Theme: Engagement**

Student engagement with the test may be affected by both the mode of the delivery and the branching model (e.g. through improved targeting of items to student ability). Student engagement was a main focus of this study.

This was done using four methods:

- Concurrent observation (CO): observing student behaviour during the test
- Concurrent think-aloud (CTA): asking students to 'think aloud' while answering an item
- Concurrent interview (CI): asking students questions about the item or item set they have just answered
- Retrospective interview (RI): asking students questions about items in an interview that is held after the test is complete.

The protocols were developed collaboratively by seven EAA staff members who had experience in cognitive interviewing as well as an understanding of either the NAPLAN Numeracy or NAPLAN Reading constructs. Three of those staff had specific experience with mathematics test development, cognitive interviews and NAPLAN Numeracy item development.

This table outlines the quantities of items for particular study:

Table 2 Targeting of items by method of study

Method		Number of items	Theme	Numeracy Items	
Concurrent	Observation	All	All	All	
	Think-aloud	4 (approx)	CBTE, Engagement	2: items A & B/D 2: items F/C/E + informally throughout test	
	Interview	4 (approx)	Branching, Engagement	Y3 & 5 2: final items of A & B/D 1: initial items of B/D 1: initial items of F/C/E	Y7 & 9 2: selected items A & B/D 1: selected items F/C/E
Retrospective	Interview	8 (approx)	CBTE, Engagement	Y3 & 5 4: selected items A & B/D 4: selected items F/C/E	Y7 & 9 2: final items of A & B/D 2: initial items of B/D & F/C/E 3: mid-testlet items F/C/E

Sampling approach

Full details of the sampling approach are given in the document *Part 1: Joint Summary Findings for Numeracy and Reading*.

Final sample

Schools

In total, sixteen schools participated in the cognitive interviews, approximately 40% of the schools in the proposed sample. Three factors contributed to a school's not being able to participate.

1. Non-return of permission slips. Several schools did not return any permission slips and consequently could not be included in the cognitive interviews.
2. Scheduling issues. Cognitive interviews were intended to take place on the same day as the main study and in some cases it was not possible to send interviewers to the school on the day designated by the main study contractor.
3. Several schools in the proposed sample opted out of the main study completely.

To mitigate these issues a number of the schools that had opted out completely were re-contacted to see if they would be willing to participate in the cognitive interviews only. In addition, one school (an independent Christian school in Sydney) which had not been listed in the proposed sample was included in the cognitive interviews.

Table 3 Actual sample by state and sector

State	Sector			Total	
	Cath	Gov	Ind		
NSW			9	4	13
QLD		1	2		3
Total		1	11	4	16

The proportion of NSW and Queensland schools was also affected, with a greater proportion of NSW schools included in the interviews. However, the proportion of metropolitan and regional schools was largely unaffected.

Table 4 Proposed sample by state and location

State	Location		
	Metro	Regional	Total
NSW	9	4	13
QLD	2	1	3
Total	11	5	16

Despite many schools not being able to participate in the study, the final sample still included a broad range of geographical areas.

Table 5 Geographical distribution of schools in final sample

State	Geographical distribution		
	Location	Area	Total
NSW	Metro	Sydney	6
		Campbelltown	1
		Western Sydney	2
	Regional	Blue Mountains	1
		Newcastle	1
		Orange	1
		Wollongong	1
QLD	Metro	Brisbane	2
	Regional	Rockhampton	1
Total			16

Within this sample of schools, three schools were single-sex schools. In all cases the single-sex schools were girls' schools.

Students

For Numeracy, a total 44 students were interviewed. In each of Years 3, 5 and 7 ten students were interviewed and in Year 9 fourteen students were interviewed.

Students were drawn primarily from government schools in metropolitan areas, but significant numbers were also drawn from regional locations and from independent schools.

Table 6 Students interviewed by location and sector

Location	Sector		
	Gov	Ind	Total
Metro	29	6	35
Regional	5	4	9
Total	34	10	44

In the original 2012 NASOP Numeracy cognitive interviews, approximately twice as many girls as boys were interviewed. In this year's study (2013) more girls than boys were interviewed but the proportional difference was less.

As in 2012, the main factors in the gender imbalance were return of permission notes, selection of students by the school and the presence of girls' schools in the sample without a balancing presence of boys' schools.

Table 7 Percentage of students by gender

Gender		
Female	Male	Total
61%	39%	100%

Pathways

Because of the importance of the tailored nature of the test, a sustained effort was made to interview students with a range of abilities. This was done firstly by including a range of schools in the proposed sample and secondly by contacting schools and asking them to identify students within given ability ranges. However, the final selection of students was dependent on the school selection, the return of permission notes and the availability of students on a given day.

Table 8 Percentage of students who followed a given pathway

Path		Numeracy
Low	ABC	32%
Middle	ABE	18%
	ADE	20%
High	ADF	30%
Total		100%

Pathways were also allocated as either “fixed” or “free”. With fixed pathways, students were allocated a predetermined sequence of testlets. With free pathways, the sequence of testlets was determined by the student’s score up to the branching point.

