Research Report – How markers respond to digital (typed) and scanned (handwritten) exam responses

NCEA Online Programme

Version 1.2

Confidentiality

The information contained in this document is proprietary to NZQA. This document must not be used, reproduced, or disclosed to others except employees of the recipient of this document who have the need to know for the purposes of this assignment. Prior to such disclosure, the recipient of this document must obtain the agreement of such employees or other parties to receive and use such information as proprietary and confidential and subject to non-disclosure on the same conditions as set out above.

The recipient by retaining and using this document agrees to the above restrictions and shall protect the document and information contained in it from loss, theft and misuse.

Table of Contents

NZQA
NEW ZEALAND QUALIFICATIONS AUTHORITY
MANA TOHU MĀTAURANGA O AOTEAROA

QUALIFY FOR THE FUTURE WORLD
KIA NOHO TAKATŪ KI TŌ ĀMUA AO!

1.   Introduction

1.1   How students complete NCEA end-of-year exams

Students complete NCEA end-of-year exams for level 1 – level 3 either:

- digitally, by typing their answers into our online exam platform, Assessment Master

- by handwriting their answers in paper exam booklets.

1.2   What is changing in 2021 for paper exam booklets?

In previous exam cycles, the students' completed paper exam booklets were physically distributed to markers who then marked them manually. The marker would then manually enter the scores into the NZQA system for management of student results and exam marking. This system is called the Electronic Qualifications Authority, or eQA.

In 2021, we will scan NCEA end-of-year paper exam responses for level 1 – level 3 so they can be marked on screen.

1.3   Research to understand the implications of change

To support the introduction of this new scanning and onscreen marking (OSM) process, a research project was proposed to understand how markers respond to digital (typed) and scanned (handwritten) exam responses. This proposal built on the outcomes of previous trials (refer Appendix 1) and research commenced in early 2021.

This research investigated how markers respond to marking student exam responses in digital typed responses and scanned handwritten responses.

The research intended to:

1. Learn about the impact on scoring and identifiable marker preference for responses that have been handwritten, typed, or handwritten and then scanned for marking; and the reliability of scores given to responses – handwritten and scanned.

2. Establish whether, or to what extent, OSM affects marking decisions.

3. Capture the experience of markers to enable an understanding of any impact that has on the scores they award.

4. Make recommendations on whether subsequent research should be conducted using a different Achievement Standard that has multiple item questions, rather than single item questions.

2.   Background

2.1   Examination marking process

Every year, NZQA appoints and trains panel leaders and markers. Panel leaders teach markers how to mark student responses. We expect markers to:

- have curriculum knowledge and teaching experience for the appropriate level, subject, and standard
- understand and have experience of standards-based assessment
- work with others
- work to stringent deadlines.

Up to and including the 2020 exam period, for all but one marker panel, markers have generally been required to live in the region where the panel is based.

Panel leaders are responsible for marker training, including an introduction to using NZQA's OSM system, RM Assessor. Guidance to support OSM is available through:

- an initial webinar presentation
- online video guides
- an RM Assessor user guide.

2.2   RM Assessor

RM Assessor allows scanned responses to be viewed onscreen.

It has a range of marking tools, including the ability to:

- zoom in and magnify the text on a page
- add notes to a response
- use colour coded tags to flag responses for additional attention
- send emails from within the system to other markers
- view one, two, or four pages of a scanned response on the screen at the same time.

From 2021, both paper and digital NCEA exams will be marked online in the marking platform, RM Assessor. Scores will be automatically uploaded into eQA. This is a significant change for a large proportion of markers who have only ever marked paper booklets.

3.   Review of the literature

The main themes identified in the literature review of studies relating to onscreen marking were:

- Marker preference
- Strong trend for acceptance of OSM over time
- Factors that impacted on the ease of use and acceptance of OSM
- Marker productivity
- OSM accuracy and reliability.

Citations for the literature covered in this review can be found in the bibliography at the end of this report.

3.1   Marker preference

Much of the existing literature about OSM considers whether marking onscreen (handwritten) rather than marking digital (typed) responses, has any influence on marker preference and mark reliability. There is little literature to date about OSM of digital (typed) responses or comparing the marking of digital (typed) and onscreen (handwritten) responses.

In 2010, Coniam commented that he anticipated over time OSM will be accepted as the marking norm, with concerns about fairness or convenience (associated with the view that paper based marking is preferable) disappearing over time (Coniam 2010).

In 2006, Verank and Tozuglu commented that the acceptance of technological innovation is inevitable, but change is not always fully embraced and teachers are noted for resisting change (Varank and Tozuglu, 2006).

In a 2018 survey of the literature, Masterman advocated for a closer investigation into the OSM of typed responses and its impact on academics' marking strategies (Masterman, 2018).

3.2   Strong trend for acceptance of OSM over time

The literature about OSM shows a strong trend for acceptance of OSM increasing over time.

Several studies in Hong Kong over a period of 10 years have found that markers had a high level of perceived ease of use and positive acceptance of OSM over time.

Yan and Coniam (2013), Yan and Coniam, (2014) and Coniam (2009, 2010) found markers were more positive about OSM post-marking compared to pre-marking, concluding that as OSM becomes more prevalent it will be accepted as the norm. (Coniam, 2009).

Coniam and Falvey in 2016 remarked that more recent findings revealed how OSM had gradually come to be accepted over the previous six years while the validation studies were being conducted.

Study after study revealed an increasing acceptance of OSM by its markers and increasing familiarity with, and acceptance of, the technology (Coniam and Falvey (2016).

When OSM of scanned responses was introduced in New South Wales, Australia, in 2009, markers were reported to have seen it as a positive development.

Some of the feedback from senior markers included 'blown away', 'instantaneous revelation' and 'just the natural progression'. (Board of Studies, Teaching and Educational Standards, 2010).

Alternatively, Kernard and Arnold commented in 2016 that markers' experiences were sometimes contradictory in nature with some praising the performance of the software and others criticising the system functionality or speed. They concluded that enforcing electronic marking would be problematic until the supporting systems, processes and policies catch up (Kennard and Arnold, 2016).

3.3   Factors that impacted on ease of use and acceptance of OSM

Factors that impacted on the perceived ease of use and acceptance of OSM have included marker, experience, gender, age and subject area.

In 2010, Coniam found that less experienced markers

- preferred OSM
- were less negative about reading off screen

- reported needing fewer breaks than more experienced markers (Coniam, 2010).

