

Summary Report:

**Digital External Assessment
Prototypes Project**

Contents

Executive Summary	3
Background	3
Project objectives.....	3
Key findings.....	4
Next steps	4
Project approach and evaluation methodology	6
Pilot components	6
Conducting the assessment	8
eMCAT marking.....	9
OPE marking.....	10
Detailed findings	11
eMCAT	11
Device usage	14
Operating systems	14
Results by ethnicity	15
User insights.....	19
Assessment learnings.....	22
Process and technology learnings.....	22

Executive Summary

Background

The Digital Assessment Transformation (DAT) workstream was established as part of the Future State Programme to help transition NZQA's current paper-based National Certificate of Educational Achievement (NCEA) assessment model to an online delivery model by the year 2020.

As a first step towards this, NZQA has undertaken trials and pilots¹ to help understand how to transition into delivering assessments digitally.² This includes testing key elements of a new operating model, and learning how that model might work in practice.

In 2015, a Digital External Assessment Prototypes (DEAP) project was undertaken, which built on learnings from a computer-based eMCAT trial undertaken in 2014. This document is focussed on the 2015 DEAP trial.

Project objectives

The objectives of the DEAP project were:

- to trial enhanced and new technology approaches for the electronic Mathematics Common Assessment Task (eMCAT)³ that built on the learnings from the 2014 trial
- to test different aspects of creating and delivering online examinations through the practice examinations
- to test logistical elements of the online examinations that will ultimately support a new model for delivering digital examinations online
- to better understand the user experience, through robust evaluation including feedback from students, schools, internal business units, and the vendor
- to help inform the end-to-end process of a new operating model.

¹ A "trial" is considered to be a practice assessment that does not count towards a student's NCEA, whereas a "pilot" is considered to be a practice assessment where a student can gain credits towards their NCEA.

² A digital or online assessment is considered to be an assessment conducted in a digital format through the use of an electronic device.

³ The "MCAT" is the Mathematics Common Assessment Task, with is an external assessment for Achievement Standard 91027 – *Apply algebraic procedures in solving problems*.

Key findings

Technology

- Seventy-six percent of students who took part in the DEAP projects used a Microsoft Windows environment.

Creation and delivery

- The range of performance results of the 2015 eMCAT was very similar to the 2014 MCAT. However, a large percentage of students who responded to the survey (which was one third of those who sat the eMCAT) reported that they still used pen and paper during the eMCAT in some way (i.e. for note taking or working out their answers before entering their final answers digitally).
- The data indicated that 5-6% of students who participated in the eMCAT achieved higher grades of Merit and Excellence over non-participating students. While 58% of students gained the same result in both assessments, 26% gained their final result from the paper-based MCAT.
- Although students are familiar with the online environment and are used to working in this context, work needs to be done to ensure participants confidence in any online assessment.

User experience

- *eMCAT* - most of the teachers involved in the eMCAT who responded to the survey reported that administering the eMCAT was a positive experience. This contrasts with the eMCAT students who completed the survey, who indicated undertaking the online assessment was a less satisfactory experience than undertaking the paper-based assessment.
- *French Online Practice Exam (OPE)* - most participants (both teachers and students) in the French OPE who completed the survey reported a positive experience with the assessment.
- *Science OPE* - the majority of participants in the Science OPE who completed the survey indicated that undertaking the online assessment was a less satisfactory experience than undertaking the paper-based assessment.

Next steps

The learnings from all the trials and pilots undertaken by NZQA in 2015 will be used to inform development and delivery of trials and pilots planned for 2016. Specific learnings to be carried forward from the DEAP project include the following.

Technology

- Any security assessments should be completed well in advance of the assessment so that adequate time is available to resolve any issues.
- NZQA needs to develop a broader set of technical specifications and make these available to schools

Creation and delivery

- Future trials and pilots should focus on incorporating a digital environment that enables participants to note-take and work on their answers, before submitting their final answer.
- Touchsreen technology *may* be important.
- A wider selection of subjects should be trialled so that there is greater testing of experience of online assessment within the sector. In particular, it is recommended that trial OPEs be developed for other languages.
- Students taking languages may be more familiar with digital pedagogy than those taking maths and/or sciences.

Logistics

- We need to gain a profile of various devices and operating systems being used (i.e. how many students use their personal laptops or tablets, schools PCs, and to what extent touchscreens are being used).
- Consideration should be given to schools limiting access for trials and pilots to students who are familiar with digital devices and/or for where the digital device is used within the teaching learning programme.
- NZQA needs to develop a case management strategy for interaction with schools, to facilitate engagement as to the requirements and implications of being involved in any digital trials.

User experience

- More time should be taken to trial online Science examinations before they are piloted.

New operating model

Future trials and pilots should focus on the understanding and use of an appropriate end-to-end model suitable for use in a digital examination context.

Project approach and evaluation methodology

The DEAP pilot was undertaken to test NZQA's critical assumptions through:

- developing digital items for use in an online examination setting
- trialling the process of delivering and managing a digital examination
- including a sufficient cohort of students to simultaneously participate in trialling the digital assessments.

The subsequent evaluation considered the outcomes, user experiences, and insights in relation to implementation, technology, and change management. An overview of the Evaluation Framework is included in Appendix 4.

Pilot components

To progress this, NZQA worked in partnership with an external vendor, Education Perfect, to undertake a second iteration of the eMCAT (following on from the 2014 trial), and two OPEs. The project developed online assessments for the following achievement standards:

1. Level 1 Mathematics Achievement Standard (AS) 91027: *Apply algebraic procedures in solving problems*
2. Level 1 French Achievement Standard 90878: *Demonstrate understanding of a variety of spoken French texts on areas of immediate relevance*
3. Level 1 Science Achievement Standard 90940: *Demonstrate understanding of aspects of mechanics.*

Mathematics AS91027 is assessed using a common assessment task (commonly referred to as the MCAT) in which NZQA develops two assessments to be sat on one of two selected dates during a week in September - schools select the date they wish to have their candidates sit the MCAT. Schools manage the assessment, including marking the assessments, sending NZQA a sample of their marked assessments for verification, and finalising their results, which they then enter into the NZQA results database. Students who completed the eMCAT also completed the paper MCAT assessment, with the best mark from the two assessments becoming their final mark.

The French and Science OPE were developed for use as practice examinations and did not count towards a student's NCEA. Schools could use the OPE for revision, preparation for the actual examination, or for their school benchmark examinations. The OPE could also be used as evidence towards a derived grade. Schools were responsible for the marking of the OPE and reporting results to the students.