Table 9 Free and fixed pathways

Path		Numeracy
Fixed	ABC	7
	ABE	6
	ADE	2
	ADF	5
Fixed Total		20
Free	ABC	7
	ABE	2
	ADE	7
	ADF	8
Free Total		24
Total		44

Platforms and computer use

Students interviewed attempted the test on a range of different platforms. However it should be noted that to accommodate the interviews schools often allocated computers that were available in quiet locations suitable for one-on-one observation. Consequently, the platform used for the interview at a given school may not have been the same platform as was used by students in the main study at the same school.

Table 10 Platforms observed

Platform	OS	Numeracy
Desktop	Mac	4
	PC	12
Big Notebook	Mac	4
	PC	16
Small Notebook	PC	6
Tablet	iOS	2
Total		44

It had been hoped that more students would have been observed using iPads, but at two schools technical issues unrelated to the platform prevented observation of students attempting the test on iPads.

At the start of the test students were asked three questions about their familiarity with computers. The following tables show the proportion of responses for Numeracy.

Table 11 Computer use

Q1. How often do you use computers or tablets at school?			
most days	some days	only for some classes	not often
34%	23%	27%	16%
Q2. How often do you use computers or tablets at home?			
every day	sometimes	not often	never
48%	41%	9%	2%
Q3. How often do you use a mobile phone with a touch screen?			
every day	sometimes	not often	never
48%	32%	14%	7%

As can be seen from the tables the majority of the students were regular users of IT in school and at home.

Findings

Theme: Branching

The major innovation of this trial was the branching process and this process formed the primary theme of the cognitive interview study.

To see if students were aware of the branching, the first question in the Retrospective Interview asked them for any differences they noticed between the test they had just completed and NAPLAN.

Q1: One difference between the test you took today and other tests you usually take (for example NAPLAN) is that you took it on the computer rather than on paper. Did the test you took today differ from the NAPLAN tests you have taken in the past in any other ways?

Most students responded that test was in some way different from NAPLAN.

Table 12 Did the test differ from NAPLAN?

Did it differ?	Numeracy
Yes	35
No	9
Other	
Total	44

Although this question overtly excluded difference due to the mode of delivery, students generally gave responses focused on the difference between computer-based and paper-based testing. Some older students gave answers which pertained slightly to the difficulty of the test but often struggled to describe differences that were not related to computers.

(N1001Y7 Fixed ABE) The test today had a few more surprises; it didn't have as many tricks as NAPLAN, things (answers) that are obvious but are not (right); NAPLAN has a lot more tricky questions; NAPLAN have more images to help understanding; this one has more text; have to read the text many times, more than NAPLAN; spent more time to read to understand.

(N1043Y9 Free ADE) Yes, it is more difficult than NAPLAN test; in NAPLAN test, there are multiple choice questions; some questions are worth different marks; for the time, the computer test does not allow to go backward or forward; on paper you can find out the questions you can and can't do.

Some older students commented on the lack of distinct calculator and non-calculator sections:

(N1033 Y9 Free ABC) In NAPLAN there was a calculator test and a non-calculator test - 2 separate tests. In this one all mixed up. That made it easier to do - like a mix in a test.

(N1044 Y9 Free ABE) It was about the same but it went back and forth between calculator and non-calculator and that made it more difficult.

No students identified unprompted that the order of difficulty of the items was in any way unusual.

Question 3 of the retrospective interview focused more specifically on the issue of test difficulty.

Q3: How difficult did you find the test? Was it more difficult than NAPLAN, less difficult or about the same?

Table 13 Was the test easier or harder than NAPLAN?

How difficult?	Numeracy
Easier than NAPLAN	12
About the same	15
Harder than NAPLAN	12
Other	5
Total	44

In response to this question students gave more information pertaining to the progression of difficulty:

(N1005 Y3 Free ADE) Some questions more difficult than NAPLAN. It got more difficult as the test went on. Division questions make a test hard.

(N1039 Y5 Free ABC) Easier than NAPLAN; NAPLAN I struggled with questions but this one I can do it.

(N1025 Y7 Fixed ABE) Easier than NAPLAN as some of the questions were easier. With NAPLAN you know it is going to get harder. I feel better because it changed and was mixed.

(N1021 Y9 Fixed ABE) Some questions were pretty hard, harder than NAPLAN (NOTE: Student found the third part very hard).

Others provided comparisons with other software:

(N1003 Y3 Fixed ABC) Not that difficult. If I was doing MathsLab and put it on Medium. Less difficult than NAPLAN.

In general students found the test easier or about the same as NAPLAN. However a number of students found the test harder. Comparing these responses against the pathways students followed shows some relation with the difficulty of the pathway. Half of the students who said that the test was harder had followed the ADF pathway and 42% of the students who had said that it was easier had followed the ABC pathway.