In 2009, Coniam found that for markers who have only ever marked on paper or who were computer-averse, the technology-related changes may not have been welcomed. He concluded that as OSM becomes more prevalent, it will be accepted as the norm (Coniam, 2009).

A UK study found marking outcomes were relatively unaffected by marking mode including amongst experienced markers (Johnson, Hopkin, Shiell, Bell, 2012). Generally, markers who have rated themselves as competent technology users have responded positively to OSM (Coniam and Yeung, 2010). Coniam also concluded after a 2011 study that tablet owners were more positive on almost every survey question, both pre- and post-marking. He found that regular use of tablet computers predisposed users towards reading on screen, and consequently appears to have had a knock-on effect of greater predisposition towards acceptance of onscreen marking (Coniam, 2013).

Gender has been considered a factor that may influence perception of OSM. Yan and Coniam (2014) found males demonstrated a higher level of perceived ease of use and acceptance of OSM in some studies than females. (Yan, Coniam, 2014). This was possibly due to males reporting higher levels of confidence regarding the use of new technology. (Yan, Coniam, 2014).

Evidence about the impact of age on marker preference for OSM is not conclusive. In the (Coniam, 2010) study, younger markers were more positive than older markers concerning the introduction of OSM and were less negative about continual reading off screen and screen resolution (Coniam, 2013).

A 2014 study found that owners of tablets were generally younger and more likely to be accepting of OSM. (Yan, Coniam, 2014). However older markers, who it was assumed were more likely to be more experienced teachers and/or markers, have higher levels of tolerance when under pressure during marking (Yan, Coniam, 2014).

3.4   Marker cognitive load

Studies on whether reading onscreen presents a higher cognitive workload than reading on paper and the impact of cognitive load on readers' ability to comprehend a text is inconclusive. Others have concluded that markers find it easier to read and mark paper responses than scanned responses (Tung-hsien, 2019).

Johnson and Nadas (2009) reported that markers found their cognitive load increased at first when marking scanned responses for the first time, but reduced as they became more familiar with the experience. (Johnson and Nadas, 2009).

Coniam (2013) cites Campbell (2005) who states that marking onscreen poses considerable cognitive demand due to the need to read from the screen. Shaw also commented that the presentation of lengthy texts onscreen, where problems of navigation and location as well as document manipulation arise, is challenging. Moreover, there is evidence that active reading behaviours are affected by the mode (Shaw 2008).

Shaw cites his earlier research in 2001, claiming that poorly handwritten responses can contribute to the increase in cognitive load as the effect of poorly handwritten answers seemed to increase when translated into a digital image. Shaw also noted that creating a digital image of a candidate's work, where handwriting is poor, results in exacerbating the effect of illegibility and is an impediment to successful screen reading (Shaw, 2008).

Masterman (2018) cites Powers, Fowles, Farnum and Ramsey (1994) as finding that typed exam responses tend to be longer than handwritten ones but give the visual appearance of being shorter

than handwritten essays, even where their word count is the same or greater. This remained the case even when typed responses were subsequently transcribed into handwriting and vice versa. This finding is important because research generally indicates a positive correlation between the length of the response and the mark achieved, whether typed or handwritten (Masterman, 2018).

3.5 Marker productivity

The question of productivity has highlighted significant differing views amongst participants, with some participants feeling electronic methods saved time overall compared to paper-based marking and others feeling it took longer to mark onscreen.

Kernard and Arnold concluded that there was consensus that economies of scale could be achieved if the tools were used over a longer period (Kennard and Arnold, 2016).

Coniam found that OSM relieves participants of administrative duties such as having to physically count, sort and check responses (Coniam, 2011) and increases work efficiency. Yan and Yang, M., Yan, Z., and Coniam (2017) concluded this is due to offering better concentration, real-time access to marking statistics, automated processes and the ability to adjust font sizes and calculate and record marks. (Yan and Coniam, 2017). General fatigue and eye tiredness were noted in studies in 2011 and 2016 (Coniam 2011 and Coniam and Falvey, 2016).

3.6 OSM accuracy and reliability

Mark reliability can be defined as different markers awarding the same or similar grades to different responses of similar quality.

Research on whether marking mode (onscreen or paper-based) affects marking accuracy and reliability is inconclusive and for the most part no association has been found between marking mode and marking reliability.

3.6.1 Accuracy

Several studies have found that marking accuracy for essays was not influenced by marking mode, concluding that markers are able to mark essays with equal accuracy on screen as they do paper.

Johnson et al found that while the examiners' marking behaviours varied when working in the different modes, it might be inferred from the comparatively similar marking accuracy levels that the markers were able to comprehend the essays equally well across paper and onscreen. They concluded that the digital marking environment provided sufficient opportunities for the examiners to employ those marking behaviours that they felt were appropriate at different points in their marking without compromising accuracy (Johnson et al, 2012).

They also pointed out that analysis of mark reliability between modes might mask the true level of marking variation because the mark disagreements between markers at the individual script level is not taken into account (Johnson and Nadas, 2009).

Similarly, in 2010, Johnson et al found that markers' profiles appeared largely unaffected by mode, with markers who tended to be more lenient on paper also tending to be more lenient on screen (and vice versa) (Johnson et al 2010).

3.6.2 Reliability

Nadas and Bell (2010) also found that marking mode did not present a systematic influence on marking reliability. The analyses compared examiners' marks with a gold standard mark for each essay and found no shifts in the location of the standard of recognised attainment across modes (Nadas, Bell 2010).

Other studies have concluded that the reliability of OSM is comparable with paper-based marking regardless of examination questions (Johnson, Nádas, & Bell, 2010, Yan et al, 2017).

A 2012 study by Johnson, Hopkin, Shiell and Bell found that markers were slightly more lenient on screen than on paper and there was no significant association between marking mode and the magnitude of marking accuracy, and only a small and extremely weak association between marking mode and the direction of marking accuracy. The findings therefore suggest that the markers were marking with similar accuracy in both marking modes (Johnson, Hopkin, Shiell and Bell, 2012).

In 2008, Mogey, Purcell, Paterson and Burk found that any variation in the mark awarded due to difference in format (handwritten vs typed) was insignificant compared to variation due to differences between markers. They found that variation between markers will remain a very important influence on marks. They concluded that although there is evidence of a small impact on marks due to the format, this is not the main source of unreliability between exam scores (Mogey, Purcell, Paterson, Burk 2008).