NZQA developed the assessments with external contractors. The marking of the eMCAT was also undertaken by external parties. The delivery and marking of the eMCAT was undertaken through Education Perfect's platform, with the marking undertaken by contracted NZQA markers. Education Perfect also enabled access to a range of familiarisation activities through their platform.

In 2015 Education Perfect provided NZQA with a set of technical specifications based on its testing and what it understood would work with the platform. These specifications were provided to schools. Schools were invited to register their interest in taking part in the

eMCAT and/or the OPE taking into consideration the technical specifications that were required before they registered.

Once registrations closed, schools provided Education Perfect with data about the students, classes and teachers who were going to take part. Education Perfect then created individual accounts for each participant and assigned attributes based on their role.

Participation

In total, 17,106 students from 208 secondary schools across New Zealand participated in these online assessments, with approximately 4000 students taking part in more than one online assessment.

eMCAT	11 375 students (31% of cohort)	146 schools
Science OPE	8944 students (25% of cohort)	113 schools
French OPE	1110 students (52% of cohort)	79 schools

Note four of the providers that undertook the eMCAT are not registered secondary schools (three study centres and one tertiary provider), but for simplicity, they are counted as a school in the above table (and excluded from the summary school data).

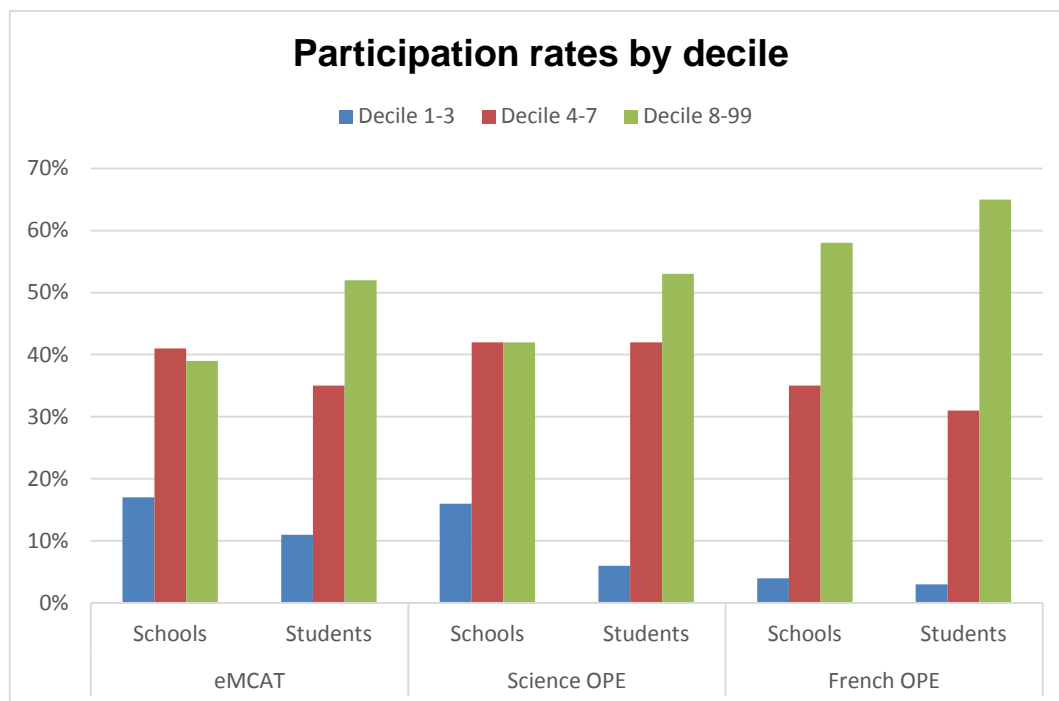
NZQA originally planned to cap the eMCAT at 5000 students in no more than 100 schools. However, registrations exceeded expectations and a decision was made to accept all schools and students, and to manage the risks associated with a scaled-up project. Risk management included the development of a second eMCAT, an additional eMCAT date, and clear communication and guidance. To maintain manageability and ensure sufficient capacity in Education Perfect's platform, NZQA assigned each school its eMCAT date (i.e. to balance the students across both dates). The dates were 8 and 10 September 2015, with schools able to set up the eMCAT between 9 am and 12 pm on their computers, or digital devices.

Summary of schools who participated

eMCAT only	65 schools
Science OPE only	34 schools
French OPE only	15 schools
All three	27 schools

eMCAT & Science OPE	30 schools
eMCAT & French OPE	20 schools
Science OPE & French OPE	17 schools
Total	208 schools

Participation rates by decile



The participation rates of schools in all three assessments in the low, middle, and high deciles, is a close approximation to the national proportion of each decile-type of school. However this changes when student participation is compared across the deciles for the eMCAT and the Science OPE, where participation rates of students in low decile schools are disproportionately lower than in the other two decile categories. It is possible that mid and high decile schools have a higher proportion of students with access to their own device for the DEAP assessment, or that the low decile schools taking part tended to be small schools whereas the high decile schools were mainly large schools. Further analysis is required to determine the reasons for this disparity.

The participation rates in the French OPE reflect the national proportion of 47% of schools offering level 1 French being high decile, and 59% of the level 1 French being from high decile schools.

Conducting the assessment

Education Perfect opened its system to students and teachers to familiarise them with the various digital tools associated with the assessment. These familiarisation activities were available from the end of July 2015 for eMCAT students, and from the beginning of September 2015 for OPE students.

Schools were responsible for conducting the eMCAT. Once the school-specific date was set and the school sent in their data, Education Perfect set up individual accounts for students to access the assessment on the set time and date. Some schools requested a variation to the set times, mainly starting or finishing half an hour earlier or later. These requests were accommodated by ensuring access to the system was open at these times.

A time buffer of 15 minutes was provided to enable schools to manage individual student issues. This meant that the programme would automatically submit the student responses at the 75 minute mark. Schools also identified students who required extra time, and Education Perfect programmed this requirement into the individual student's account.

For the OPE, schools were responsible for when and how the assessment was to be used. It was made available in two time-windows; the last two weeks of term three, and the first two weeks of term four (14–25 September 2015 and 12–23 October 2015 respectively).

The assessment was managed through a dashboard developed by Education Perfect. The dashboard provided the invigilator with the status of all students in the classroom undertaking the assessment, including whether:

- they had logged on
- they had commenced the assessment
- they had submitted the assessment
- the student had a network issue (i.e. issues with internet connection)
- the student exited the assessment and accessed part of their computer or another part of the internet, and the length of time the student spent doing so.

Anti-cheating software

Education Perfect trialled some anti-cheating software in the eMCAT. If the student “lost focus” by changing the mode, accessing a file or app on their computer (including a calculator), or by accessing the internet, a report was immediately provided to the examiner. A dialogue box also appeared to the candidate, informing them they had “lost focus” and advising them to return to the assessment.