Table 14 Easier/Harder by pathway (percentage of row)

How difficult?	Path				
	ABC	ABE	ADE	ADF	Total
Easier than NAPLAN	42%	17%	17%	25%	100%
About the same	33%	13%	27%	27%	100%
Harder than NAPLAN	8%	25%	17%	50%	100%
Other	60%	20%	20%	0%	100%
Total	32%	18%	20%	30%	100%

Question 4 probed the issue of the progression of difficulty further.

Q4: Overall did the test seem to get harder as you worked through it, or easier or was it more mixed?

In general, students thought the difficulty progression was more mixed than any general progression from easy to hard or vice versa.

Table 15 Student perception of difficulty progression by domain

Difficulty change	Numeracy
Got easier	3
Got harder	14
Mixed	27
Total	44

All pathways received some ‘Mixed’ responses but overall there was some relation between the pathway and the perception of difficulty progression.

Table 16 Difficulty progression by pathway (percentage of row)

Difficulty change	Path				
	ABC	ABE	ADE	ADF	Total
Got easier	67%	33%	0%	0%	100%
Got harder	14%	7%	29%	50%	100%
Mixed	37%	22%	19%	22%	100%
Total	32%	18%	20%	30%	100%

ABC pathway:

Many students who went through ABC pathway (fixed or free branching) recognised the overall changes in difficulty as general decrease in difficulty. Almost all students on this pathway showed visible relief when getting into C testlet. Several students commented on how easy the questions were. Most of them agreed that the easiness of last part made them feel better about the test. Year 7 and Year 9 students generally did not notice the existence of branching points.

(N1035 Y9 Free ABC) Easy then harder and then easier again. Middle bit got harder.

(N1006 Y5 Fixed ABC) First was easy, second one - some were hard, then the third one was easier.

(N1007Y5 Fixed ABC) It seemed to get harder in the middle test and then easier again for the last one. Last one not as hard as middle one.

ABE and ADE pathways:

Some students on both pathways recognised that the test got harder as working through it. Some felt the questions getting harder and harder.

(N1001 Y7 Fixed ABE) It's kinda like a wave, up and down all the time; it's definitely getting harder

(N1005 Y3 Free ADE) It got harder and as it got harder didn't understand what some of the questions were asking

(N1044 Y9 Free ABE) Around the middle half it gradually got harder.

(N1041Y9 Free ADE) [It went from] easy to hard; I like hard questions but not very very hard ones.

ADF pathway:

Many students going through ADF pathway (fixed or free branching) perceived the overall changes in difficulty as progressive with 'up and down' in the middle.

(N1002 Y7 Free ADF) At the very start most questions are easy, they throw in a couple of hard ones, but I did feel it gradually went harder, bit up and down in the middle

(N1010 Y3 Free ADF) A bit mixed up. At the beginning-easy and then got harder. Part 2 was mixed up while Part 3 had a few easy ones but also a few hard ones.

(N1013 Y9 Free ADF) There are some hard ones every now and then, but they definitely get harder in the end; I just guessed most of them. Also there's no multiple choice questions towards the end, that makes it hard.

The branching points between testlets were not consciously recognised by Year 7 and Year 9 students and some students felt the test was in two parts (easy and hard) instead of three.

(N1013 Y9 Free ADF) Can't really tell there are three parts, looks more like two parts to me, the easy and the hard.

Two students in Year 5 were allowed to progress to the fourth testlet because they were quite fast and keen to continue. Both students found it more difficult than previous testlets.

The sequence of questioning on this issue was designed so that students would receive an increasing amount of prompting on the issue of branching. By question 5, interviewers had explained to the students that the test was intended to be in three parts of varying difficulty. Students were asked whether they had noticed the changes.

Q5: Did you actually notice the test changing in any way when you moved from part 1 to part 2? Did you notice the test changing in any way when you went from part 2 to part 3?

Not all students responded to this question but this table summarises the responses of the students who did.

Table 17 Students noticing change between testlets

	Notice change from 1 to 2?	Notice change from 2 to 3?
Ambivalent	18	16
No	12	12
Yes	12	14
Total	42	42

(N1005 Y3 Free ADE) Each part was harder than the one before. Yes it was noticed.

(N1006 Y5 Fixed ABC) Yes, the questions got harder in the second one then easier in the third one.

(N1001 Y7 Fixed ABE –response to Q5 Retrospective) It was more like two parts; easy beginning, halfway gets harder.

(N1044 Y9 Free ABE) The type of maths changed when moving between parts. It went from problem solving. It wasn't that noticeable - about half and half in terms of difficulty.

Branching: Summary

Many students could perceive the difficulty progression of items through different pathways. However, interviewers felt that students primarily noticed this in retrospect. When students were asked initially about any differences they may have noticed between NAPLAN and the online tailored tests, no students identified difficulty progression as an issue.

Student responses to direct questions regarding the difficulty of items around branching points suggest that they were aware of significant shifts in difficulty of items around branching points. It is notable that students generally did not seem to regard this as a significant issue or something that distracted them from their test-taking experience.

Theme: Computer-based test experience

The term ‘Computer-based test experience’ (CBTE) relates to those aspects of the test experience that are particular to a computer-based test or which are significantly affected by computer-based delivery of the test. This theme was explored in a number of ways during the Concurrent and Retrospective phases of the interview. In general, unsolicited comments from students were more likely to pertain to technical issues or general computer-based experience issues than to the other two themes in the study.

