

Dual modes for written assessments: Examining validity and reliability

Jan Eyre

<http://dx.doi.org/10.18296/am.0009>

Abstract

Computer-based testing is on the increase, especially for large-scale assessments where there are advantages for manageability, administration, and reporting. In Aotearoa/New Zealand, NZQA has announced plans to move NCEA examinations to an online platform, with a dual-assessment model available for selected levels and subjects from 2016. The dual-assessment approach appears to reflect an assumption that results in either mode are comparable, and that both approaches test the same set of skills. However, there is evidence that differences between computer- and paper-based modes exist, and that these have the potential to compromise the validity and reliability of these dual assessments. This article argues that these differences, especially when students are required to demonstrate their understanding through written responses, have the potential to create an unfair situation that disadvantages some groups of students. This may lead to a widening of the gap between our lowest and highest achieving students.

Introduction

The New Zealand Qualifications Authority (NZQA) has announced its intention to move towards “assessment anywhere, anytime, online and on demand”, with paper examinations becoming “a thing of the past” (NZQA, 2014a). As a first step towards digital assessment, NZQA plans to offer a dual-assessment model for selected external assessments in 2016. This will allow a choice between paper-based or online versions of the assessment, both involving the same set of questions. It is intended as an interim measure until all external assessments move completely online.

NZQA’s move to online assessment for NCEA appears to be a response to the pervasiveness of technology both globally and locally. Here in Aotearoa/New Zealand, many of us use technology routinely to accomplish everyday tasks, both at home and at work (Gibson, Miller,

Smith, Bell, & Crothers, 2013). Our use of mobile internet services doubled over the period 2009–2012 as we increasingly used devices such as laptops, smartphones, and tablets to connect with the web “on the go”. Much of our communication is now through screen-based digital media rather than traditional print-based forms (Warschauer, 2006). For many students in our schools, technology is an integral part of their lives; they are “digital natives” (Prensky, 2001), having grown up with technology-based communication as the norm. Such students are likely to spend significant amounts of time out of school “accessing the internet, learning and exchanging new information—often via their social networks” (Johnson et al., 2013, p. 2).

Against this backdrop, it is natural that schools are using a range of technologies, accessed through a variety of digital devices, to support and enhance teaching and learning (Leeson & Hattie, 2009). There is also interest in exploring how assessment can make use of the digital environment, both to harness the benefits of technology and to match the digital learning experiences of students. However, schools differ in the extent to which they have incorporated technology-based learning and in their readiness to move to an online assessment platform. NZQA has recognised this by offering schools the choice between online and paper-based assessment for external examinations. Before introducing this dual-mode approach, however, it is important to consider how it might impact on validity and reliability of assessments. Would a student’s score remain constant, regardless of the mode? Do the two modes test the same skills? Is there potential for some students to be disadvantaged by a particular mode?

Studies that compare computer-based and paper-based assessments provide some insight into these questions and highlight potential impacts on validity and reliability. This article will draw on existing research to argue that there are likely to be differences in students’ scores in online and paper-based modes, particularly for assessments that require written responses. These differences, related to students’ levels of digital literacy, compromise the validity of the dual-assessment model. Further, the differences between online and paper-based modes mean there is the potential to widen the gap between our lowest and highest achieving students if the move to online assessment is not carefully managed.

Computer-based testing is a first step towards transformative use of technology

The New Zealand government clearly believes in the potential for ICT to transform teaching and learning. *The New Zealand Curriculum* outlines the expectation that schools will use ICT in innovative ways: “Schools should explore not only how ICT can supplement traditional ways of teaching but how it can open up new and different ways of learning” (Ministry of Education, 2007, p. 36). Part of this exploration involves considering how technology can be used to support assessment of learning.

The use of technology for assessment can be represented as a continuum, with computer-based testing at one end and assessment that is seamlessly embedded in learning tasks at the other (Redecker & Johannessen, 2013). Bunderson, Inouye and Olsen forecasted four distinct stages in the development of computerised assessment: computerised testing, computerised adaptive testing, continuous measurement, and intelligent measurement (1989, cited in Redecker & Johannessen, 2013, p. 81). The first of these stages, computer-based testing, involves transferring traditional, paper-based tests to a computer platform. This is seen as an inevitable first step in the journey towards a new paradigm in which technology will support new, dynamic, and interactive ways of assessment that are not possible with traditional pencil and paper (Leeson & Hattie, 2009).