Masterman concluded in 2018 that evidence shows that reliability between markers may be improved by the marking of typed responses. The largely insignificant differences between the marks achieved in typed versus handwritten exams suggest that the risk of grade inflation or deflation resulting from the change of tool is negligible (Masterman, 2018).

Mogey, Purcell, Paterson, and Burk came to an opposing conclusion in a 2008 study, finding that markers may be influenced by format – and that difference might be worth almost two marks (in favour of typing) to the average student (Mogey, Purcell, Paterson, Burk, 2008).

The literature also supports a contrary view where markers are more likely to demonstrate empathy towards, and reward, handwritten responses. Masterman (2018) again cites Powers, Fowles, Farnum and Ramsey (1994) who refer to a "reader empathy effect" between the marker and a student who handwrites their exam script, with the marker feeling "closer to the writer" of a handwritten script. Powers, Fowles, Farnum and Ramsey (1994) suggest that the marker may give the student the benefit of the doubt over illegible patches or interpret crossings-out as evidence of the student's attempts to revise their work (and reward the student accordingly) (Masterman, 2018).

In 2014, Charman reported that the question of handwriting and its interpretation can pose a problem for establishing the comparability of computer and paper-based writing tests (Charman, 2014).

4. Methodology

4.1 Sample responses

A total of 1,000 responses from the Level 2 English Achievement Standard 91098 answer booklets were randomly selected from the 2020 NCEA end-of-year exam paper responses.

Initial analysis was undertaken using a unique identifier to ensure the random selection of responses represented the full range of NCEA achievement levels – from Not Achieved to Excellence.

The responses were scanned into RM Assessor[1] (refer Figure 1).

Thumbnail images of each page were shown on the screen and participants could navigate through the responses by clicking on the individual pages. Each page had to be clicked on before a score could be entered.

Figure 1: Sample of scanned handwritten response
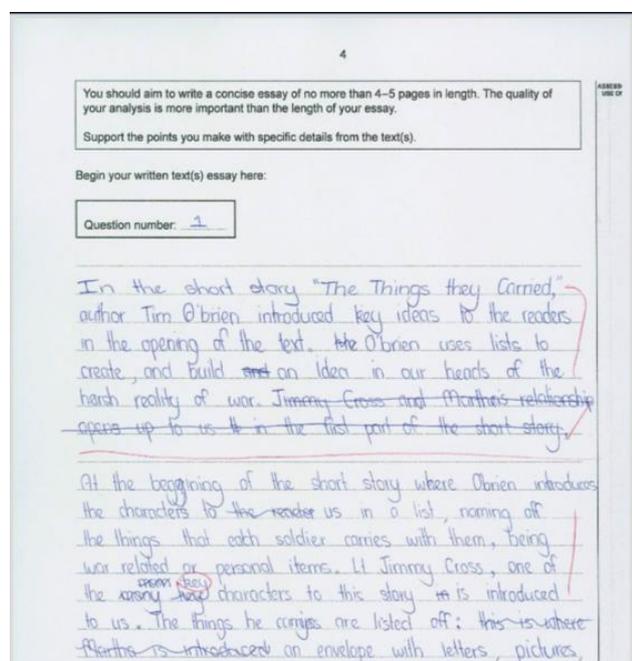
The photocopied booklets contained the score that had been given by the allocated marker. In an attempt to hide the score prior to scanning, NZQA used black marker to obscure the score from markers. Due to the high specification scanners used, the original score was not completely obscured on some scanned images as the scanners could capture the indentations from the pen used by markers to record the score.

A random selection of 1,000 digital (typed) responses from the same standard were loaded into the RM Assessor (see overleaf).

---

[1] Photocopies of the original responses were scanned as the original 2020 paper responses had been returned to students in early 2021. For the 2021 exam cycle, responses will be scanned from the original script.

Figure 2: Sample of digital response[2]
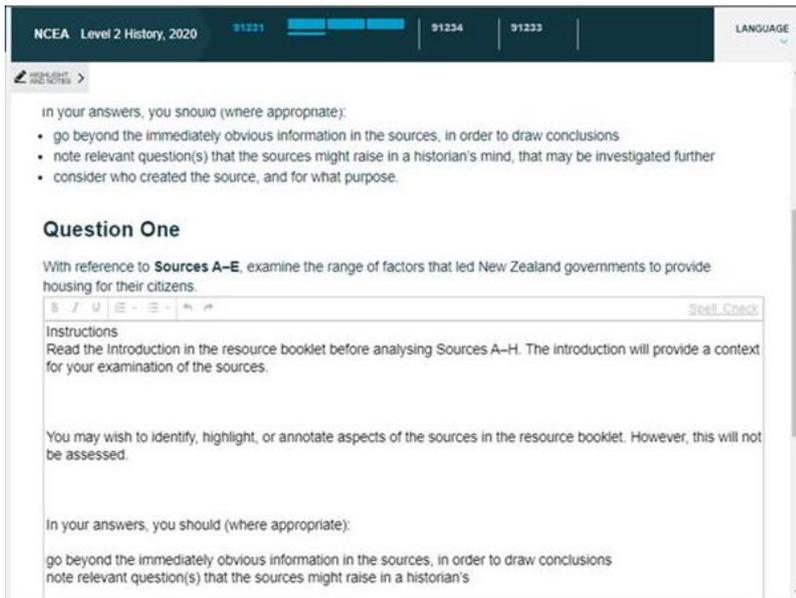
4.2   Research participants

Members of a marking panel who marked the NCEA Level 2 English 91098 standard (typed and handwritten responses) in 2020 were invited to participate in the trial. Information about the research was sent to panel members (refer Appendix Two) along with a consent form to complete. The participating panel comprised of:

- one panel leader

- two check markers

- four markers.

Participants were asked to confirm their consent to participate in the research and complete a short survey before attending the workshop. The intention of the survey was to gather demographic data and gauge the markers' level of comfort with technology and marking. The survey consisted of 14 questions providing information about:

- demographic characteristics including age, gender and ethnicity

- experience with marking

- level of comfort with marking and technology

- marking preferences

- use of technology.

Questions were either multi choice or used a five-point Likert scale where 1 denoted a positive response and 5 a negative response.

---

[2] Sample is not from an English examination and is used to illustrate the different modes of student responses.

4.3   Data collection

A four-day research workshop for research participants was held in the NZQA head office in April 2021. Three participants (panel leader and two check markers) arrived on the first day, and the remaining participants (markers) joined on the third day of the workshop.