When asked about the differences between the test they had just sat and NAPLAN, 14 students gave responses pertaining to the ease or difficulty of entering their responses. These responses could be roughly divided into three categories.

Table 18 Difference from NAPLAN: responding to items

Entering data	Numeracy
Easier to correct	2
Easier to respond	11
Entry mistakes more likely	1
Total	14

‘Easier to correct’ responses included comments such as:

(N1027 Y7 Fixed ABC) In paper (test) you have to write more often, sometimes if you make mistake, you can't rub it out you could still see the dot, then you come back to the question you see the dot you think yes I have finished question when you didn't.

‘Easier to respond’ responses included comments such as:

(N1030 Y7 Fixed ABE) Computer was faster; less writing, typing the answer was quicker; I prefer using computer; feel more focused on computer because less distraction; computer can have disadvantage as well, like graphics not displaying properly.

The third category covered this response from a student concerned that they may press the wrong button:

(N1040 Y5 Free ADE) With computers, sometimes you can't work it out properly; can get confused. Might press wrong thing, paper is easier to work out on.

Other CBTE issues raised related to some of the navigation and timing features of the test.

Table 19 Assorted navigation/timing issues

Navigation	Numeracy
Can't go back	2
Isolated items	1
Timer	2
Total	5

In Year 7 and Year 9 Numeracy no back button was available because calculator and non-calculator items were interspersed. The two students in Numeracy who were unhappy about navigation were in Year 9.

(N1017 Y9 Fixed ADF) NAPLAN is able to look back, it usually recommends us to look at the whole paper first I think it's a better way to take a test because you know that you got more right at the end instead of stuck half way; the risk of computer loading the next question can be a problem; font on paper text is usually small; computer screen text and image are bigger.

(N1043 Y9 Free ADE) Yes, it is more difficult than NAPLAN test; in the NAPLAN test, there are multiple-choice questions; some questions are worth different marks... The computer test does not allow to go backward or forward; on paper you can find out the questions you can and can't do.

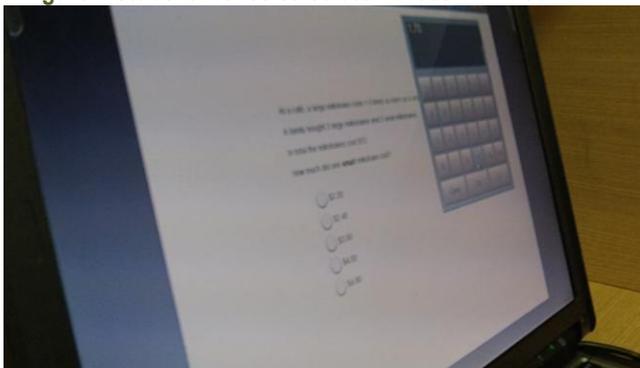
Apart from those two students, the restricted navigation was not a major issue among students.

‘Isolated items’ relates to responses from students (though primarily in Reading rather than Numeracy) raising the issue that in the computer-based tests each item is presented in isolation from the other items.

Those students who mentioned the presence of the timer generally regarded it as positive feature.

Many students at various stages of the Numeracy interview responded positively to the on-screen calculator. Many students were comfortable moving the calculator to more convenient positions.

Figure 2 Student moves calculator N1030 Y7 Fixed ABE



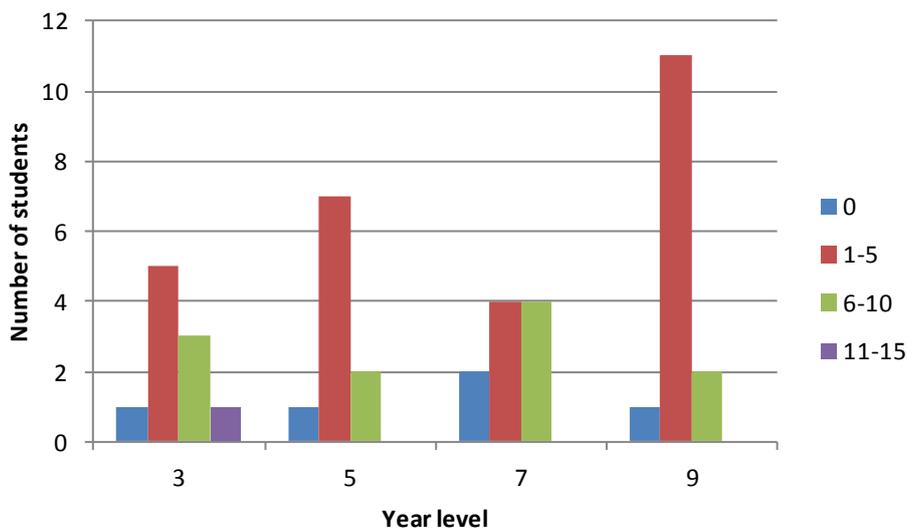
Many students made positive comments pertaining to the presentation of items. Students responded favourably to larger graphics and occasional use of colour.