While widespread use of innovative models of assessment is some way off, computer-based testing is relatively well-established and growing. Computers offer particular advantages for the manageability and administration of large-scale, standardised testing (Maguire, Smith, Brallier, & Palm, 2010). Computer-based tests can reduce the time needed to set up, grade, and report on assessments. They can be used in flexible ways, offering “on demand” assessment at times to suit individual students and educators (DeSouza & Fleming, 2003). Marking may be done automatically, especially with formats such as multiple-choice. Many computer-based tests offer a range of automatic reporting options, with data able to be exported to or integrated with other systems (Leeson & Hattie, 2009). When scoring and reporting is automated, feedback

to students, educators and other stakeholders is available promptly on completion of the test and can be given in a range of formats (Zandvliet & Farragher, 1997).

The advantages that computers offer for manageability, administration, and reporting have seen many large-scale, standardised assessments being transferred to an online platform. As a transitional stage to cater for those who do not have the capability for online assessment, some assessments are offered with a choice between paper-based and online assessment modes (“dual-mode” assessment). Dual-mode assessments include tools already used in Aotearoa/New Zealand schools and tertiary organisations, such as the e-asTTle tests, the Progressive Achievement Tests (PATs), and the Literacy and Numeracy for Adults Assessment Tool. In these assessments, teachers may create the assessment in either paper-based or online mode. Both options draw from the same questions (items) and are scored against the same scale. The underlying assumption appears to be that the mode of delivery does not affect student performance: both versions are deemed to assess the same skills and abilities, with student scores in either mode located on the same scale. Studies that compare the results of assessments offered in both paper-based and computer-based modes will help establish whether this assumption is correct, and whether there are impacts on validity and reliability.

Computer-based assessments may not measure the same skills as paper-based assessments

For an assessment to be effective, it needs to be valid and reliable. To be valid, an assessment task should provide robust information about the specific area of learning in question. Darr (2005a, p. 56) draws on the work of Wiggins (1998, p. 32) to suggest the following two questions as helpful in establishing an assessment’s validity:

- Could the student do well at the task for reasons that have little to do with the desired understanding or skill being assessed?
- Could the student do poorly at the task for reasons that have little to do with the desired understanding or skill?

To illustrate this point, Darr gives the example of a mathematics test where the reading demands of questions present obstacles for less-able readers.

In this case, validity is compromised because reading comprehension is being assessed either instead of, or alongside, ability in mathematics.

To be reliable, an assessment should provide consistent results. The measurement of student achievement it provides should be stable, regardless of the time at which the assessment is taken, the tasks that are used, and the person (or machine) that is doing the marking (Darr, 2005b). An assessment might be said to be reliable if equivalent assessments (of the same area of learning) produce similar results. We can apply these concepts of validity and reliability to the issue of dual-mode assessments by drawing on research that considers “mode” effects in assessment.

Some studies have shown that identical paper-based and computer-based tests may not obtain the same results. Clariana and Wallace (2002) compared computer-based and paper-based test scores for a multiple-choice test for first-year tertiary students in a business programme in the USA. The computer-based group (54 students) scored a mean of 83 percent (a B overall), while the paper-based group (51 students) scored a mean of 76 percent (a C overall). The differences were greatest for higher achieving students. The authors suggest that such a difference leads to an unfair assessment with low validity, and that measures should be put in place to remedy this: “test developers must show that computer-based and paper-based test versions are equivalent, and/or must provide scaling information to allow the two to be equated” (Bugbee, 1996, cited in Clariana & Wallace, 2002, p. 601).

Even where studies have found little difference in the scores of equivalent paper-based and computer-based assessments, the need for careful analysis of results is signalled. For example, Schroeders and Wilhelm (2011) found no significant differences between scores on a test of reading and listening comprehension for German high-school students who were learning English as a foreign language. However, the computer-based version of this test had been carefully designed to minimise potential sources of difference identified by previous research. For example, shorter passages of text were selected to eliminate the need for scrolling, and only multiple-choice items were used. The authors of this study concluded that under different conditions (for example, if the test had required written responses), different results may have been obtained. They see “no theoretical or empirical framework that guarantees that measures would

be invariant across test media” (p. 866). As each test situation is different, careful statistical analysis is always required to establish validity and equivalence between modes.