The software was a trial version, and it was accepted that it was at the lower end of security, however it was considered as complementary to good invigilation, and it worked well with most operating systems.

Candidates were also required to undertake the assessment in full screen mode.

eMCAT marking

A marking panel consisting of a panel leader, four check markers and 25 markers was established and based in Auckland.

Although the plan was for markers to ‘strip mark’ a particular item, all were trained to mark each of the item variations in one eMCAT. This enabled the panel leader to assign markers to a particular item on a daily basis to ensure balanced progress across the eMCAT.

Education Perfect developed a marking portal for markers to access the submitted eMCATs. Markers signed on as ready to mark, inputted the item and variation they had been assigned, and were then allocated student responses to mark. The assessment schedule (marking guide) for the item and variation accompanied the candidate responses. Markers reported that they found the process simple and intuitive to use, and very quick to mark.

Note current marking processes block markers from marking their own students or family, however there are subjects in which the students' responses can be anonymised making it impossible to identify students' assessments. This enables NZQA to simplify the marking process by removing the "forbidden provider or candidate" rule when marking a digital examination, for such subjects.

The Education Perfect software provided the marker with a recommended result, but some markers reported that they didn't use it, as they preferred to read the candidate response and make a judgement themselves.

At the conclusion of marking, Education Perfect provided NZQA with a file containing the NSN of every candidate who took part, their result for every item, and the overall response. NZQA extracted the paper MCAT results and compared them with the eMCAT results. The best result was then entered into the candidate's record of achievement.

OPE marking

Schools were responsible for the marking of the French and Science OPEs.

An assessment schedule for each OPE was included in the teacher's version of the assessment. The schedule was broken down by item, and each item judgement was aligned with the specific item in the assessment. Breaking down the assessment schedule item-by-item proved useful for the French OPE and the eMCAT, but problematic for the Science OPE, as Science markers preferred to take a more holistic view of the candidate evidence, and required access to the entire assessment schedule.

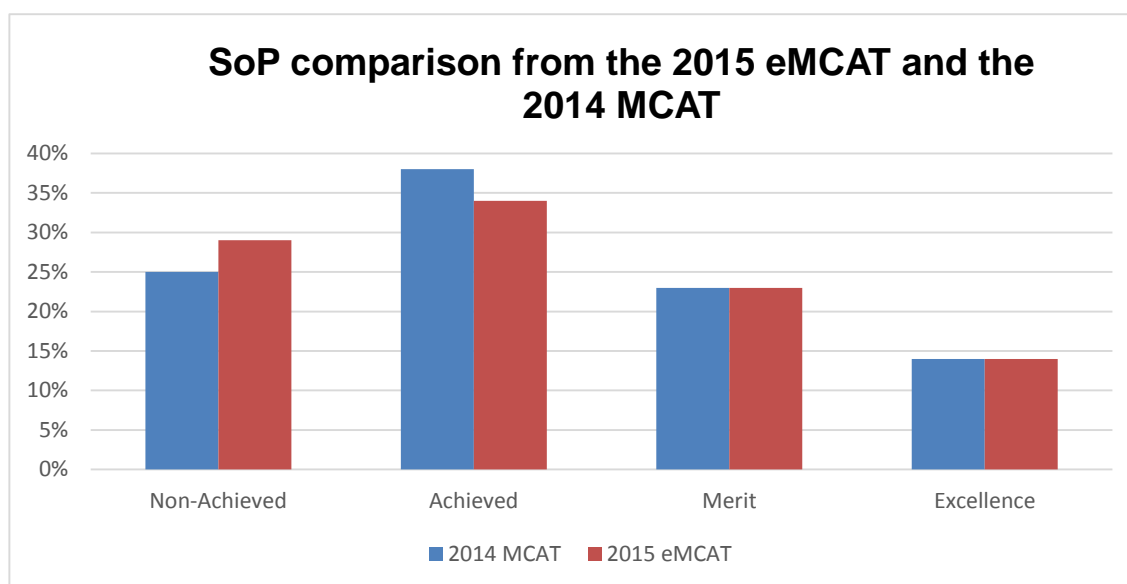
Detailed findings

The data in the first section of this Part four is drawn from the entire MCAT cohort and makes comparisons between the Spreads of Performance (SoPs or singular, SoP) of a range of assessments. The SoP is used to compare results on a year by year basis and is intended to be used as part of a series of tools to minimise inter-year variability. The SoP will demonstrate the percentage of students who gained grades of Non-Achieved (N), Achieved (A), Merit (M) or Excellence (E) from all students who sat the assessment.

*e*MCAT

The 2015 MCAT format was changed from 2014, and this format change was also reflected in the two *e*MCATs.

The graph below compares the SoP of the 2014 MCAT with the SoP of the 2015 *e*MCAT results. The 2015 *e*MCAT results are not the final result and do not include the results of the 2015 MCAT.

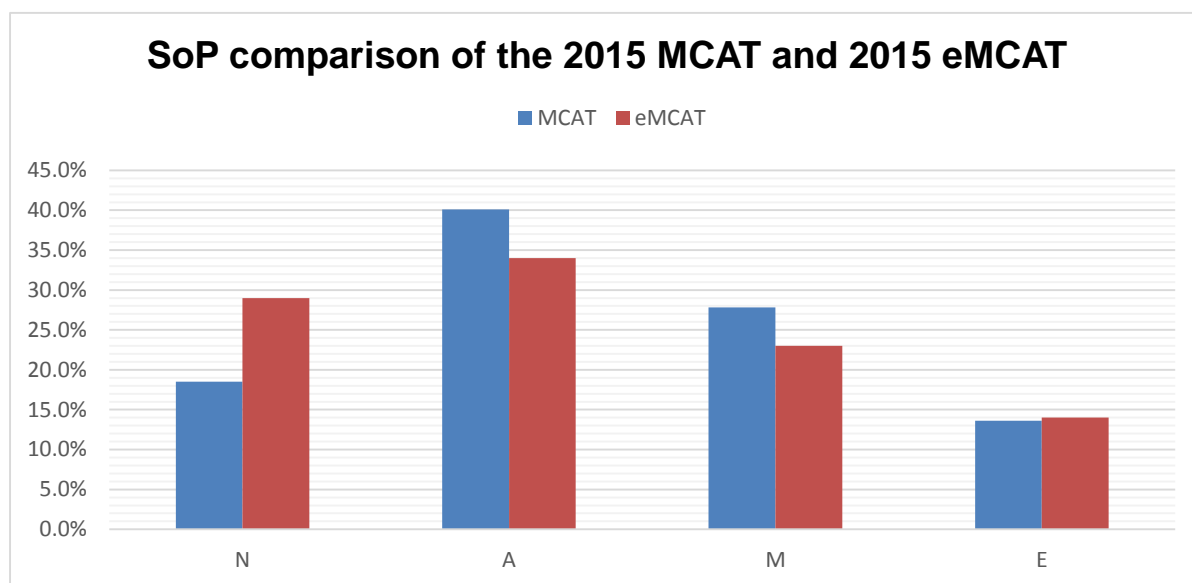


Comparisons between the SoPs of the 2014 MCAT and the 2015 *e*MCAT are as follows:

- 25 percent of participating students gained N in the 2014 MCAT against 29% in the 2015 *e*MCAT, a difference of 4 percent.
- 38 percent of participating students gained A in the 2014 MCAT against 34 percent in the 2015 *e*MCAT, a difference again of 4 percent.
- The 2014 MCAT and the 2015 *e*MCAT resulted in identical achievement percentages for grades M (23 percent) and E (14 percent).