Interaction with graphics (IWG)

“Interaction with graphics” refers to how students respond to graphics and diagrams presented on-screen. Interaction with graphics was observed in all year levels, but mainly for questions involving maps, direction, measure of distance and certain graphs. Some students moved the cursor over the text as they read. However much of the time students read the screen and selected their response without overt interaction with graphical elements.

The following graph shows the number of students in each year level by the amount of graphical interaction observed during the tests. The groupings refer to the number of items observed.

Figure 3 Graph showing the number of students by frequency of IWG observed



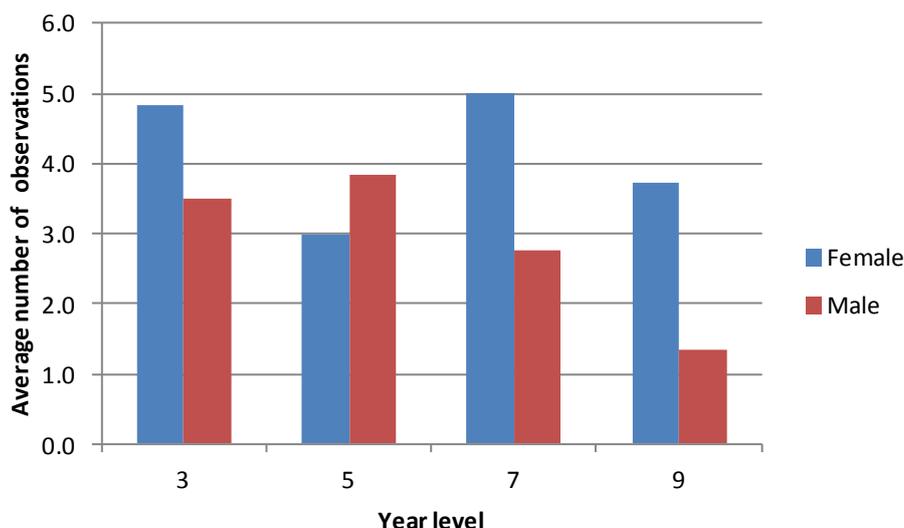
The greatest amount of interaction with graphics was observed in Year 7 but this may be primarily due to the choice of items in the Year 7 testlets. One Year 3 student used the mouse or finger to count or indicate many objects on screen, leading to 13 observed items with some interaction with graphics. However, most students interacted overtly only with a small number of items.

Table 20 Mean IWG observations per student by year level and pathway

Year	Path				Total
	ABC	ABE	ADE	ADF	
Year 3	1.0	n/a	5.8	3.8	4.3
Year 5	4.2	n/a	3.0	2.5	3.5
Year 7	4.0	1.8	7.0	5.8	4.1
Year 9	3.3	3.8	0.0	3.5	3.2
Total	3.5	2.8	4.3	4.2	3.7

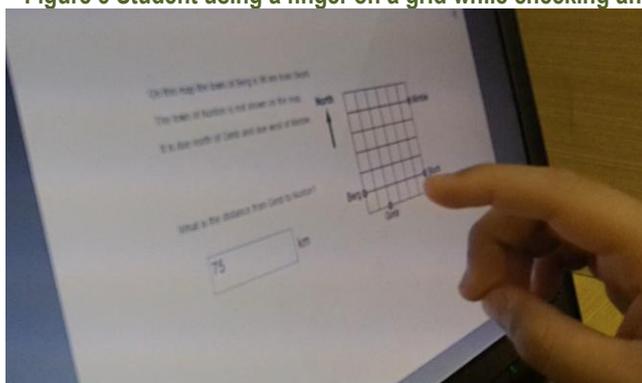
There was some variation by gender but given that sample was not well-balanced with regard to gender this should be treated with caution.

Figure 4 Mean IWG observation by year level and gender



In general, the range of responses was similar to the previous study. Students used cursors, fingers or pens to count objects on screen or to follow particular lines or objects.

Figure 5 Student using a finger on a grid while checking an answer N1030 Y7 Fixed ABE



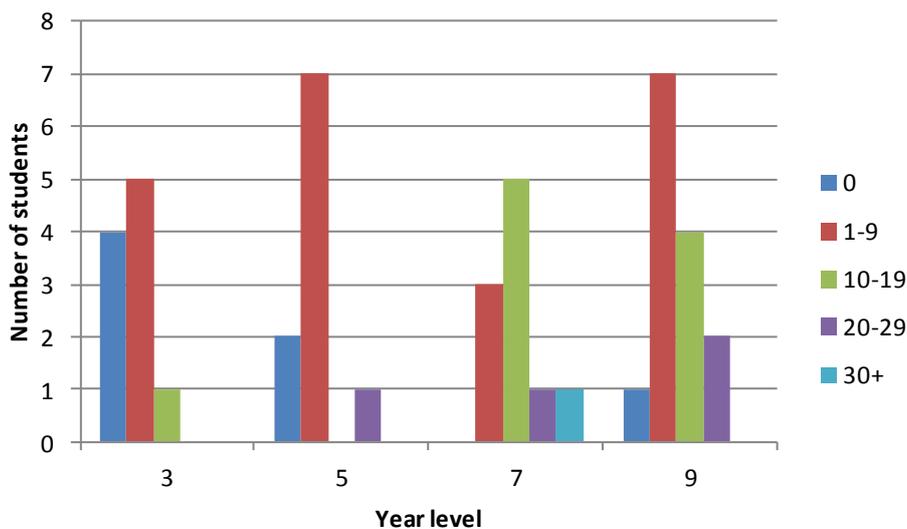
The presentation of diagrams on screen did not present a significant obstacle to students.