One of the few studies that involved younger children was conducted as part of a project in Hungary that aimed to develop online diagnostic assessment for primary-school students. Molnar, Toth and Csapo (2011) conducted a comparability study with a large group (5,000+) of students aged 6–7 years. The students took a paper-based test of inductive reasoning. Two years later, a similar group took a computer-based version, designed to be as similar as possible to the original test. Students who took the paper-based test performed significantly better than those who completed the test on the computer, with the equivalent of a year’s difference in their results: “students in the first grade achieved on (paper-based) test like second graders ... in an online environment” (p. 5). Students who took the computer-based test were therefore significantly disadvantaged. The authors concluded that ancillary factors related to computer use were interfering with the test’s validity.

The studies mentioned in this section have identified potential validity and reliability issues for dual-mode assessments that use a multiple-choice format. They suggest that multiple-choice assessments must be carefully constructed to minimise mode effects, taking into account relevant research on test design. Such assessments will also require trialling followed by thorough statistical analysis of the results, to establish equivalence between the two modes. If equivalence cannot be established, scaling information will need to be developed so that scores in either mode can be compared.

The potential mode effects identified for multiple-choice formats are likely to be amplified for assessments that require written responses, such as external NCEA examinations. When these are offered online, students have to compose (“construct”) their responses on computer, rather than on paper. Typical questions ask students to demonstrate their understanding by explaining, discussing, describing or justifying their ideas (NZQA, 2014b). This requires use of academic writing skills, including the ability to compose answers with a clear, logical structure; concise, relevant content; correct grammar; and appropriate vocabulary, including academic and subject-specific vocabulary. Such expectations for written work are signalled in the writing standards for Years 1–8 (Ministry of

Education, 2009). This kind of written assessment is a “step up” from computer-based tests that rely on multiple-choice responses, representing another shift along the continuum of computer-based assessment.

Concerns about possible mode effects have guided decisions in the United States about national assessments that require written responses. With reference to the National Assessment of Educational Progress (NAEP), which aims to provide nationally representative information about students’ achievement in core subjects such as mathematics, reading, and writing, Horkay, Bennett, Allen, Kaplan and Yan (2006) observed that:

in most ... state online testing programs, multiple-choice items are exclusively used because the test delivery software for their presentation is more evolved than that for constructed-response delivery and because of concern that some groups of students would be unfairly disadvantaged by having to answer constructed-response questions on computer. (p. 4)

Technology has continued to evolve over the years since this study, with automated assessment of writing becoming more commonplace. The NAEP writing assessments for Grades 8 and 12 have already moved to online delivery, and the assessment for Grade 4 students (9 year olds) will follow “in the near future” (National Assessment Governing Board, 2010). However, concerns about the validity and reliability of online written assessments remain.

A computer is a tool that can enhance academic writing

As a first step in considering possible validity and reliability issues associated with dual-mode written assessments, it is important to consider how academic writing composed on a computer might differ from academic writing that is composed on paper.

Writing on a computer or other digital device usually incorporates aspects of word processing such as the ability to move, copy, paste, delete, and insert text. Often, there is access to tools such as a spelling and grammar check, or an online thesaurus. There has been a great deal of interest in how these functions of word processing can support the teaching of writing. Relevant research is summarised in three meta-analyses carried out between 1993 and 2007. All three meta-analyses found that word processing has the potential to lift the quality of students’ writing.

The first, by Bangert-Drowns (1993), analysed 33 studies that provided quantitative data on the effects of word-processing on the writing process. In these studies, two groups of students received the same instruction in the writing process, except that one group was allowed to use a word-processor while the other group wrote by hand. The majority of the studies were concerned with college students, although a few presented results from younger students in elementary or pre-school settings. The meta-analysis focused on the impact of word processors on writing quality, writing length, students' attitudes towards writing, and frequency of revision.

The results of this meta-analysis showed that word-processing had a "small positive effect" on writing quality, especially for basic or lower achieving writers (Bangert-Drowns, 1993, p. 7). There was also a positive effect on length of writing. Length and quality of writing were positively related, as confirmed by other, more recent studies (Morphy & Graham, 2012; Lovatt, Lewandowski, Berger, & Gathje, 2010). Impacts on attitude to writing and frequency of revision were inconclusive, although the author suggested that lower achieving writers find writing on a computer engaging, which may contribute to increases in the quality of their work. The motivational aspect of computer-based writing for struggling writers has been reported in other studies (Boyd & McDowall 2001; Morphy & Graham, 2011). However, to maximise any potential benefits, students must be given direct instruction in how to use the features of word processing. Access to word processing is not sufficient: students need to be taught to use it as a tool at all stages of the writing process (MacArthur, 2009). While students today probably have more online experience than previous groups, we cannot assume that this means they are more skilled in composing writing in digital formats.