Semi-structured group interviews were held with participants before they commenced marking processes. Participants asked about their experiences with marking and expectations about marking scanned responses. One of the researchers led the groups and the other took notes to summarise the findings of the groups. Audio recordings were transcribed following the focus group discussions.

To replicate the normal marking experience as much as possible, participants worked through their usual processes when marking. This included the panel leader and check marker participants convening to discuss the criteria for scoring the responses, benchmarking and classifying responses in advance of the remaining participants arrival on day 3.

The panel spent the third and fourth days marking in the NZQA office. Participants attended a second group interview before leaving the office to ascertain their views about the experience of OSM. Not all responses had been marked during the research workshop, and participants continued marking when they returned home.

Marking was completed by 8 May 2021 when all sample responses had been allocated a score.

Once marking was complete, a semi-structured phone interview was conducted with each member of the panel. This allowed the researchers to record any marker observations made during marking after they had left the NZQA office and had returned to their usual location for marking. Any trends the researchers had observed in the data were also probed.

Transcribed marker focus group responses were coded in vivo using NVivo 12 data analysis software.

Data was also collected from RM Assessor. This included:

- time taken to mark each response by marker by scanned and digital
- time taken on average to mark each response by marker by scanned and digital
- number of responses marked by digital and scanned responses
- scores allocated to the digital and scanned responses.

5.   Research outcome

5.1   Overview

The research outcomes can be grouped into these main areas:

- Participant characteristics
- Participant comments prior to OSM
- Participant comments post OSM
- Analysis of marking data

5.2 Participant characteristics

5.2.1 Participants' demographics and experience

Research participants were asked to provide demographic information in the pre-trial survey. Of the seven participants:

- six were female and one was male

- five were New Zealand European (or European) and two were Māori

- five were in the 40 – 49 year age group; one was in the 30 – 39 year age group; and one was in the 50 – 59 year age group.

Four participants had had more than 10 years' experience of marking NCEA, two had been marking for between 5 and 10 years and one had been marking for between 1 and 5 years - this participant had also only marked NCEA Online assessments for one year. All the other participants had had more than one year's experience of marking NCEA Online.

All participants with more than 10 years' experience stated they were extremely confident with marking NCEA external assessments. Participants with between 1 and 10-years' experience said they were somewhat comfortable marking. When asked about marking online, the responses were the same.

There was no correlation between gender and marker confidence and with just one male completing the survey it was not possible to draw any conclusions about gender from the data.

All participants said they marked at least some of the time at home.

When asked how comfortable they were with using digital technology, six markers said they were somewhat comfortable using digital technology, with one marker stating they were very comfortable and at expert level. The participant who stated they were at expert level had more than 10 years' marking and was female.

Six participants said they used a laptop for general technology activities (not marking) most of the time and did not use a monitor. Two participants used a desktop with monitor some of the time. All participants used a smart phone. There was no correlation between technology use and marking experience or confidence.

When marking, three participants used a laptop without a monitor some of the time, one used a laptop or desktop with a monitor some of the time and six of the seven marked at least some of the time using a laptop without a monitor. One participant always used a laptop with no monitor for marking.

When asked when they preferred to mark there was a mix of answers. None of the participants had only one time during the day during which they would mark. All seven participants marked during a mix of before and after school, in the evenings, and during the weekend.

5.3 Participant comments prior to undertaking OSM

5.3.1 Comments about things that make marking easier

During the pre-marking discussion group, research participants identified that a disciplined approach and good organisation made the marking process easier. This included having clear timelines about panel meetings, and when marking should be completed. Other demands on time from home and

family during the exam marking period can impact on participants' marking capacity, so proactive time management was identified as an important success factor. One participant noted:

> *"Having the outline of the [responses] per day is crucial, and I always do more than what it says to, like I try and frontload it a little so that I've always got some days breathing space".*

Participants agreed that taking a methodical approach and marking several exam responses at a time (at least 10 at a time) meant that there were regular breaks to keep participants fresh. Marking in the morning was preferred, so that participants were less fatigued. Those participants who marked digital responses using a laptop stated that having access to a larger screen was also helpful.

Exams are marked out of school hours, mostly at participants' homes, and being able to have uninterrupted focus time also made marking easier. For one participant this meant setting clear boundaries:

> *"[My children] know that they must knock on the door, and if I say, "go away," they must go away unless the house is burning down, don't come in. They've grown up with that rule, and I'm just marking".*

There were some efficiencies noted when marking digital responses compared to paper responses. Paper responses are 'contained' whereby each script is reviewed and any blank space crossed out to 'contain' the student's response. This process is not necessary for digital responses, which show typed student responses. One participant said:

> *"Not having to contain is a big [advantage]. I also find [digital marking] faster, I guess because with the typed responses you're not needing to decipher, so that is the difference thus far".*

5.3.2 Comments about things that make marking harder

Participants identified more factors that made marking harder related to paper responses than digital responses.

Participants described the effort by markers that goes into managing, sorting, marking, completing administration, and returning marked responses before the due date. Paper responses are individually wrapped in plastic and delivered to markers by courier. Markers must ensure that the boxes are securely stored. One participant said:

> *"[For paper] there's a lot of handling, whereas where a digital they're [processed] once and then it's gone. You don't have to check it's in the box and then contain it and then mark it and then send it off for check moderation and then find where it went and then double enter it and then put it in a box and then send it off".*

Delays in receiving paper responses due to markers' distributed locations caused frustration:

> *"Last year I had one bag [of paper responses] that got picked up in the morning. So, you had the front office person has taken a picture of the courier person taking it away, and it was like nine o'clock in the morning on the Monday, and it didn't get to me until the Wednesday afternoon".*

Participants noted that marking paper responses by students with poor handwriting and digital responses with phonetic spelling made marking harder:

> *"[I had difficulty with the] spelling, trying to figure out what that word is, you know, you can read it out loud… Once you figure it out you go back, and you read it like it's natural".*

Spending too much time marking responses without taking a break was viewed negatively by participants. Each participant expressed an individual view of the number of responses that were acceptable to be marked in a single sitting, ranging from a finite number of responses (e.g. 10 at a time) to a finite length of time (e.g. no more than 2 hours at a time). Breaks were described as physical e.g. getting up and moving around, cognitive e.g. shifting between different modes of marking (paper and digital) or finding a 'void' or blank paper response. One participant said:

> *"I think you need to mix up the online with the physical, so you might do 45 minutes online and then balance it [with marking paper responses] because otherwise [marking is] quite straining".*