The scrap-paper problem

The “scrap-paper problem” relates to some students feeling that they cannot use paper to help work out problems when the problem is presented on-screen. Scrap-paper was used to solve various types of questions. Some high-ability students (especially Year 3 and 5) were able to work out many questions mentally, whereas others (usually medium and low-ability students) would quickly pick an answer by skimming questions superficially without further investigation. Lower ability students tended to use scrap-paper to work out simple calculations.

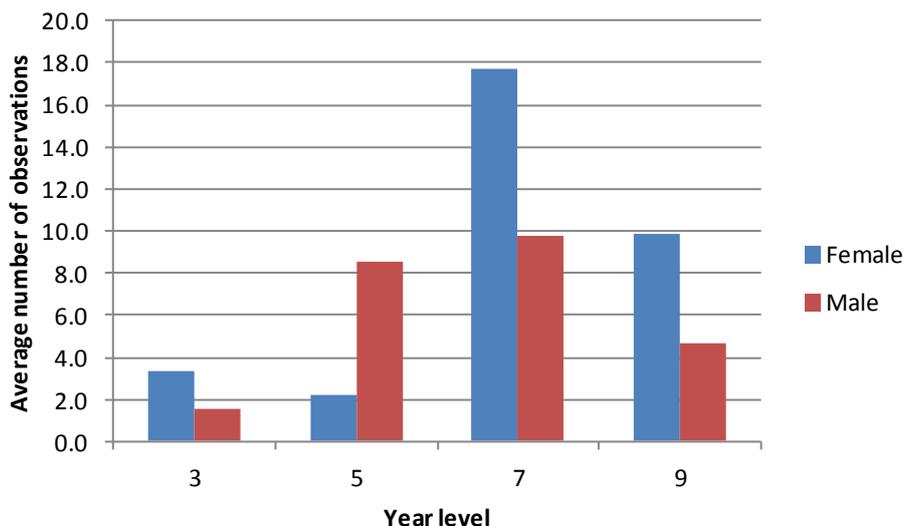
The following graph shows the number of students in each year level by the frequency with which they used paper for additional working. The groupings refer to the number of items observed.

Figure 6 Graph showing the number of students by frequency of scrap-paper use observed



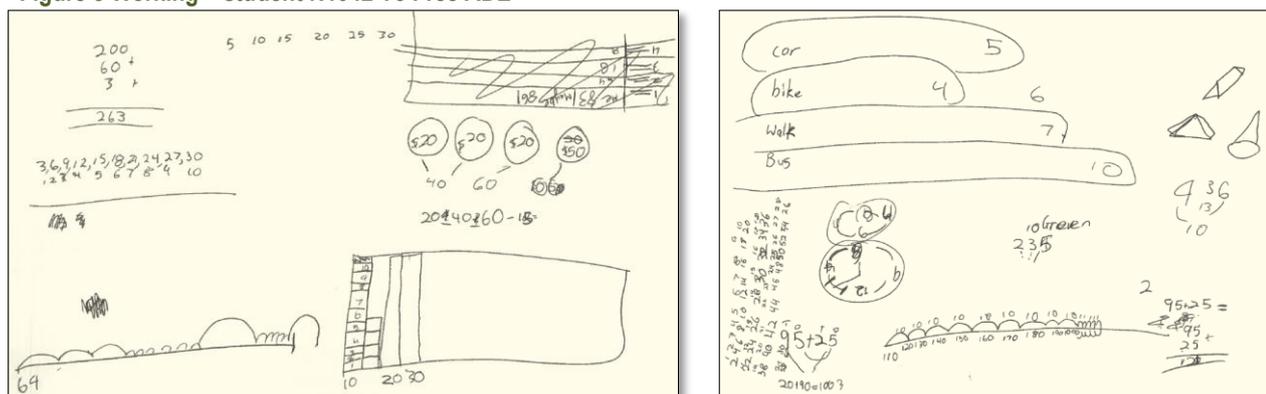
Scrap-paper use was lowest in Year 3 but very variable across all year levels. The students who used scrap-paper most frequently were in Year 7. There was some variation across gender also.

Figure 7 Mean scrap-paper observations by year level and gender



In some cases working involved not only arithmetic but also graphical elements to help with calculations and drawings of shapes and objects.