Building on Bangert-Drowns' work, Goldberg, Russell, and Cook (2003) used both quantitative and qualitative techniques to analyse 26 studies that compared writing produced by word processor with paper-based writing. The studies, published between 1992 and 2002, concentrated on students in primary and secondary (K-12) education. The focus was on the impact of word-processing on quantity of writing, quality of writing, and number of revisions. As with the earlier study by Bangert-Drowns, the analysis showed a positive effect on both length and quality of student writing; these effects "tended to be larger for middle and high school students than for elementary students" (p. 14). Partly because there were only a small

number of relevant studies, it was not possible to estimate the effect of word processing on the number of revisions. However, Lenhart, Arafeh, Smith, & Rankin Macgill, in a 2008 United States-based survey of 12 to 17 year olds' attitudes to writing and technology, found that most students appreciate the ease with which they can revise and edit text on a computer, and believe that using computers increases the likelihood that they will do so.

Finally, in a report to the Carnegie Corporation, Graham and Perin (2007) conducted a meta-analysis of studies of adolescents' writing. Their goal was to identify specific teaching strategies that research suggested would be effective in lifting the quality of student writing. Word-processing was one of the eleven strategies identified, and was found to have a "consistently positive impact on writing quality", with a greater impact for low-achieving writers (p. 17).

The meta-analyses described above confirm the potential for computer-based writing to be qualitatively and quantitatively different from paper-based writing. Word processing is a tool that can enhance writing, and students who are able to use it effectively can produce longer texts and texts of higher quality than they would on paper. These findings have implications both for the teaching of writing and for the validity and reliability of assessments that offer a choice of composing text either by hand or on the computer. We can now compare these findings with studies that look specifically at dual-mode written assessments.

Experience with computers is related to level of achievement in online written assessments

Horkay et al. (2006) investigated the implications of electronic delivery for a writing assessment that was part of the NAEP in the United States. The focus of the investigation was the "comparability and fairness of scores" (p. 4) across paper-based and computer-based versions of the assessment. For the study, a nationally representative group of 2,878 eighth-grade (13–14 year old) students first took a paper-based version of either the writing assessment or a reading assessment. At least 3 weeks later, approximately half these students completed an online version of the writing assessment. Both versions of the writing assessment required students to complete two essays.

Before taking the computer-based test, students completed a short tutorial, which covered basic navigation through the assessment and offered practice in using the word-processing functions. Students' levels of computer familiarity were also measured. Computer familiarity was divided into two aspects. The first was general experience with computing (defined as "extent of computer use") and the second was hands-on computer-proficiency (defined as keyboarding speed, keyboarding accuracy, and editing skill, including command of basic word-processing functions). Measuring these two aspects separately acknowledged that students may have extensive experience with computers without being either fast or accurate keyboarders.

The results of the writing assessments were analysed to see if there were differences in student performance between the two modes, whether type of computer (school computer or NAEP-supplied laptop) affected performance, and whether students who were relatively unfamiliar with computers performed differently from students who were more familiar with them. Results showed that computer familiarity had the most significant effect on students' performance.

The impact of computer familiarity

The level of students' computer familiarity, in particular hands-on computer proficiency, had a significant effect on students' overall writing score when the assessment was completed online:

Hands-on skill was significantly related to online writing assessment performance: students with greater hands-on skill achieved higher WOL (writing online) scores, holding constant their performance on a paper-writing test. Computer familiarity added about 11 percentage points over the paper writing score to the prediction of WOL performance. (p. 35)

Thus, students were likely to score significantly higher in the online mode than the paper-based mode if they had high levels of hands-on computer skills, such as fast and accurate typing and ability to use basic word-processing functions.

In discussing the implications of their investigation, Horkay et al. (2006) proposed that students who are familiar with using computers for writing, and particularly those who have received instruction on using the

computer to assist the writing process, would be able to produce writing of higher quality than the writing they could produce on paper. This finding aligns with the meta-analyses of research into word processing discussed above, and has implications for the validity and fairness of assessments that require students to write online:

it is conceivable that, for a given level of paper writing performance, students with greater computer facility score higher on WOL [writing online] because they write better on computer than on paper (relative to their less technologically-experienced peers). And, the reason they write better on computer than they do on paper may be because the computer offers them a tool that makes it possible to do so ... students with little practice writing on computer will not score as highly in an online writing test as their peers who word-process routinely. And that lower relative performance will not necessarily be because the former students are less skilled writers, but because they are less skilled writers on computer. (p. 36)