5.3.3 Comments about marking both typed and handwritten responses

The novelty of marking both types of typed and handwritten responses together was commented on by participants:

> *"Traditionally we've always done, written, you know, paper [responses] first, and then we've gone onto digital; and so, from what I understand it could be changing back and forward. So, you would never know if it's going to be a digital or a scanned one".*

There was discussion about whether no longer being able to hold a paper script would affect how participants felt emotionally connected to the marking process:

> *"I just get more of a sense of some personality from a script than I do from a screen… but that's how I feel. I don't know if it's a holistic thing, but [when marking all digital responses] I found myself wanting to go back to paper [responses] and maybe that's the difference between reading a book and having the hard copy, and reading something online, but you do get used to reading online in saying that."*

Participants assumed that the marking process would be faster when marking onscreen. Responses can be accessed online quickly after the exam, and therefore participants assumed marking panels can be convened earlier for check marking and benchmarking[3]. One participant said:

> *"We can do the benchmark within days rather than having to wait, like six days. We do the normal panel stuff, and then we send everyone away and say, 'You mark these five [responses] before you do anything else.'*
>
> *And then you have the conversation with your check marker.*
>
> *As a check marker I go, 'okay, this person is interpreting the standard this way', and so, I have that kind of logged in the back of my brain".*

Without the need for physical distribution of paper responses, markers could be located anywhere in New Zealand.

> *"I'm quite excited about marking fully online because now I live in the South Island[4], but I don't know whether it's going to be harder to 'get your eye in' [and mark consistently]".*

---

[3] The marking process includes checks and balances to ensure all examination results are accurate and fair. The panel leader and a senior marker (check marker) mark student scripts from the current year and compare them with student scripts from the previous year(s) to benchmark the examination. https://www.nzqa.govt.nz/assets/About-us/Publications/Brochures/NCEA-factsheet-4-July-FINAL.pdf

[4] This panel comprised markers located in the North Island.

A perceived benefit of onscreen marking was that if any markers are scoring outside the expected range, this can be observed and responded to more quickly through the marking platform, unlike when marking paper responses.

> *[As a panel leader] that's always for me a little bit of a nervous time because we say, "stop marking" [paper responses] but they don't. So, therefore by the time it got to me, I got six lots of 20 [extra paper responses to review].*

Markers' access to responses on the system can be controlled by the panel leader in real time, so that adjustments to scoring can be resolved early.

5.3.4 Comments about OSM having greater flexibility and security

Onscreen marking was envisaged to give markers greater flexibility about where and when marking could be undertaken.

Security for marked responses was perceived to be related to logins and passwords, rather than space. Not having to manage the security and administration of physical responses was expected to be a great advantage by participants. One comment was:

> *"You've got those annoying ones in plastic bags. Oh, yes, that'll be great to get out of them".*

The administration for paper responses was reported as being onerous, and onscreen marking removed those tasks. One comment was:

> *[For paper responses] I'll have my document and I'll have all the NSNs written down, all the student response numbers, I'll have the grades, and I'll have the comments, and that's time consuming. Not having to double enter [grades] at the end of the [onscreen marking] process will be quite significant".*

5.3.5 Assumptions about OSM prior to marking

Participants expressed concern about the readability of scanned responses. One comment was:

> *"I'm quite apprehensive about it. I'm like, the whole screen combined, like if you get one with a doozy of a handwriting, then it just being on a screen might compound it".*

Participants perceived digital responses were easier to mark because:

- responses are typed
- students can edit text
- students can use spellcheck.

Concerns were also expressed about how markers will deal with handwriting compared to digital responses when marking onscreen. This was both a functional concern about legible writing, and the impact on scoring. One comment was:

> *"Number one is making sure that those students who are writing are not going to get penalised because their work might be harder to read and making sure that there's a fairness is maintained".*

When asked about what would happen to scores when responses were marked a second time, participants felt that they were treating this as a completely new marking exercise, with new

benchmarking established, and therefore they did not expect that scores would be matched. One comment was:

> *"I think [scores] won't be the same because we've had a period of separation from what's gone on prior.*
>
> *There's pressure but not the intense pressure that we are usually under.*
>
> *I think the most important thing is that we don't pressure ourselves to come up with the same mark that was originally there, because we've actually changed the parameters somewhat".*

Other assumptions related to physical implications of marking everything onscreen, such as:

- the impact of looking at a screen for long periods
- how breaks will be managed
- the ability to mark an increased number of responses because the management of completed paper responses would no longer be necessary.

5.4   Participant comments after undertaking OSM

At the end of the workshop, participants were interviewed in a group about their experience of OSM, and at the conclusion of all marking, were interviewed individually over the phone.

5.4.1   Participants were very positive about OSM

Overall, participants comments were very positive about the experience of OSM. The word cloud in Figure 3 derived from NVivo 12 analysis of the transcript from the final group interview prominently features the word 'like'. The larger the size of the words shown in the word cloud, the more frequently the word was used by participants during the interview.

Figure 3 – Word cloud

Participants were very positive about the experience of marking handwritten responses onscreen.

> *"I remember one of the things I was worried about, I think based on a question that you asked us, was that the scan ones might be more difficult. I thought they would be, but they weren't at all"*

The ability for the system to zoom in on text mitigated previous concerns about whether it would be difficult to read handwriting onscreen.

> *"I like making it bigger, zooming, because when we have [paper] in front of us, we can't zoom.*
>
> *Sometimes that made it easier to read the scrawl".*

Another positive aspect described by participants related to not having to deal with paper responses.

> *"For me from an organisational side of things, I like things to be quite neat and tidy when I'm marking. And so, submitting it [in the marking system], and it's whipped away, I feel satisfaction, but I also feel tidy.*
>
> *I don't have to put it on another pile in my office area, it's done and tidy and organised."*

Participants were very positive about the ability to collaborate on check marking by messaging through the marking system. This significantly increased responsiveness for guidance from check markers who could be contacted in real time rather than having to wait for hard copy papers to be distributed by couriers.

> *"Yesterday I had an interesting paper, and my check markers could see it at the same time as I was seeing it on my screen and I was thinking about that in terms of real marking and so forth, [another marker] could theoretically message me and say, "can you hop on and look at this paper?" right then and there. We could be looking at it at the same time and having that conversation and talking about that paper while we're both reading it".*

Participants indicated that this would improve marking efficiency, as resolution could be sought while markers still had their 'eye in' and were marking consistently, rather than having to delay responding at a later time, out of context.

