Typed *versus* handwritten essay exams: is there a need to recalibrate the gauges for digital assessment?

Elizabeth Masterman
University of Oxford
United Kingdom

In moving from handwritten to typed essay-based examinations (e-exams), the salient issue from an academic perspective is equivalence: can exams in the two modes be considered the same? This paper reports the findings of a literature survey addressing this question, conducted prior to a trial of e-exams at a leading university. The survey sought to establish whether the move results in 1) changes in students' strategies when composing an exam answer, and the resulting product, and 2) changes in academics' perception of typed exam scripts and their strategies in marking on screen. The research team concluded that the two modes of exam are not equivalent, even though differences in the marks achieved by students may be statistically insignificant. Recommendations arising from the analysis include moving to e-exams as the sole mode; supporting students and academics to develop IT proficiency for assessment; and capitalizing on the analytics available in e-exam tools to minimize the inequities that arise when exams are conducted in a single mode.

Keywords: analytics; digital assessment; e-exams; essay-based exams; on-screen marking

Introduction

Invigilated essay-based examinations, in which students write their scripts by hand, are traditionally the primary means by which students' knowledge gain is measured in tertiary education (Moore, 2018). Where such assessments are conducted in a face-to-face setting (i.e. an examination hall or other room allocated to the purpose), there is an emerging trend towards *e-exams*, in which students type their scripts either on their own laptops or on institutionally provided devices. The benefits include reduced physical discomfort compared with handwriting, the potential to incorporate multimedia elements into questions and the streamlined management of assessment (Hillier & Fluck, 2013; Sindre & Vegendla, 2015). Furthermore, typing is now the norm for students' coursework, and so handwritten exams may be both unnatural and anachronistic (Moore, 2018).

E-exams have been standard practice in US law schools for over 20 years (Augustine-Adams, Hendrix & Rasband, 2001; Mogey, Paterson, Burk & Purcell, 2010). They are becoming commonplace in Scandinavian countries, where Sindre and Vegendla (2015) have forecast a large-scale shift towards e-exams during the coming decade. In the UK there have been isolated trials of e-exams since at least 2008 (for a summary, see Masterman & Fresen, 2017). Many initiatives remain largely at the course or departmental level (Newland & Martin, 2016), but at least one university is undertaking an institution-wide implementation (Brunel University, n.d.). In Australia, pioneering work has been carried out by Hillier, Fluck and colleagues into e-exams conducted on students' own devices (referred to in their recent publications as eExams: Fluck & Hillier, 2017).

While much attention has been paid to the practical and policy aspects of implementing e-exams, such as technology, security, resources and procedures, the salient question from an academic perspective is that of the *equivalence* between typed and handwritten exams. As Noyes and Garland observe, since two different presentation and response modes are being used there is a "need for equivalence to be determined fully to ensure that overall performance outcomes are matched" (2008, p. 1357). They suggest that this is especially the case with "non-standardised, open-ended tasks" (p. 1371), i.e. essay-based exams, in contrast to "bespoke and closed" tasks such as objective tests, which can be made more similar in both digital and paper-based modes.

This paper explores, through a survey of the currently available evidence from the research literature, whether the cognitive processes in, and intellectual outputs from, e-exams can be considered the same as for handwritten exams and, if not, whether institutions should adjust their marking schemes and other measures. To adopt a maritime metaphor, it investigates the extent to which the move to e-exams is "plain sailing" for the students sitting them and the academics marking them, or whether we need to recalibrate the existing gauges with which we navigate the high seas of assessment.

The paper starts by outlining the context in which the literature survey was conducted. It then offers a model of the writing process as an introduction to coverage of research on the process and product of writing an exam essay, before turning its attention to the influences on academics as they mark students' scripts. It then looks at the evidence for differences in students' performance between typed and handwritten exams. It concludes by reflecting on the issue of equivalence and recommending possible ways forward in practice and research.

Methodology

The context in which the literature survey was conducted was a trial of e-exams at the University of Oxford, UK. Preparatory to the trial, the project team conducted a "landscape" study of a) the research literature on the cognitive and academic aspects of changing from handwritten to typed exams (reported here), and b) reports on the software and processes adopted by other institutions running e-exams, the resources required to run exams on computers, and the practical benefits derived for both students and academics (Masterman & Fresen, 2017). The study was intended to inform the design of the trial itself, which was conducted with 35 undergraduate and 30 taught postgraduate students in three subjects in April 2018.