Figure 8 Working – student N1042 Y3 Free ADE



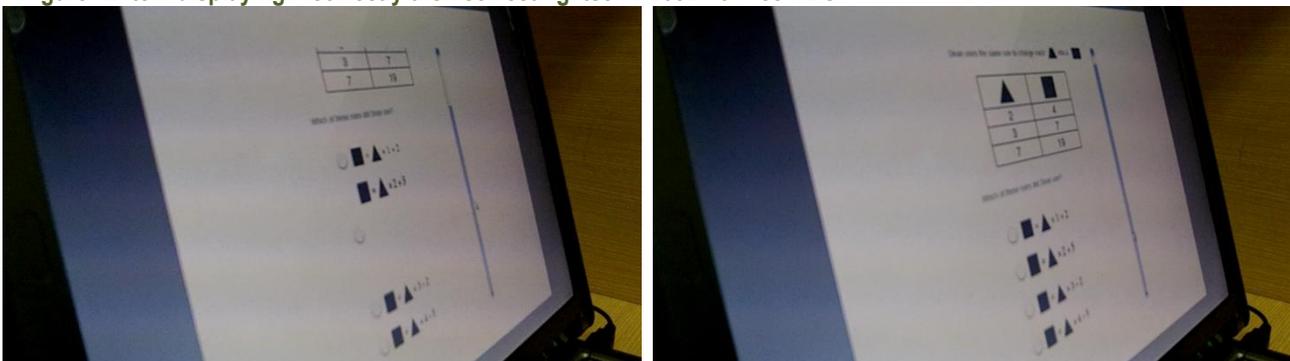
Technical issues

This study was not intended to be an evaluation of the test-delivery platform, but when considering the computer-based test experience, technical problems have to be considered.

The primary technical issues encountered by interviewees and students related to logging into the system. In some cases this was often just an issue of finding the right way to turn off pop-up blockers. In other cases logging-in was delayed by network issues either at the school or at the server side of the system. In general, students reacted to such issues stoically, presumably because technical issues are a not-uncommon feature of using computers in a school environment. On occasions when students had to close their browser because of technical problems, the system typically returned students to the last item they had attempted when they logged back in. This feature helped mitigate the effect of technical glitches on the student's experience of the test.

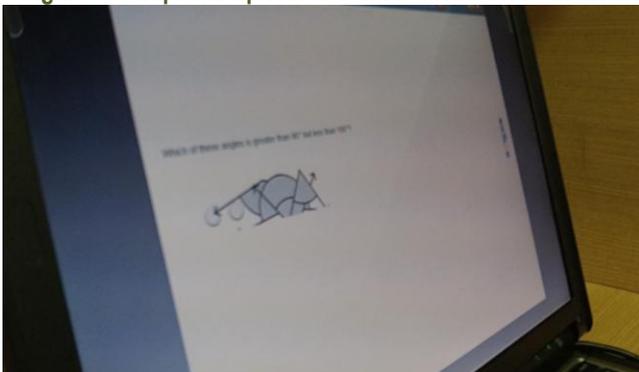
More generally, variations in hardware and software meant that student experiences of the test were highly variable. In the case of the cognitive interviews, this issue was exacerbated by students being allocated computers that were 'spare' machines – often older machines with less than up-to-date software. In one case a student attempting the test on a computer with unusual screen resolution settings (which couldn't be changed due to permission restriction) identified a division symbol (\div) as an addition symbol (+). In other cases, items displayed unusually, with options poorly aligned or with additional scroll bars.

Figure 12 Item displaying incorrectly then correcting itself N1032 Y9 Free ABC



In some cases graphical options became completely jumbled.

Figure 13 Graphical options scrambled N1030 Y7 Fixed ABE



However these experiences were relatively unusual and in general items displayed consistently.

CBTE: Summary

Students found many positive aspects to taking the test on computer. Large numbers of students reported preferring the computer-based test on the simple grounds that it was more interesting because it was on a computer. Features such as the timer and the on-screen calculator received favourable responses. Similarly, graphics were often regarded as being clearer and more colourful. Some lesser presentation issues occurred but these were primarily due to incorrect settings on specific machines.

Theme: Engagement

Both the mode of delivery and the improved targeting of items may improve student engagement with the test. Interviewers monitored engagement during the Concurrent Observational phase and also asked students questions about their level of engagement during the Retrospective Interview phase.

Question 2 in the Retrospective Interview phase asked students to compare their level of interest between the NASOP test and NAPLAN.

Q2: How interesting or engaging did you find the test? Was it more engaging than NAPLAN, less engaging or about the same?

Table 21 Engagement

How interesting?	Numeracy
Less than NAPLAN	2
More than NAPLAN	24
About the same	18
Total	44

Of those who said that it was more interesting than NAPLAN, 9 students cited reasons related to simple computer delivery as being the primary reason why it was more engaging.

(N1015 Y3 Fixed ADF) More interesting on computer as I had fun typing. Questions much the same

Other students cited content-specific reasons:

(N1035 Y9 Free ABC) They were interesting questions - likes doing graphs and picture ones like shapes, also number questions with substitutions.