The minimal difference between the two profiles suggests the mode of assessment was irrelevant to outcomes.

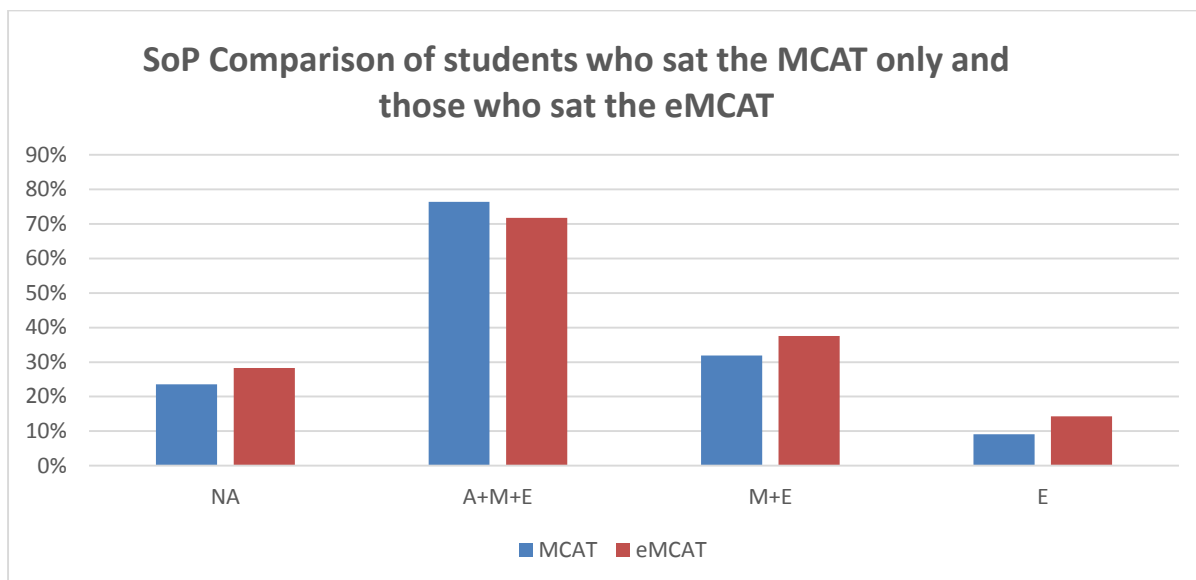
The additional SoP for the students who participated in the 2015 MCAT and 2015 eMCAT below assists in providing a more evident comparison of results.



Comparisons between the SoPs of the 2015 MCAT and the 2015 eMCAT are as follows:

- 18.5 percent of participating students gained N in the 2015 MCAT against 29% in the 2015 eMCAT, a difference of 10.5 percent.
- 40.1 percent of participating students gained A in the 2015 MCAT against 34 percent in the 2015 eMCAT, a difference 6.1 percent.
- 27.8 percent of participating students gained M in the 2015 MCAT against 23 percent in the 2015 eMCAT, a difference 4.8 percent
- 13.6 percent of participating students gained E in the 2015 MCAT against 14 percent in the 2015 eMCAT, a difference of only 0.4 percent.

Note that candidates undertook the eMCAT a week before the MCAT and the 'practice effect' (i.e. students may have performed better in the MCAT after having had "practice" in an electronic version a week prior) should be taken into account when considering these results.



The above graph compares the results of those students who sat the 2015 MCAT only (they did not sit the eMCAT) with the results for those who sat the 2015 eMCAT only.

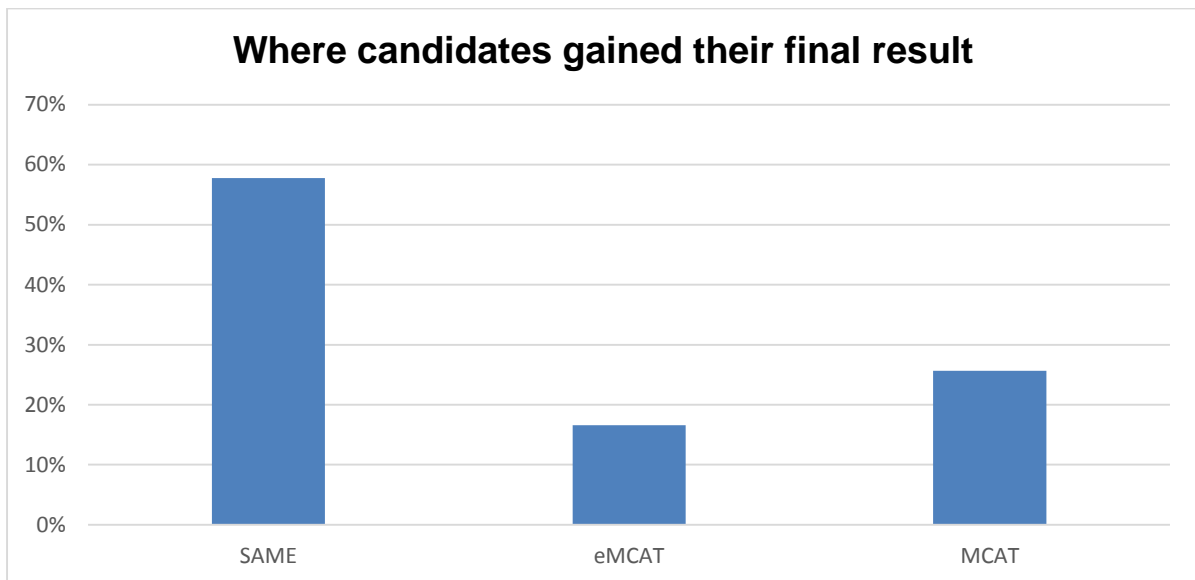
These figures are based on of the following numbers of student participants:

- MCAT, 24,001
- eMCAT, 10,739

From this graph it would appear that those students who sat the paper-based MCAT only had a slight advantage in terms of gaining the standard (76% against 72%) but this changes when a comparison is made at the higher grades. When Merit and Excellence is combined 31% of the MCAT cohort gained an M or E compared with 37% of the eMCAT cohort. The difference is similar at Excellence where 9% of the MCAT cohort gained an E against 14% of the eMCAT. The graph would indicate that the students who sat the eMCAT were more likely to gain a higher grade than those who sat the MCAT. The difference is not large (6% when Merit and Excellence is combined and 5% at Excellence) but it would appear at first glance that those who sat the eMCAT had a slight advantage in attaining the higher grades. The reasons for this disparity would require further research and could be related to decile rating, school characteristics such as rural or urban, confidence in the use of technology or even the nature of the devices used.