The workshop was held at NZQA national office away from participants usual marking environments, using the participants own laptops in meeting rooms that did not have adjustable furniture, which contributed to fatigue. When asked, participants commented that they needed to be mindful of taking breaks to mitigate any physical effects of OSM. For example:

> *"I found that my eyes got sore when I was using my laptop, and I felt I was looking down [too much]"*

Any prolonged repetitive activity can impact on wellbeing, whether marking is on paper or onscreen. RM Assessor limits the number of responses that can be accessed in one marking session, which will naturally encourage breaks to be taken consistently.

5.4.2   OSM seen as more efficient

Onscreen marking was seen as a more efficient way of working than marking on paper, removing much administration associated with processing physical responses.

> *"When you get [paper responses] the initial check that you've got to do, and check.*

Recording which ones, you send for check marking, and keeping those piles, when the piles will come back.

…and some days, let's say you've just marked a hundred and then you've got to go and enter those hundred in. That's painful too. That'll save that time".

Marking handwritten and typed responses within the same time period and system was reported as providing a cognitive break during marking, as opposed to creating greater cognitive demand.

"Have them mixed in with the digital ones. I thought that might be annoy[ing] but it was good because if you have too many digital ones in a row, I don't know, it just felt good to me to have it mixed up.

It kind of forces you to pay attention to what you're reading because sometimes you've got to decipher writing and stuff; so, you're waking up again".

An advantage of having all responses securely stored online within the marking system, was that any requests from NZQA about responses could be dealt with by NZQA, rather than being the responsibility of the marker. One comment was:

"I was thinking about sometimes NZQA requests the paper [response] and so forth, and that can be a mission because you're like, 'is that one of the papers that you've marked?', or it's in the pile that you haven't marked and you're fossicking through hundreds of papers to find that paper. I was just thinking, 'well, if NZQA needed a [response] well they would just whip it on out [of the marking system]'".

5.4.3   Advice from participants about induction for markers new to OSM

Participants provided advice as to what should be considered when inducting markers to the OSM process for the first time.

This included:

- having access to guidance documentation
- learning opportunities
- ability to practise with a version of the OSM system
- consideration of managing physical wellbeing while undertaking OSM.

Participants identified a functional problem using the system, whereby handwritten responses did not load immediately, and this was identified by NZQA as a defect in the test environment that has been analysed and resolved for end of year marking.

5.5 Analysis of marking data

The sample of responses provided for participants during this trial was randomly selected from the body of marked responses that had been originally marked by the same marking panel. Analysis of the responses showed that there were no papers re-marked by the same panel member participating in the research.

5.5.1 Average marking time

The average time (in seconds) taken by all participants to mark allocated typed and handwritten responses was analysed over the seven-day marking period (30 April to 6 May).

The data shows that on average across all participants, typed responses took slightly longer to mark than handwritten responses.

Both Figures 4 and 5 show that over the seven-day period, participants' marking speed improved, with a 20% reduction in the time taken to mark as participants became more practised.
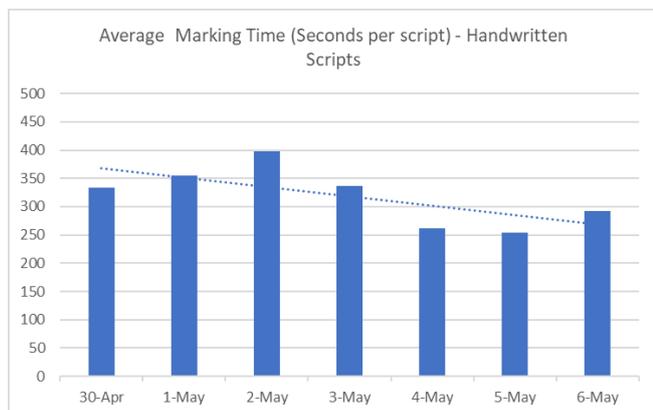
Figure 4   *Figure 5*
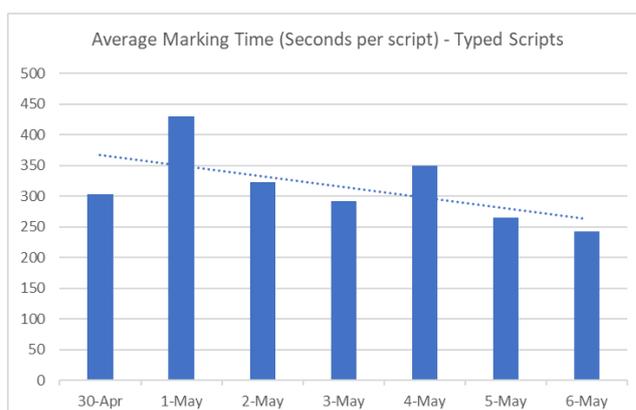
5.5.2 Analysis of scores

There were some aspects of the research design that impacted on the outcome of scoring:

- Research participants treated OSM as if the responses had not been previously marked. Therefore, the handwritten and digital responses were assessed anew, with new benchmarks set, and marking was aligned with the new benchmarks. This meant that comparisons between original scores and scores from OSM were invalid and unreliable.
- The sample of paper response booklets was randomly selected and were reviewed to exclude 'void' responses, i.e. booklets with no content. However, the digital responses sample was not reviewed, and included 20% 'void' responses.
- As noted previously in section 4.1, the original score was not completely obscured on some scanned images. While participants reported that they did not intentionally look to see if they could make out the obscured score, they were aware that some scores may have been visible.

Due to these factors the research team determined that individual scores would be unreliable, and therefore could not be used for reporting.

6. Conclusion

This research aimed to illuminate marker perceptions about the change to marking handwritten responses onscreen, as opposed to marking typed digital responses.

The intentions for this research and accompanying outcomes are described in the following table.