A similar finding was reported by Chen, White, McCloskey, Soroui, and Chun (2011), who investigated mode effects in an assessment of adults' functional writing skills (defined as writing to achieve everyday goals). This investigation used data collected as part of the development process for the National Assessment of Adult Literacy (NAAL) in the United States. The participants, who were rated as low, medium, or high users of computers on the basis of their responses to a background questionnaire, completed three writing tasks either on paper or on computer. Before completing the computer tasks, the participants had the option of completing a short tutorial on how to navigate through the assessment. Chen et al. found that the adults performed better on paper than on computer. Level of computer experience was related to level of writing performance: those with a low level of computer experience scored significantly higher when writing on paper than when writing on a computer, meaning that they "were more disadvantaged by the computer mode" (p. 67). Although concerned with adults (aged 16–65+) rather than school students, and with functional writing rather than academic writing, this study lends weight to the finding from the Horkay et al. (2006) study that an individual's ability to compose on computer is affected by his or her level of familiarity and experience with computers.

The influence of computer familiarity on the quality of student writing has implications for the validity levels of dual-mode assessments. As Horkay et al. (2006) point out, paper-based assessments, no matter what the subject content, provide information about how well students write on paper. Similarly, computer-based assessments provide information about how well students write on computer. As we have seen, writing on computer involves a particular skill set, and paper-based and computer-based assessments of writing offer “qualitatively different experiences” (McDonald, 2002, p. 309). Assuming that the two are equivalent compromises validity and results in an unfair assessment if either mode can be used.

Mode effects on the spread of scores

Computer-based writing may show more variability, both in length and quality, than paper-based writing (Horkay et al., 2006; Chen et al., 2011), resulting in a wider spread of scores. A wider spread does not necessarily affect the mean score in either mode, but it does have the potential to extend the gap between low- and high-achieving students. Students who are not familiar with computers may achieve at lower levels than they would on paper, and those who routinely use computers may achieve at higher levels. The distance between high and low achievers is likely to be greater in the computer-based mode than in the paper-based mode when assessments involve written responses.

Mode effects on marker reliability

Some studies have found that mode of presentation, handwritten or typed, affects the score given to a written assessment. Higher scores may be given to writing composed by hand (Russell & Tao, 2004a and 2004b, cited in Horkay et al. 2006). Other studies have reported no significant differences in scoring, although Lee (2004) found that computer-based responses were scored more consistently between markers than handwritten responses. Bias in scoring affects assessment reliability, and so must be carefully considered when assessments of writing are offered in dual modes. It may be possible to eliminate bias through careful training of the markers (Russell & Tao, 2004a and 2004b, cited in Horkay et al., 2006). Use of analytic rather than holistic scoring methods, such as rubrics that focus on individual elements of the task, may also help to reduce bias (Chen et al., 2011).

Match between assessment mode and classroom experience

One measure of an assessment's validity is the extent to which it matches the students' classroom learning experience. Where there is a mismatch between the assessment activity and the nature of the learning (and assignments) that students have completed in the classroom, the assessment could be said to have lower validity than one that more closely matches students' learning experiences.

The validity of a written assessment is affected by the extent to which it matches students' usual writing practices. If students have completed their classroom learning using paper and pencil, then using a computer to assess this learning has a lower level of validity than a paper-based assessment. The opposite case is also true. In a white paper completed for Pearson Assessment, part of a global, commercial educational publishing and assessment service, Way, Davis, and Strain-Seymour (2008) make a case for the validity of computer-based assessments of writing. They argue that as students' use of technology is increasing, both in and out of school, and as writing is increasingly being taught through the medium of computers, then computer-based assessment of writing has higher validity than paper-based assessment. To establish whether this is true in Aotearoa/New Zealand, we need to examine the extent to which schools make use of computers to support their writing programmes.

Use of computers or other digital devices as part of the writing process varies from school to school and teacher to teacher in Aotearoa/New Zealand. *The New Zealand Curriculum* (Ministry of Education, 2007) outlines a general expectation that schools should use ICT as part of effective pedagogy, but makes no explicit statements on its use to support the writing programme. Associated curriculum documents such as the National Standards (Ministry of Education, 2009) and the Literacy Learning Progressions (Ministry of Education, 2010) outline expectations for student performance in writing at different stages of schooling from Year 1 to Year 10. Again, the use of technology is not specified; where annotated exemplars of student writing at each level are provided, the writing is composed by hand. It is likely that such documents influence teachers' expectations for written work. In some classes, particularly at primary and intermediate level, it is likely that computers and other

digital devices may be mainly used for publishing writing that has already been composed on paper (Boyd & McDowall, 2001).