Marks by assessment type

Each candidate's final result was determined by awarding the highest result of the two assessments. Out of 10,980 candidates, 58 percent received the same result from both assessments. Furthermore, 26 percent received their final result from the MCAT; this may be due to the "practice effect".



Device usage

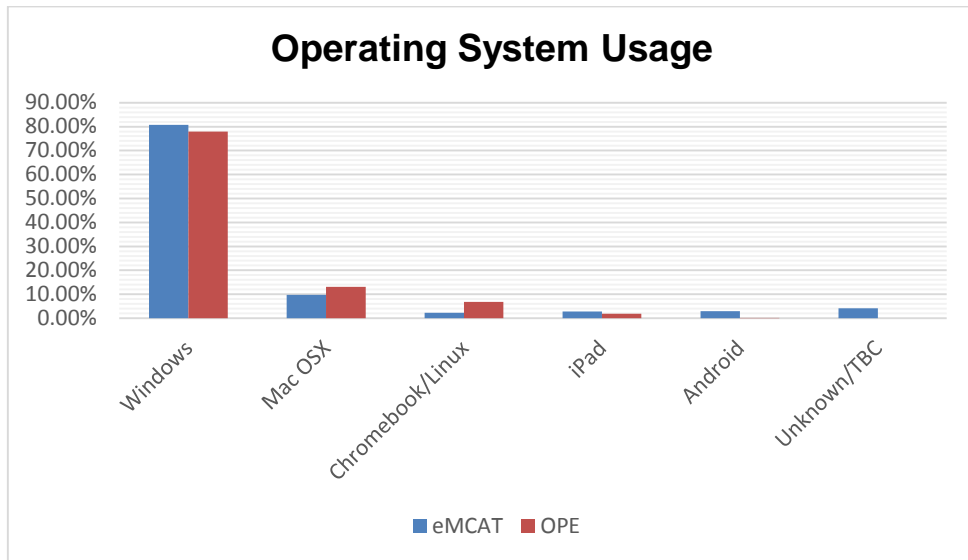
Students were asked how often they used electronic devices in class and of the 3312 students who responded, 80% indicated they used them at least once a week. What is not known is how the students were using their devices. Anecdotal evidence indicates that students were accessing a range of algebraic activities online, especially through Education Perfect, and completing them using pen and paper.

Feedback from teachers suggests that until classroom practice changes to enable more use of electronic devices in mathematics classes, pen and paper will still be used in assessments. This raises a question as to the benefit of undertaking an eMCAT in 2016.

Computer usage in class was similar across all three subjects that were part of the DEAP project, with students tending to use computers slightly more in French, and slightly less in Science.

Operating systems

The majority (76 percent) of the 17,106 students used Microsoft Windows to access the assessment. However, other systems were also used, which reinforces the need for online examinations to be device-agnostic and accessible using more than one operating system.

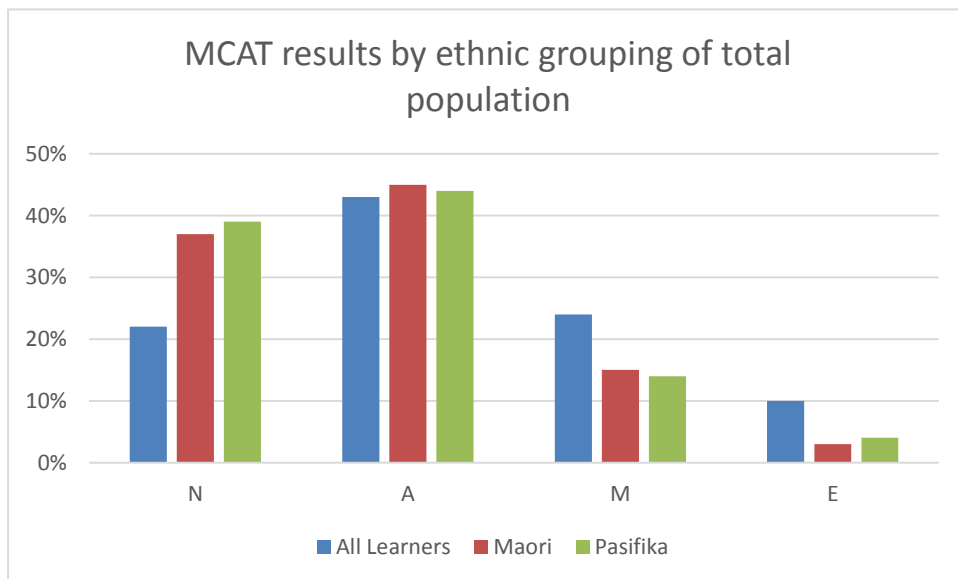


The eMCAT could be completed using a touchscreen, with software converting candidates' handwriting (via stylus or finger) to text for marking. While exact numbers of students who used touchscreen is not known, those who used the iPad and or Android device could only use a touchscreen. This mean that at least 5 percent used a touchscreen for the eMCAT and 2 percent used a touchscreen for the OPE. What is unknown is how many used the touchscreen capabilities of Microsoft Windows 10.