Research intention	Outcome (Met/Not Met)
1. Learn about the impact on scoring and identifiable marker preference for responses that have been handwritten, typed, or handwritten and then scanned for marking; and the reliability of scores given to responses – handwritten and scanned.	**Partially met.** From the participants' perspective there was an unexpectedly positive response to marking onscreen because of the reduction in administrative effort to manage paper responses (section 5.1). However, due to the limitations of the research noted in 5.5.3, the reliability of marking at the level of individual scores could not be assessed.
2. Establish whether, or to what extent, OSM affects marking decisions.	**Met.** Markers did not report being influenced by the response mode when determining marks.
3. Capture the experience of markers to enable an understanding of any impact that has on the scores they award.	**Met.** Markers' positive experience with OSM aligned well with their expressed preferences as to how they go about marking. Distribution curves for results of 2020 and 2021 handwritten and typed responses are closely aligned, indicating that the benchmarking processes used in marking of handwritten and typed responses respectively produce similar outcomes. This supports the research team's view that the RM Assessor standardisation functionality used for benchmarking scanned and digital responses performs as reliably as NZQA's traditional benchmarking process used for paper responses
4. Consideration of, and recommendations on, whether subsequent research should be conducted using a different standard that has multiple item questions rather than single item questions.	**Met.** Refer section 7.

While participants were experienced markers for both paper and digital responses, marking handwriting onscreen, as well as marking digital and handwritten responses together was an entirely new approach.

The process of standardisation within RM Assessor for OSM of both handwritten and typed responses produced a similar outcome to benchmarking processes for paper and digital benchmarking. This

should give markers confidence in achieving comparable marking outcomes through OSM irrespective of marking mode.

Participants' perceptions about OSM were consistent with those reported in the literature, with analysis of group interview transcripts showing greater positivity after the participants had completed the marking trial. Because of the small size of the group, it was not possible to determine correlations between marker age, gender and experience and their preference for OSM.

Participant comments were also consistent with the literature in relation to cognitive demands of onscreen reading (cf Campbell 2005), whereby they reported concerns prior to OSM about viewing responses onscreen. However, after experiencing OSM participants reported that the shift between handwritten and typed responses created a cognitive break that was a positive experience.

Consistent with (Shaw 2008), participants commented on challenges about navigating through long typed texts and following the thread of handwritten texts that were not sequential. However, as participants became familiar with OSM, their average marking time decreased by around 20%.

Participants identified the greatest positive benefits related to removing the administration associated with paper responses, similar to Coniam (2011). Most of the negative aspects of marking reported by participants related to paper-based processes, such as administration, security, distribution delays, and checkmarking paper responses. The ability to collaborate in real-time when needing guidance from other panel members was seen as a great advantage of OSM.

More NCEA subject achievement standards are currently offered as paper-based rather than digital assessments. Consequently, a larger number of markers will be scoring handwritten responses online than scoring both handwritten and typed responses onscreen, and will receive the benefits of OSM, including improved efficiency, flexibility, and a reduction of administration effort.

To enable these benefits to be experienced, induction and support for markers new to OSM will need to be carefully managed.

7.   Recommendations

The recommendations from this research report are:

1. Give markers access to a range of learning and development opportunities to support their effective use of RM Assessor for OSM.

    Learning and development opportunities could include:

    - user guides

    - online learning

    - face to face learning

    - an online environment to practise using RM Assessor.

2. Make available a range of handwritten responses from both text and symbol-based subjects to help with OSM practice.

3. When inducting markers, share the benefits of OSM described in this report.

4. Publish this final report on the NZQA website (Research, Innovation and Enhancements page) so we can share markers' positive experience of OSM.

5. Undertake further research on OSM outcomes using the data from 2021 onwards to more thoroughly explore the implications of marking scanned paper and digital responses using this mode of marking.

8.   Bibliography

Board of Studies, Teaching and Educational Standards, NSW, Monday 22 Mar 2010, 'Onscreen marking', *BOSTE News*, Accessed from https://news.bostes.nsw.edu.au/blog/2010/3/22/onscreen-marking

Coniam, D., & Yan, Z. (2016). A comparative picture of the ease of use and acceptance of onscreen marking by markers across subject areas. British Journal of Educational Technology, 47(6), 1151-1167. doi: http://dx.doi.org.helicon.vuw.ac.nz/10.1111/bjet.12294

Coniam, D., & Yeung, S. A. (2010). Markers' perceptions regarding the onscreen marking of liberal studies in the hong kong public examination system. Asia Pacific Journal of Education, 30(3), 249-271. Retrieved from https://search-proquest-com.helicon.vuw.ac.nz/docview/762466011?accountid=14782

Charman, Melody. "Linguistic Analysis of Extended Examination Answers: Differences Between On-screen and Paper-based, High- and Low-scoring Answers." British journal of educational technology 45.5 (2014): 834–843. Web.

Coniam, D. (2009). A comparison of onscreen and paper-based marking in the Hong Kong public examination system. Educational Research and Evaluation, 15(3), 243-263. doi: http://dx.doi.org.helicon.vuw.ac.nz/10.1080/13803610902972940

Coniam, D. (2011). A qualitative examination of the attitudes of liberal studies markers towards onscreen marking in Hong Kong. British Journal of Educational Technology, 42(6), 1042-1054. doi: http://dx.doi.org.helicon.vuw.ac.nz/10.1111/j.1467-8535.2010.01136.x

Coniam, D., & Falvey, P. (2016). Validating technological innovation : the introduction and implementation of onscreen marking in Hong Kong. Springer. https://doi.org/10.1007/978-981-10-0434-6Review of Research into Onscreen Marking and a Description of the Methodological Approaches and Analytical Tools Used in the OSM Validation Studies, Pages 43-55 BACKGROUND INFO

Coniam, David. (2009). Discrepancy essays: natural phenomenon or problem to be solved? [online]. Melbourne Papers in Language Testing; v.14 n.2 p.1-31; 2009. Availability: <https://search-informit-com-au.helicon.vuw.ac.nz/fullText;dn=194324;res=AEIPT> ISSN: 1327-0311. [cited 20 Oct 20].

Coniam, D. (2013). The increasing acceptance of onscreen marking—The 'tablet computer' effect. Journal of Educational Technology & Society, 16(3), 119-129. Retrieved from https://search-proquest-com.helicon.vuw.ac.nz/docview/1515991060?accountid=14782

Johnson, M. & Nádas, R. 2009. Research Matters: an investigation into marker reliability and some qualitative aspects of on-screen marking. Cambridge, United Kingdom.