The survey of literature on the cognitive and academic aspects addressed three questions:
1. For students sitting e-exams: Does moving from handwritten to typed exams change the intellectual process of responding to an exam question and, hence, the length and stylistic features of the resulting response?
2. For markers of e-exams:
    2.1 Does moving from marking handwritten exam scripts to typed scripts change the marker's perception of, and attitude towards, students' responses?
    2.2 Does moving from marking exam scripts on paper to marking scripts online change the intellectual process of marking an exam question?
3. If the move results in change, is there a difference in the marks achieved/awarded, and should the differences be considered important?

The literature was gathered primarily from an intensive search of online materials, conducted in June 2017 using Google Scholar. The search terms used to locate documents include:

computer + essay + exams	digital assessments	online exams
computer-based exams	digital exams	typed exams
computerised exams	essay exams	

We initially restricted our search to works published from 2000 onwards, largely because older papers described software with outdated functionality. Also, several authors included pre-2000 studies in their literature surveys, and we often felt it unnecessary to read the cited works as well. However, we made exceptions where the findings of pre-2000 studies proved to be particularly relevant to our purposes. Additional works of interest were located in the references of articles found through the Google Scholar search, and also on journal websites which listed related items alongside the article being read.

We collected a total of 46 works, comprising peer-reviewed journal articles and full-length conference papers; non-peer reviewed literature (reports of projects, evaluations and surveys; also overviews of the field); and blog posts. We categorized them thematically according to the areas of interest identified in the research questions. The author took responsibility for the detailed reading and analysis of the materials, and her analysis was subsequently critically reviewed by her co-researcher. The search was repeated in July 2018 in order to identify relevant papers published since the original survey; five such papers were located.

Typing versus handwriting exam responses

An overview of the writing process

In order to appreciate the potential differences between handwriting and typing exam responses, an overview of the fundamental metacognitive processes involved in text composition may be helpful. Peverly's (2006) survey of models of writing competence provides a basic understanding of these processes. Under exam conditions, the processes are:
- planning (goal setting, generating and organizing content),
- retrieving knowledge,
- translating (turning ideas into text) and
- revising the text produced so far.

Translating one's ideas into text on the paper or screen entails two further processes: text generation and transcription. Text generation involves "translating generated ideas into language in working memory and then translating those temporary mental representations into more permanent external representations using the symbols of the writing system." Transcription involves "retrieving letter forms and familiar word spellings from long-term memory, strategically spelling novel words, and motor planning to produce the letters [using the tool at hand]" (Peverly, 2006, pp. 199–200).

An individual's capacity to carry out these high-level processes depends in part on the efficiency, or fluency, of the lower-level processes involved in outputting the text onto paper or screen. Summarizing a number of models, Peverly suggests that:

> Writers must (a) be fluent in generating ideas that can be written down and (b) write these ideas down quickly before they are forgotten. If writers are efficient in executing (a) and (b), they will be able to use the metacognitive processes … and other cognitive resources (e.g., genre and content knowledge…) to create reader-based prose. (2006, p. 199)

The pressurized environment of an exam may place additional pressure on the capacity of a student's metacognitive processes, with a resultant impact on the quality of their responses (Connelly, Dockrell & Barnett, 2005). Introducing technology into the exam room may generate further pressures: namely, computer anxiety (Walker & Handley, 2016) and anxiety about technical failure (Hillier, 2014).

Intellectual process and product

Insights into differences between handwriting and typing in the intellectual processes of writing an exam response are derived largely from students' self-reports. Hand-writers (i.e. students writing their exam responses by hand) in Lee's (2002) study reported that they spent more time planning their responses before starting to write than did the typists (i.e. students typing their responses on the computer). Typists reported that they composed their responses in a rough form first, then went back and expanded them; they also paused more while they were actually writing (i.e. they may have needed more time to think while producing the text because they had spent less time planning).

However, studies are inconsistent, or even contradictory regarding the differences (Lee, 2002). For example, participants in Kohler's (2015) study stated that they re-read and revised their writing more while typing than handwriting. This finding stands in contrast with the finding by Hillier (2015b), in whose study similar proportions of hand-writers and typists reported that they went back over their responses before submitting. The exact nature of differences in the process may differ from student to student (Lee 2004); furthermore, these differences may be insignificant from a methodological perspective (Mogey & Paterson, 2013).

Differences between typing and handwriting are more clearly discernible in the finished product. The salient difference is in length, with typists generally producing longer responses than hand-writers in several studies (Charman, 2014; Hillier & Lyon, 217; Kim, Bowles, Yang & Chung, 2018; Kohler, 2015; Lee, 2002; Mogey et al., 2010; Whithaus, Scott & Midyette, 2008). However, it is important to note that the length of a student's exam response depends on their content knowledge and analytical skills as well as on their typing or writing speed (Augustine-Adams et al., 2001).