Results by ethnicity

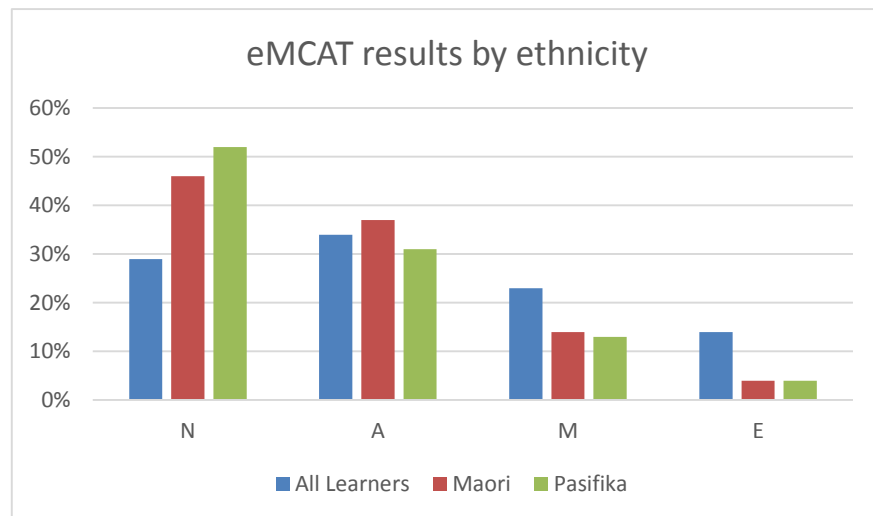
The following graph shows the SoP for the 2015 MCAT only (no eMCAT results are used in this graph). These results provide a comparison of Māori and Pasifika students against all learners undertaking the MCAT and the eMCAT. These figures are based on of the following numbers of student participants:

- All learners 34739
- Māori 5221
- Pasifika 2777



The graph shows that in the grade of:

- *N*;
 - Māori and Pasifika students gained a higher proportion of this grade than All learners.
 - Māori students gained a 14 percent higher N rate than All learners
 - Pasifika students gained a 17 percent higher rate than All learners
- *A*;
 - all three population groups were similar, with no significant difference.
- *M and E*;
 - The graph indicates that Māori and Pasifika student participants did not achieve significantly in these higher grades.
 - Overall, 35 percent of All learners gained M or E, whereas 18 percent of Māori and 17 percent of Pasifika students gained these grades.



Analysis of the eMCAT results indicated a similar trend. The graph shows that in the grade of:

- *N*;
 - rates for Māori and Pasifika students were much higher than All learners in the eMCAT
 - Māori students' grades were 17 percent higher than All learners
 - Pasifika students' grades were 23 percent higher than All learners

- A;
 - rates across all three population groups in the eMCAT had a spread of 6 percent, indicating no significant difference across the three groups.
- M and E;
 - the disparity in participating student grades is more pronounced, with 37 percent % of all non-Māori and Pasifika student participants gaining these grades. Māori and Pasifika student participant grades were less than half this rate (Māori student participants at 18 percent and Pasifika at 17 percent).

This next graph demonstrates the results of the 2015 eMCAT only. These figures are based on of the following numbers of student participants:

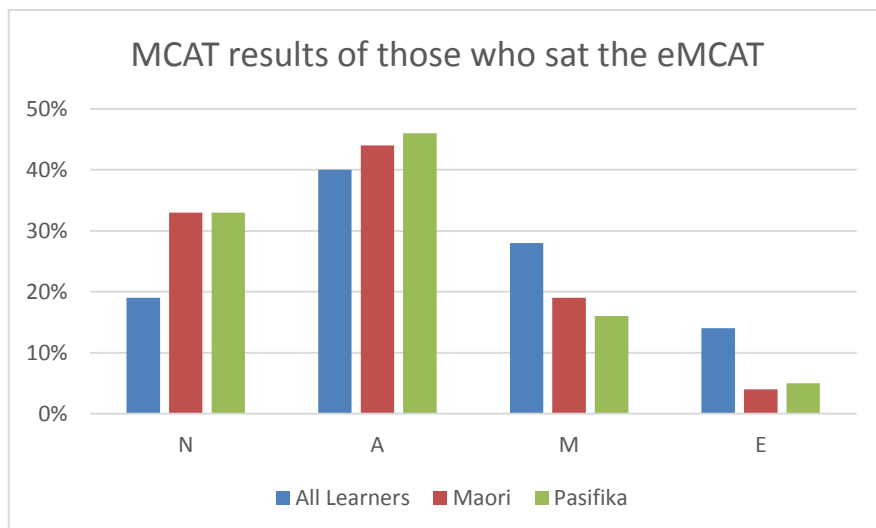
- All learners 10980
- Māori 1312
- Pasifika 809

The graph below shows an analysis of the MCAT results by those who sat the eMCAT indicate a similar trend as the MCAT.

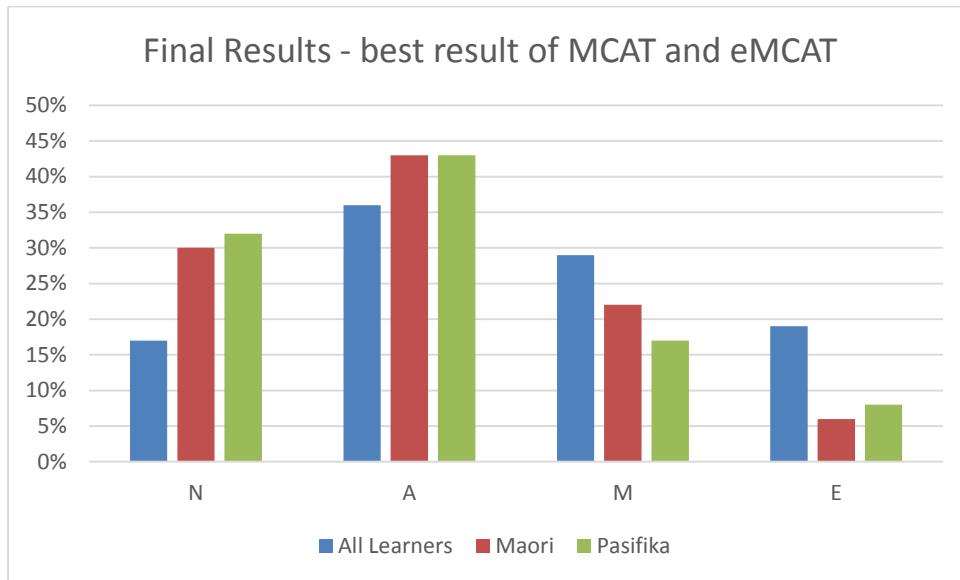
These figures are based on of the following numbers of student participants:

All learners 10739

- Māori 1247
- Pasifika 778



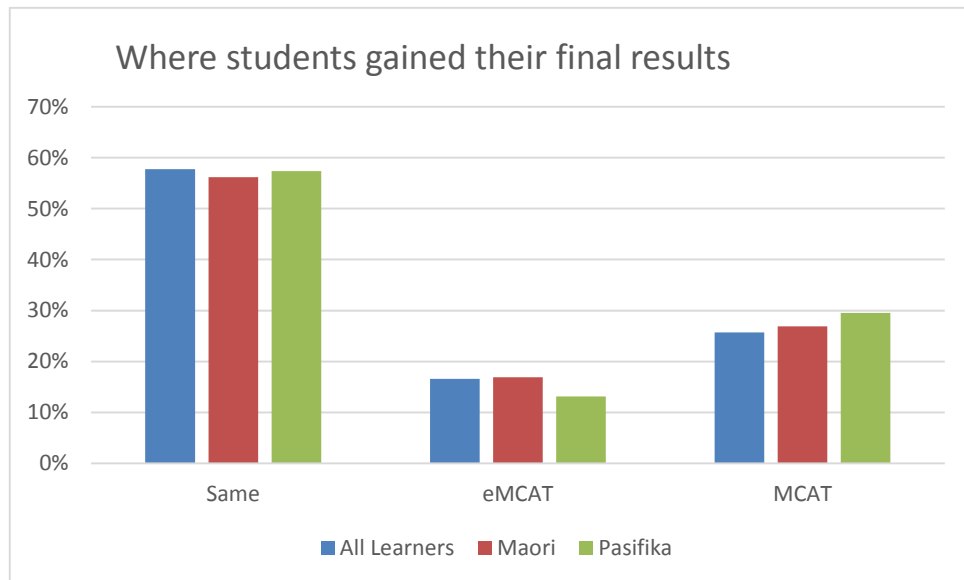
The next graph shows the SoP of the final result. The overall result for an individual student was taken from the best result from the MCAT and the eMCAT.



The graph reflects similar trends in the disparity of grades between Māori and Pasifika participating students, compared to those of non-Māori and Pasifika participating students:

- a higher proportion of Māori and Pasifika participating students did not achieve the standards compared to non-Māori and Pasifika participating students
- fewer Māori and Pasifika participating students gained the higher grades M and E compared to non-Māori and Pasifika participating students.

There was no significant difference across the three population groups in regards to whether their final results were gained from either the eMCAT or the MCAT, with close to half the students gaining the same result from both assessments.



User insights

To gain insight from those involved in the pilot, a survey was undertaken with participating students and the teachers that were involved. The student survey was attached to the end of the OPE and eMCAT assessments, and a link to a teacher survey was emailed to participating schools.

The following data is based on the results of the survey:

- 31.2% of students who participated in a DEAP assessment (eMCAT and/or OPE) completed the survey, noting that the numbers who answered individual questions varied.

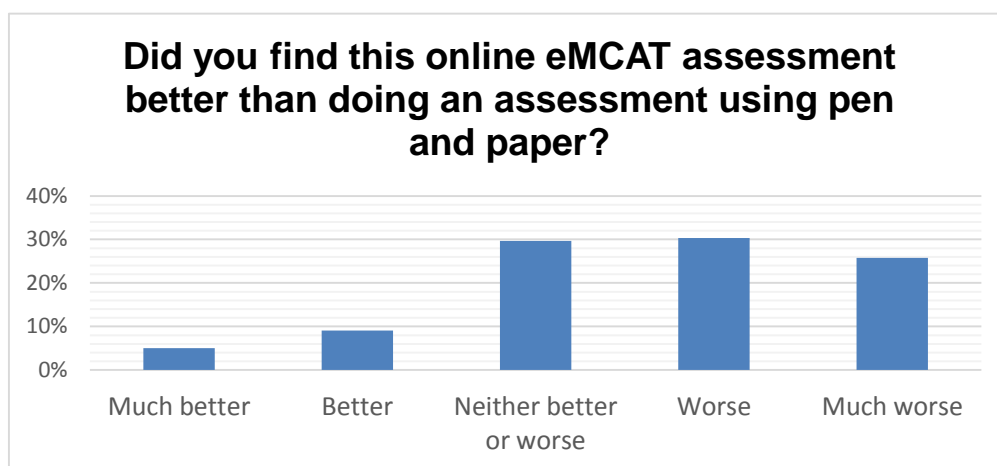
A total of 300 teachers completed the survey.

Student perceptions – eMCAT

Feedback from students who completed the survey was mixed. Some felt that the online version of the assessment was less threatening than an unfamiliar paper examination. However, some students felt that the online assessment was less satisfactory than a paper-based assessment.

The graph below is based on of the following numbers of student participants:

- 5,344 students from a total of 17,106 students completed the survey (31.24 percent)
- 3,586 students from the 11,375 students who undertook the eMCAT completed the survey (31.53 percent)



Of the 31% of students *who completed the survey* at the end of the eMCAT:

- 30% judged it as neither better nor worse
- 60% of eMCAT respondents indicated they felt it was worse or much worse.

Comments from students included the following.

- “I am a faster writer than typist; also some things aren't on the keyboard so you have to move the mouse, which is quite annoying in the test. I also struggled since there

was no area where I could figure out my thinking. Because if you typed it. You have to delete and this caused me to lose my train of thought.”

- “Worse, because I worked the equations out on paper then had to type it all into the computer again.”
- “It takes longer to do maths on the computer. I didn’t have enough time to complete all things I could.”

The majority of students (91.5 percent) *who completed the survey* reported that they also used pen and paper during the eMCAT. This ranged from using paper only as a scratch pad, to deciphering answers on paper and transposing these into the digital format.

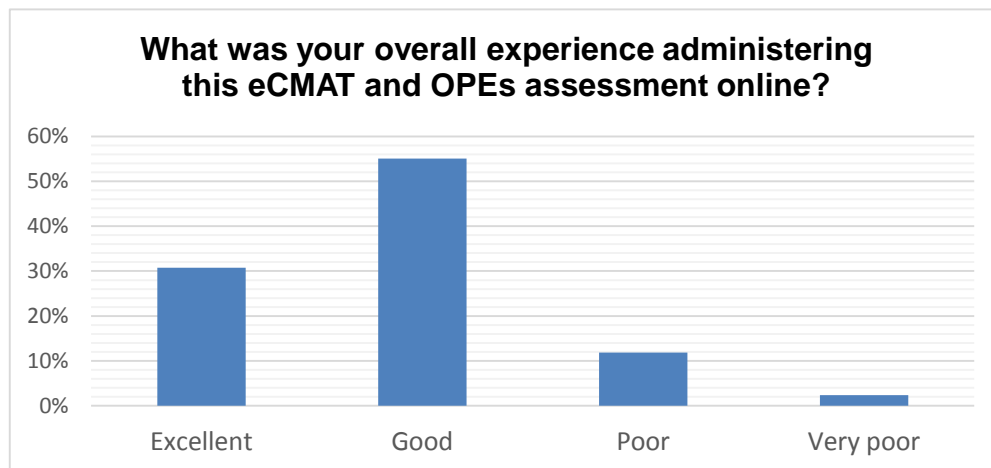
Anecdotal feedback from teachers suggested that until classroom practice, enabled by technology, changes to enable more use of devices in mathematics classes, pen and paper will still be extensively used in assessments. This raises a question as to the benefit of undertaking an eMCAT in 2016.