Johnson, M., Hopkin, R., Shiell, H., & Bell, J. F. (2012). Extended essay marking on screen: Is examiner marking accuracy influenced by marking mode? Educational Research and Evaluation, 18(2), 107-124. doi:http://dx.doi.org.helicon.vuw.ac.nz/10.1080/13803611.2012.659932

Johnson, M., Nádas, R. and Bell, J.F. (2010), Marking essays on screen: An investigation into the reliability of marking extended subjective texts. British Journal of Educational Technology, 41: 814-826. https://doi-org.helicon.vuw.ac.nz/10.1111/j.1467-8535.2009.00979

Kennard, C.; Arnold, L. Staff and Student Experiences of Electronic Marking and Feedback. International Journal of Assessment & Evaluation, [s. l.], v. 23, n. 3, p. 35–47, 2016. DOI 10.18848/2327-7920/cgp/v23i03/35-47. Disponível em:

http://search.ebscohost.com.helicon.vuw.ac.nz/login.aspx?direct=true&db=eue&AN=119120514&site=ehost-live Acesso em: 27 out. 2020.

Masterman, E. (2018). Typed versus handwritten essay exams: is there a need to recalibrate the gauges for digital assessment? In M. Campbell, J. Willems, C. Adachi, D. Blake, I. Doherty, S. Krishnan, S. Macfarlane, L. Ngo, M. O'Donnell, S. Palmer, L. Riddell, I. Story, H. Suri & J. Tai (Eds.), Open Oceans: Learning without borders. Proceedings ASCILITE 2018 Geelong (pp. 204-213).

Mogey, Nora; Purcell, Mike; Paterson, J.S.; Burk, John (2008): Handwriting or typing exams – can we give students the choice?. Loughborough University. Conference contribution. https://hdl.handle.net/2134/5555

Shaw, S. (2008). Essay marking on-screen: Implications for assessment validity. E-Learning, 5(3), 256-274. doi:http://dx.doi.org.helicon.vuw.ac.nz/10.2304/elea.2008.5.3.256

Tung-hsien, H. (2019). The impact of computers on marking behaviors and assessment: A many-facet rasch measurement analysis of essays by EFL college students. Sage Open, 9(2) doi: http://dx.doi.org.helicon.vuw.ac.nz/10.1177/2158244019846692

Varank, Ilhan and Dogan Tozoglu. "Why Are Teachers Resistant to Change? Key Issues and Challenges in Technology Integration." (2006).

Yan, Zi & Coniam, David. (2013). Assessing the ease of use in the environment and markers' acceptance of onscreen marking: A Rasch measurement perspective. Educational Research and Evaluation. 19. 461-483. 10.1080/13803611.2013.793604.

Yan, Z., & Coniam, D. (2014). The effects of key demographic variables on markers' perceived ease of use and acceptance of onscreen marking. Assessment in Education: Principles, Policy & Practice, 21(4), 464-480. doi: http://dx.doi.org.helicon.vuw.ac.nz/10.1080/0969594X.2014.953910

Yang, M., Yan, Z., & Coniam, D. (2017). A qualitative study of markers' perceptions on onscreen marking in five subject areas. Educational Research and Evaluation, 23(7-8), 290-310. doi: http://dx.doi.org.helicon.vuw.ac.nz/10.1080/13803611.2018.1446836

Appendix One: List of preceding Innovation Trials

2020 Innovation Trials

2020 scanning innovation trial

In 2020, we completed a small-scale scanning innovation trial. The trial validated a proof of concept supporting an end-to-end process of:

- scanning the completed paper exam booklets

- marking the completed paper exam booklets

- returning the completed paper exam booklets back to students.

2020 RM Assessor benchmarking and check marking feature (standardisation) trial

A separate, small-scale, complementary innovation trial in 2020 tested the benchmarking and check marking feature within RM Assessor (our OSM platform).

The trial found that all OSM processes could be performed in the one digital environment.

2021 Scanning Trial

This trials showed that scanning students' completed paper exam booklets is technically feasible. However, the assessment implications when scoring scanned handwritten paper response booklets – alongside digital (typed) exam responses, in the same digital environment – were unclear.

Appendix Two: Information provided to invited participants

Information: Research on marker responses to digital and scanned examination artefacts

NZQA will be scanning paper exam responses for NCEA's Level 1 – Level 3 end of year external examinations during the 2021 exam cycle. The scanned responses will be marked online, along with digital (typed) responses in the marking application, RM Assessor. NZQA refers to this as onscreen marking (OSM).

You have been invited to participate in a marking trial at the end of April 2021, to ensure that all of marking panel processes can be completed successfully after OSM of scanned and digital scripts. At the same time, we are researching how markers respond to onscreen marking.

What is the aim of the project?

The aim of the project is to investigate markers' responses to onscreen marking of scanned handwritten and digital (typed) student scripts to one English Level 2 standard. The research will involve:

- a focus group type discussion with markers before, and after, the scripts have been marked to elicit your perceptions and experiences of marking.
- comparing the scores from the marked trial scripts with the original scores from the paper and digital scripts completed in 2020.

This research was approved by the NCEA Online Programme Steering Committee in March 2021.

How can you help?

You have been invited to participate because you are a member of the marking panel for the scanning and standardisation trial.

If you agree to take part, you will participate in a focus group at the beginning, and at the end of the standardisation trial, at NZQA's National Office in Wellington. The focus group will take up to 90 minutes. We will audio or video record the focus group discussion, with your permission, and write it up later. You can choose to not answer any question, or stop participating at any time, without giving a reason. You can withdraw from participating in the research by contacting the researchers at any time before 24 April 2021.

What will happen to the information you give?

This research is confidential. This means that the researchers named below will be aware of your identity, but the research data will be combined, and your identity will not be revealed in any reports, presentations, or public documentation. While your identity will be kept confidential, there is a possibility that you may be identified by people you have worked closely with.

Only the named researchers will read the notes or transcripts of the focus groups, and we may discuss them. The focus group transcripts, summaries and any recordings will be kept securely and in confidence.

What will the research produce?

The information from this research will be used for communications, change management, and guidance instructions for markers. The final report may also be used for academic publications and conference presentations.

If you accept this invitation, what are your rights as a research participant?

You do not have to accept this invitation if you don't want to. If you do decide to participate, you have the right to:

• choose not to answer any question;

• ask for the recorder to be turned off at any time during the interview;

• withdraw from the study before 24 April 2021;

• ask any questions about the study at any time;

• receive a copy of the focus group interview recording or transcript*;*

• read any reports of this research by emailing the researchers to request an electronic copy.

If you have any questions or problems, who can you contact?
If you have any questions, either now or in the future, please feel free to contact either:

Researchers:

Rose Cole, Strategic Change Manager rose.cole@nzqa.govt.nz

Annette Edwards-Hill, Senior Business Analyst Annette.edwards-hill@nzqa.govt.nz