Until computer-based writing becomes the norm in Aotearoa/New Zealand classrooms, it is important that teachers and students are able to choose the mode of assessment that most closely matches their classroom experience.

Computer-based assessments have the potential to widen the achievement gap

There is a view that that our students are “digital natives”, who live and breathe technology (Tapscott, 2012, cited in NZQA, 2014a). However, the reality is that students in Aotearoa/New Zealand have differing levels of experience with and access to technology, and this differential access is related to socioeconomic status.

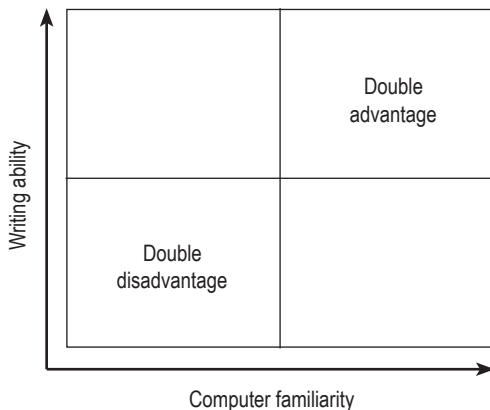
The phrase “digital divide” has been used to describe the gap between those who have “ready access to the tools of information and communication technologies, and the knowledge that they provide access to, and those without such access or skills” (Cullen, 2001, p. 1). In Aotearoa/New Zealand, lower household incomes are associated with lower use of the internet (Statistics New Zealand, 2012; Gibson et al., 2013). Twenty-five percent of New Zealanders in low-income households (less than \$35,000 a year) do not use the internet. Non-use of the internet is higher among Māori and Pasifika New Zealanders than for NZ Europeans and Asians: “even *amongst* non-users, Māori and Pasifika appear to be more digitally disadvantaged than NZ Europeans” (Gibson et al., 2013, p. 32). If digital disadvantage impacts on computer familiarity as defined by Horkay et al. (2006), there is the potential for Māori and Pasifika students to be over-represented in the group that will be adversely affected by computer-based written assessments.

Those who have low levels of computer familiarity and are also struggling writers are likely to be doubly disadvantaged in computer-based assessments of writing. In 2012, the National Monitoring Study of Student Achievement (NMSSA) focused on student achievement in writing across Aotearoa/New Zealand at Years 4 and 8. The results showed that writing achievement was “lower for Māori and Pasifika students respectively, and was lower

for students from lower decile schools” (Ministry of Education, 2013, p. 9). These results are consistent with findings reported by the National Education Monitoring Project (NEMP) in 1998, 2002, and 2006. Students from Māori, Pasifika and lower socioeconomic groups are over-represented both amongst lower achieving writers, and amongst those who are likely to have low levels of computer familiarity. It is therefore likely that the performance of Māori, Pasifika and those from lower socioeconomic groups will be significantly affected by the computer-based mode for written assessments. Those with already low levels of performance may achieve at even lower levels in the computer-based mode, for reasons that are not associated with their general academic writing ability, but are related to their level of experience with computers.

The likely interplay between computer familiarity and writing skills is represented in Figure 1. Students who are in the lower-left quartile, with low levels of both computer familiarity and writing ability, are likely to be doubly disadvantaged by an online mode. Those in the upper-right quartile, on the other hand, are likely to be significantly advantaged.

Figure 1. Computer familiarity and writing ability: Impacts when assessment is online



Preparing for the move to online assessment

It is inevitable that assessment will eventually move to digital platforms as technology becomes the norm for teaching and learning. However,

paper-based and online modes may test different skills. For example, an online written assessment such as an NCEA external examination is likely to involve academic literacy, subject knowledge, and digital literacy, whereas a paper-based assessment involves only the first two of these. To increase validity and reliability, it will be important to carry out thorough trials in both modes before implementing the assessments. This will enable student performance in each mode to be compared at individual question level and separate reporting scales developed if necessary.

The move towards digital assessment signals an assumption that digital literacy is an important skill for our students. If this is the case, then it is important to make sure that digital literacy is firmly embedded in the curriculum. The current *New Zealand Curriculum* (Ministry of Education, 2007), while stating that schools should explore how ICT can support and enhance learning as part of effective pedagogy, does not position it within any one particular learning area. Schools and individual teachers have freedom to decide how they will incorporate technology in their teaching and learning. However, if the expectation is that students will be completing assessments in an online medium by their middle high school years, it would seem important to begin instruction in the necessary skills at primary school. At the least, it would be useful to have exemplars of computer-based writing in key documents such as the National Standards and Literacy Learning Progressions, to communicate a clear message about the importance of digital literacy.