Others cited reasons covered in the CBTE section above, such as the items being presented one at a time:

(N1017 Y9 Fixed ADF) It is more engaging, because you always want to know what next questions are, on paper you know you got so many questions left; you feel not excited but less focus in a way. This one you know you're up to question 20, just take a risk, keep going and try the answers.

(N1001 Y7 Fixed ABE) I found it bit more engaging, because of the way it is set out... like one question on screen each time.

No student cited issues that were clearly related to either the targeting or the branching aspects of the test. However this does not mean that the improved targeting of the tailored test did not generate an overall improvement in the level of engagement.

Students were asked in Question 7 which of the three parts (i.e. testlets) they enjoyed the most.

Q7: Which of the three parts did you enjoy the most?

Many students had no particular preference and final testlets were not particularly more popular.

Table 22 Preferred testlet

Part enjoyed most	Numeracy
1. A	13
2. B/D	16
3. F/E/C	9
No preference	6
Total	44

Reasons for selecting testlet A were primarily because it was seen as easier:

(N1012 Y7 Fixed ADE) Probably the first part, because it's the easiest, everyone likes easy things; it's more interesting and enjoyable; when you spent a lot of time on one question, it's not enjoyable any more.

Even some students on the ABC pathway cited testlet A as being preferred because it was easier.

Similarly students who preferred testlets B or D often cited the balance of hard and easy items in that testlet as a reason for their preference.

(N1025 Y7 Fixed ABE) Second part. The questions were medium but they were things you could work out.

(N1030 Y7 Fixed ABE) Part 2; some questions are easy, some are hard and make you think; I liked the mixed way.

(N1041 Y3 ADE) I like the medium (part 2); it's a bit challenging.

The final testlets were the most diverse in terms of difficulty. Unsurprisingly students who preferred the third testlet on ABC pathway preferred testlet C because of its ease and students on the ADF pathway preferred testlet F because of its more challenging nature.

Of particular interest was whether testlets C and F had any effect on engagement. Consequently students on ABC and ADF pathways were asked an additional question.

Q8: For testlet C: Did the easiness of the last part make a difference to how you liked the test/this part of the test?

Q8: For testlet F: Did the difficulty of the last part make a difference to how you liked the test/this part of the test?

When asked this specific question, students gave some indication that the targeting of the final testlet may have had some effect.

Table 23 Effect of difficulty of third testlet

Make a difference?	Numeracy
Ambivalent	2
No	9
Yes	14
Total	25

Of those students who responded that the difficulty of the final testlet had made a difference, about two-thirds saw it as having a positive effect.

Table 24 Effect of difficulty of third testlet on preference

Positive or Negative	Numeracy
Negative	5
Positive	9
Total	14

However several students did cite some negative effects – primarily for testlet F:

(N1007 Y5 Fixed ABC) I actually like harder tests better so the easiness didn't make me like it more.

(N1013 Y9 Free ADF) There was no multiple choice towards the end, that put a lot of pressure on me, it made me feel pretty dull.

(N1016 Y7 Fixed ADF) I felt good at the beginning, thinking most tests get harder but this seems to be easy; the questions are supposed to be hard but not too hard, but really got stuck towards the end. They are bit too hard for me.

(N1017 Y9 Fixed ADF) If it's continued to be hard, you got to look at your time management, if you got to spend to spend too much time on this question, you better just guess and continue on.

(N1038 Y7 Free ADF) When it's hard, I do feel bad; when it's hard, you look at the questions and have mental block.

Features of individual items that students liked included:

- questions with interesting/bright graphics
- questions about food
- graphs (although these were also actively disliked by some students)
- calculator allowed questions.

Features of individual items that they disliked included:

- equations and formulas
- large number
- patterns (harder examples)
- fractions.

There was no clear data on item difficulty and engagement. This is perhaps unsurprising as it is likely that the most engaging items are those that are best targeted to the student's level of ability.

Engagement: Summary

The primary improvement in engagement appears to derive simply from the test being delivered on computer. Students were less aware, in general, of branching and targeting effects. In so far as they were aware, the effect on engagement appeared to be positive. Larger graphics and the occasional use of colour seem to have helped engagement also. Technical problems were the primary cause of disengagement.

Conclusions and recommendations

Branching/Tailored test design

No cognitive or engagement problems with the tailored test design were identified in this study. Students largely ignored shifts in difficulty and treated the test as just another test.

When the branching and targeting was drawn to students' attention, they largely regarded it positively.

Computer-based test experience

Students responded positively to computer-based delivery. Features such as the timer, on-screen calculator and presentation of a single item at a time were regarded favourably.

Restriction on navigating back to previous items was largely not regarded as an issue by students. However teachers may regard it less favourably and students may need some preparation to adjust to tests where they cannot review their answers at the end.

Students are increasingly solving mathematical problems in on-line environments. Students need to be aware that cognitive aids such as working-out paper remain important regardless of the mode of delivery.

Engagement

New item types, clearer and colourful stimulus all appear to aid engagement. Exploiting the advantages of the technology is likely to be appreciated by students.