Discrepancies exist in the findings of research into the length and organization of sentences in students' responses. Mogey and Hartley (2013) found that typists produced more, but shorter, sentences and arrange them into a smaller number of paragraphs; Kohler (2015) also observed that typists wrote fewer (i.e. longer) paragraphs than their handwriting peers. In contrast, Lee (2002) and Mogey and Paterson (2013) found that typists produced longer sentences.

Research into the linguistic features of typed and handwritten responses reveals further differences, although these are not necessarily significant (Kim et al., 2018). Charman's (2014) detailed analysis of responses produced by high-school students revealed greater lexical variation (range of vocabulary) in typed responses, but a slightly greater lexical density (proportion of meaning-bearing words to functional words) in handwritten responses. Mogey and Hartley (2013) also observed greater lexical density in students' handwritten responses. Both measures, variation and density, tend to be higher in writing than in speech. These findings led Charman, and Mogey and Hartley, to suggest that students may write in a more informal style on the computer than on paper. Once again, the findings are not consistent across studies: for example, Whithaus et al. (2008) found that handwritten exam scripts were in a more informal style than typed ones.

Influence of the tool

The speed of the motor act of transcription – whether handwriting or typing – can determine how much of a student's working memory is available for the higher-level actions involved in text composition (Peverly, 2006). Indeed, combined with exam pressure, speed can have an impact on a student's exam performance, as Connelly et al. (2005) point out in relation to handwritten exams:

> … it is only when cognitive load is high that handwriting fluency becomes an important predictor of writing quality. This does not bode well for handwritten exams, where it is presumed that the quality of knowledge produced reflects the learning of the student, not simply how fluently they can write. (p. 106)

In view of research suggesting extensive computer use can impair fine motor skills including handwriting (Sülzenbrück, Hegele, Rinkenauer & Heuer, 2011), the argument in favour of a move to e-exams would appear strong. It is further reinforced by the expectation that much coursework is typed; as a result, students may have little or no practice writing essays by hand (Mogey et al., 2008), except in subjects in which students are required to handwrite formulae or hand-draw diagrams (such as mathematics and the sciences). Indeed, a number of authors conclude that typing proficiency has a stronger influence on students' e-exam performance than the amount of their computer experience (e.g. Bridgeman & Cooper, 1998; Kohler, 2015). Furthermore, in a *dual-option* situation, where students are given the choice between handwriting and typing an exam, students who are proficient typists may be more willing to type their scripts (Mogey & Fluck, 2015).

Although it has been observed that students generally type faster than they can handwrite (Augustine-Adams et al., 2001) and that they may complete their exams more quickly (Truell, Alexander & Davis, 2004), neither observation is universally the case. Furthermore, students' typing speeds vary. Slow, two-fingered typists may report that the effort of typing interferes with the process of composition or forces them to write more concise responses (Fluck, Pullen & Harper, 2009; Lee, 2002). In contrast, those who have been trained in typing and can type faster are in a stronger position to perform well (Kohler, 2015). Kohler (2015) raises the additional possibility that using inefficient keyboard techniques may hamper students:

> … lack of fluency in lower order cognitive processes such as keyboarding or handwriting constrain higher order cognitive processes … To this end, it might make sense that less fluent typists would be forced to spend more time on lower order processes as opposed to higher order processes that have to do with the content and organization of their ideas in essays. (pp. 140–141)

It is also not necessarily the case that faster typists necessarily produce lengthier exam responses. Indeed, Mogey and Hartley (2013) found no association between speed and the number of words produced. This may be explained, at least in part, by pauses for thinking and a greater time spent on revision.

Given that handwriting speed and style (printed or cursive) can have a similar impact on the finished product (Graham, Weintraub & Berninger, 1998; Connelly et al., 2005), providing the dual option to students can prove problematic for them. On the basis of their statistical analysis of handwritten and typed exam scripts, Augustine-Adams et al. (2001) advise that students should type their exams if they are proficient typists; otherwise, it is preferable to spend time studying the substance of the course rather than improving their typing skills.

Speed is not solely a function of an individual student's typing proficiency; it can also be adversely affected by the device used. This is particularly the case where students take their exams on unfamiliar institutional devices (Hillier, 2015a; Lee, 2002; Walker & Handley, 2016). Indeed, Walker and Handley draw a distinction between "digital proficiency – reflected in the effective day-to-day use of technology for learning (e.g. from email to essay writing) – and IT proficiency for assessment, reflected in the capability to use unfamiliar technology under time pressure in computer-based exams" (2016, n.p.). A solution is for students to bring their own devices to the exam, but this raises a number of practical issues including computer security (Masterman & Fresen, 2017).