Given the importance of what Warschauer (2006) has described as “computer literacy” and Horkay et al. (2006) describe as “computer familiarity”, it would seem important that we ensure students develop these skills in a systematic way. One way of doing this would be to incorporate computer literacy in the teaching of writing skills, giving students specific instruction in how technology can be used at all stages of the writing process from planning and drafting through editing and proofreading to final publication. This has the potential to lift the achievement of struggling writers. It may also be important to consider how we help students to develop the fast and accurate text entry skills that are part of hands-on computer expertise. NZQA (2014a) claims that, generally, students “can type faster than they write”—but this is something that we should not leave to chance. Teaching keyboarding skills alongside handwriting is one possible solution.

Technology is a powerful tool with the potential to transform teaching and learning. Many believe it also has the potential to reduce educational disparities. As Warschauer, 2006, observed:

Many believe that technology in schools can ... be a lever for overcoming inequality between students in rich and poor communities, by giving low-SES students powerful learning tools that can help them leapfrog out of educational disadvantage (p. 5).

Technology offers a tool for writing that, with direct instruction, can help all students achieve at higher levels. However, transferring large-scale, summative assessments to online platforms without thoroughly investigating the possible mode effects raises questions about validity and aspects of reliability and may create unfair situations that will widen the gap between low and high-achieving students. It is important that we take time to research and evaluate the way we implement technological changes. In this way, we will ensure that the benefits for improved administration and manageability do not override principles of valid and fair assessment.

References

- Bangert-Drowns, R. L. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational Research*, 63(1), 69–93. <http://dx.doi.org/10.3102/00346543063001069>
- Boyd, S., & McDowall, S. (2001). *Techno-magic: Whizz or fizz? The relationship between writing mode, editing process, and writing product*. Wellington: New Zealand Council for Educational Research.
- Bunderson, C. V., Inouye, D.K., & Olsen, J. B. (1989). The four generations of computerized testing. In R. Linn (Ed.), *Educational measurement* (pp. 367–407). New York, NY: Macmillan.
- Clariana, R., & Wallace, P. (2002). Paper-based versus computer-based assessment: Key factors associated with test mode effect. *British Journal of Educational Technology*, 33(5), 593–602. <http://dx.doi.org/10.1111/1467-8535.00294>
- Chen, J., White, S., McCloskey, M., Soroui, J., & Chun, Y. (2011). Effects of computer versus paper administration of an adult functional writing assessment. *Assessing Writing*, 16, 49–71. <http://dx.doi.org/10.1016/j.asw.2010.11.001>
- Cullen, R. (2001). Addressing the digital divide. *Online Information Review*, 25(5), 311–320. <http://dx.doi.org/10.1108/14684520110410517>

- Darr, C. (2005a). A hitchhiker's guide to validity. *Set: Research Information for Teachers*, 2, 55–56.
- Darr, C. (2005b). A hitchhiker's guide to reliability. *Set: Research Information for Teachers*, 3, 59–60.
- DeSouza, E., & Fleming, M. (2003). A comparison of in-class and online quizzes on student exam performance. *Journal of Computing in Higher Education*, 14(2), 121–134. <http://dx.doi.org/10.1007/BF02940941>
- Gibson, A., Miller, M., Smith, P., Bell, A., & Crothers, C. (2013). *The internet in New Zealand 2013*. Auckland: Institute of Culture, Discourse & Communication, AUT University.
- Goldberg, A., Russell, M., & Cook, A. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002. *The Journal of Technology, Learning, and Assessment*, 2(1). Retrieved from <http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1661/1503>
- Graham, S., & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools*. Washington, DC: Alliance for Excellence in Education.
- Horkay, N., Bennett, R. E., Allen, N., Kaplan, B., & Yan, F. (2006). Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *Journal of Technology, Learning and Assessment*, 5(2). Retrieved from <http://www.jtla.org>
- Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., & Ludgate, H. (2013). *NMC horizon report: 2013 K-12 edition*. Austin, TX: The New Media Consortium.
- Lee, H.K. (2004). A comparative study of ESL writers' performance in a paper-based and a computer-delivered writing test. *Assessing Writing*, 9(1), 4–26. <http://dx.doi.org/10.1016/j.asw.2004.01.001>
- Leeson, H., & Hattie, J. (2009). *The use of technology for assessment* (Commissioned Review Paper 12). Auckland: Visible Learning Labs, Faculty of Education, University of Auckland.
- Lenhart, A., Arafeh, S., Smith, A., & Rankin Macgill, A. (2008). *Writing, technology and teens*. Washington, DC: The National Commission on Writing.
- Lovatt, B. J., Lewandowski, L. J., Berger, C., & Gathje, R. A. (2010). Effects of response mode and time allotment on college students' writing. *Journal of College Reading and Learning*, 40(2), 64–79. <http://dx.doi.org/10.1080/10790195.2010.10850331>
- MacArthur, C. A. (2009). Reflections on research on writing and technology for struggling writers. *Learning Disabilities Research & Practice*, 24(2), 93–103. <http://dx.doi.org/10.1111/j.1540-5826.2009.00283.x>
- McDonald, A. S. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers and Education*, 39, 299–312. [http://dx.doi.org/10.1016/S0360-1315\(02\)00032-5](http://dx.doi.org/10.1016/S0360-1315(02)00032-5)