User perceptions - OPEs

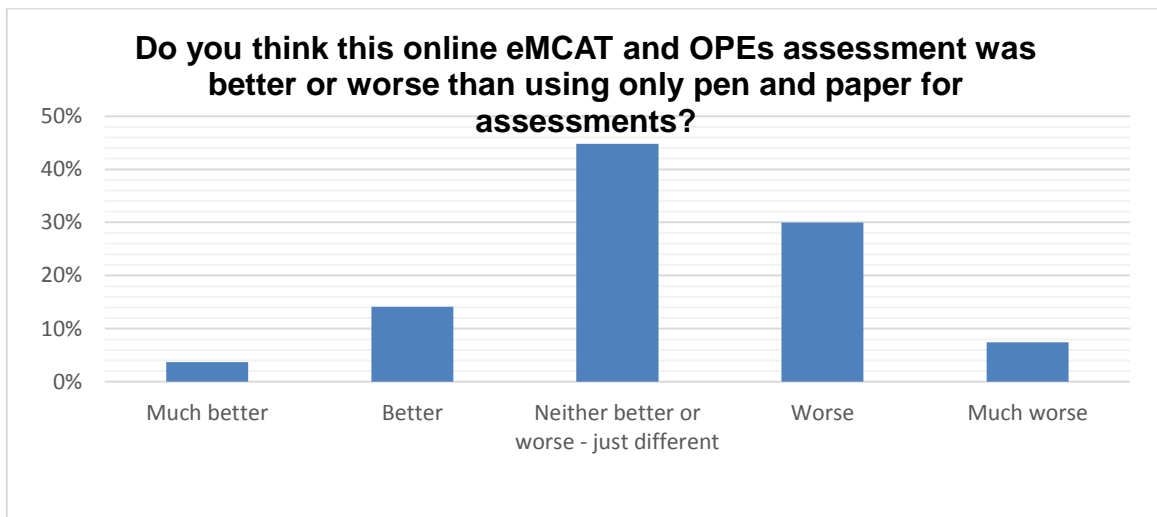
Feedback for the online French Practice Examination was positive. It was reported that teachers and students were familiar with the format, and with using a device as part of their learning. This is possibly as many of the schools who took part already use the Education Perfect online resources.

Teacher perceptions

Teachers were also asked to rate their overall experience, with 296 of the 300 responding, as seen in the table below. This related to both the eMCAT and the OPEs.



Nearly half (45 percent) of the 297 respondents indicated they viewed it as neither better nor worse.



Some teachers also had an expectation that the technology for the eMCAT platform would allow certain innovation associated with online examinations (e.g. copying and pasting formulae from the provided information), which was not possible for a pilot setting.

eMCAT markers' perceptions

Marking was organised along traditional external assessment processes. A panel leader and four check markers undertook benchmarking (i.e. initial marking by comparison with the standard), and then two days training 25 markers. Markers used a portal designed by Education Perfect, which was simple to use and very practical. Markers could easily access the eMCATs they needed to assess and could do so on an item-by-item basis or as an entire assessment. The marking portal recorded their judgements and enabled access by the check marker. Markers could access the assessment for a specific item.

The marking software provided the marker with a recommended result, which the markers could override. Several markers reported that they looked at the recommended result after they had made their judgement as a means of confirmation, but some elected not to use it as they felt it constrained their judgements

Markers were trained to strip mark, which significantly increased the speed of marking. Marking was completed within four days, two of which were spent marking together and two marking at home.

Overall, the markers reported that their experiences were positive. They liked the strip marking (where a marker marks only one particular question of each examination), although one marker commented that they did not get to see the overall performance of the student. Central marking (i.e. a panel of markers working together instead of in isolation) at the start was viewed positively, and as a great opportunity to fine-tune the assessment schedule for everyone.

Markers included feedback:

- It would be useful to be able to annotate the assessment for their own reference, and also for the check marker, to explain their reasoning

- some processes that markers found valuable would not be available to school-based markers, however it may be possible to use a webinar to brief teachers in the future
- some feedback was provided on how best to configure the marker interface.

All markers said they preferred online marking, but some did not like the automatic marking.

Assessment learnings

eMCAT

The MCAT was selected as a pilot because, in terms of external assessment, it is an outlier – it sits outside the normal examination cycle and is only one achievement standard. These two factors made the MCAT a useful vehicle for piloting online assessments.

The standard was developed to assess a candidate's algebraic ability without the use of a calculator, and most of the teaching and learning methods used are paper-based. Every attempt was made to ensure the assessment items were of the highest quality and designed to challenge the candidate to demonstrate their ability: they were essentially paper-based activities replicated in a digital environment.

French and Science OPE

The French OPE contained a series of sound files which the candidate accessed and then answered a series of written questions about. The candidates had complete control over the sound files – they were able to play them when they wished, replay parts, and pause it when they wanted. This contrasts with the paper-based examination, where the sound file is played externally for all candidates, for a set number of times to a specific formula.

A critical element of the standard is the ability to understand spoken French in a natural environment. Feedback from the OPE indicated that the ability of the candidate to play the sound file when they wished was viewed very positively and enabled more flexibility than the current paper-based process. However some teachers commented that allowing the candidate to play the sound file as many times as they prefer, and pausing when they wanted, does not completely replicate a natural environment.

To resolve this for future online assessments, NZQA will seek subject teachers' and HODs' expertise in determining the appropriate protocols for the management and playing of sound files.

All the assessments (eMCAT and OPEs) were developed by contractors and then passed onto Education Perfect to develop as a digital assessment. This somewhat limited the iterative development process for assessments.

Process and technology learnings

The experience with the eMCAT and the Science OPE, where the assessment was a paper-based-based one transferred to a digital medium drew some criticism.

To ensure comparability, the items designed for the eMCAT were very similar to the MCAT, and given the way algebra is taught, were very paper-oriented. Similarly, although the

Science OPE examination development team wished to introduce a range of items that would make use of some digital tools such as animations, the platform was unable to support such assessment without significant investment and time. Therefore, the Science OPE development team stayed with one-dimensional items. This still enabled the testing of process, which was the main objective of the trial.

The layout of the eMCAT was an improvement on the 2014 version, using lessons learnt from pilots that year. There was one item per screen, and the candidate interface was simple, colourful, and engaging. The French OPE had a static sound file and text boxes that increased in size depending upon the candidate response. A suggestion has been made to see if the sound file (and other stimulus) can “float” on the screen so that the candidate has access to it without the need to scroll up and down to start or stop the audio.