- Ministry of Education (2007). *The New Zealand curriculum*. Wellington: Learning Media.
- Ministry of Education (2009). *The New Zealand curriculum reading and writing standards for years 1–8*. Wellington: Author.
- Ministry of Education (2010). *The literacy learning progressions: Meeting the reading and writing demands of the curriculum*. Wellington: Author.
- Ministry of Education (2013). *Wānangatia te putanga taura—National monitoring study of student achievement: English: Writing 2012*. Wellington: Author.
- Molnar, G., Toth, K. R., & Csapo, B. (2011). *Comparing paper-based and computer-based testing in the first grade*. Paper presented at the Round Table on Uncovering the Promise and Pitfalls of Computerized and Adaptive Testing in Action, AERA, New Orleans, Louisiana, USA.
- Morphy, P., & Graham, S. (2012). Word processing programs and weaker writers/readers: A meta-analysis of research findings. *Reading and Writing*, 25(3), 641–678. <http://dx.doi.org/10.1007/s11145-010-9292-5>
- National Assessment Governing Board (2010). *Writing framework for the 2011 National Assessment of Educational Progress*. Washington, DC: Author
- NZQA. (2014a). Future state presentation to SPANZ. Retrieved from <http://www.nzqa.govt.nz/assets/About-us/News/SPANZ-presentation-2014.pdf>
- NZQA. (2014b). Subject resource pages. Retrieved from <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/subjects/>
- Powers, D. E., Fowles, M. E., Farnum, M., & Ramsey, P. (1994). Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays. *Journal of Educational Measurement*, 31(3), 220–233. <http://dx.doi.org/10.1111/j.1745-3984.1994.tb00444.x>
- Prensky, M. (2001). Digital natives, digital immigrants: Part 1. *On the Horizon*, 9(5), 1–6. <http://dx.doi.org/10.1108/10748120110424816>
- Redecker, C., & Johannessen, O. (2013). Changing assessment—towards a new assessment paradigm using ICT. *European Journal of Education*, 48(1), 79–96. <http://dx.doi.org/10.1111/ejed.12018>
- Schroeders, U., & Wilhelm, O. (2011). Equivalence of reading and listening comprehension across test media. *Educational and Psychological Measurement*, 71(5), 849–869. <http://dx.doi.org/10.1177/0013164410391468>
- Statistics New Zealand. (2012). *Household use of information and communication technology: 2012*. Retrieved from http://www.stats.govt.nz/browse_for_stats/industry_sectors/information_technology_and_communications/HouseholdUseofICT_HOTP2012/Commentary.aspx
- Warschauer, M. (2006). *Laptops and literacy: Learning in the wireless classroom*. New York, NY: Teachers College Press.

- Way, W.D., Davis, L.L., & Strain-Seymour, E. (2008). *The validity case for assessing direct writing by computer: A Pearson assessments & information white paper*. Retrieved from <http://www.pearsonassessments.com>
- Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass.
- Zandvliet, D., & Farragher, P. (1997). A comparison of computer-administered and written tests. *Journal of Research in Computing Education*, 29(4), 423–439. <http://dx.doi.org/10.1080/08886504.1997.10782209>

The author

Jan Eyre is a Senior Researcher/Resource Developer at the New Zealand Council for Educational Research.

Email: jan.eyre@nzcer.org.nz