Marking e-exams

Perceptual and attitudinal influences on the marking of exam scripts

The research studies reviewed for this report suggest that a number of perceptual and attitudinal factors come into play as markers read and mentally process students' exam scripts.

Lee (2004) reports the negative impression created by "severely illegible" handwriting, which participants in his study felt "interrupted the smooth flow of reading and impaired their focus on content." He suggests that, in comparison with typed scripts, handwritten scripts may exercise a negative influence where markers find themselves in a "time-constrained testing condition" (pp. 13–14). In contrast, Powers, Fowles, Farnum and Ramsey (1994) 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 (p. 221). Powers et al. 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).

Although typed exam responses tend to be longer than handwritten ones, typed scripts give the visual appearance of being shorter than handwritten essays, even where their word count is the same or greater. In Powers et al.'s (1994) study, this remained the case even when typed scripts 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 (Augustine-Adams et al., 2001; Charman, 2014; Hillier & Lyon, 2018; MacCann, Eastment & Pickering, 2002).

A number of studies have reported on suspected differences in markers' overall expectations of typed and handwritten scripts (Lee, 2004; MacCann et al., 2002; Mogey et al., 2008; Mogey, Cowan, Paterson & Purcell, 2012; Powers et al., 1994; Whithaus et al., 2008). It has been suggested that markers may expect typed scripts to be qualitatively different from handwritten scripts, as Whithaus and colleagues explain:

> It is tempting to think of the writing ability that is measured on a high-stakes exam as first-draft writing and therefore not subject to differences in composing materials. … Having the exams keyboarded seems to have shifted readers' expectations away from first-draft writing toward higher expectations associated with texts that have been more thoroughly revised. (pp. 12, 14)

Students in Mogey and colleagues' (2008) study suggested that typing their exams might lead markers to expect their responses to approach the same standard as their typed coursework. This situation can be of concern to them as the standard of work expected an exam situation differs substantially from the standard expected from coursework (Mogey et al., 2008; Mogey et al., 2012): i.e. there may be a difference in the marks they achieve.

On-screen marking: influence of the tool

If students type their exams and submit scripts in digital format, it arguably follows that academics should mark the scripts on the computer. A number of studies address the experience of on-screen marking (OSM), but only in relation to handwritten scripts that have been scanned into the computer. Even so, some of the findings may also be relevant to the marking of scripts in typed directly onto the computer.

Shaw (2008), and subsequently Johnson and colleagues (Johnson & Nádas, 2009; Johnson, Nádas & Shiell, 2009), investigated whether cognitive processes differ between marking on screen and marking on paper. In addition to slower reading speeds on screen (Shaw, 2008) and a greater cognitive load at first (Johnson et al., 2009), differences were found in reading strategies, navigation and awareness of spatial relationships within individual scripts, and annotation practices.

For example, Shaw's (2008) work suggests that the mode in which an exam script is read (i.e. on paper *versus* on a computer screen) has an impact on the marker's reading strategy. Participants in his study reported that they were more likely to read in a linear fashion on paper and in a haphazard fashion on the screen. They also found it harder to skim-read scripts on the screen in order to elicit the salient themes of individual responses; instead, they had to resort to multiple readings. Shaw comments:

> If examiners construct meaning by processing at different levels concurrently in an interactive way then they must be able to apply simultaneously, elements such as context and purpose together with lexico-grammatical and discoursal features. If mode affects their ability to do this, then not only will a different reading strategy be employed on-screen but recovery of the intended meaning of a candidate's answer might be compromised when marking on-screen. (p. 267)

Differences in performance

Students' performance

As with other aspects of e-exams, research comparing the actual marks achieved in typed and handwritten responses has yielded contradictory results. Among the studies read in this survey, higher marks were awarded to handwritten responses in the work reported by Bridgeman and Cooper (1998), Kohler (2015), Lee (2004), Mogey et al. (2010) and Powers et al. (1994). Higher marks were awarded to typed essays in the work reported by Augustine-Adams et al. (2001), Charman (2014), MacCann et al. (2002) and Whithaus et al. (2008). Lee (2004) found that handwritten responses received higher marks than typed responses when scored holistically, but the situation was reversed when responses were scored analytically (i.e. using a set of specified criteria). MacCann et al. (2002) also report discrepancies when scripts we marked holistically and analytically. However, in almost all of the studies the differences between the marks awarded were not statistically significant.

Reasons put forward for the higher scores awarded to handwritten essays include the greater visibility of errors in typed scripts (Kohler, 2015; Lee, 2004; MacCann et al., 2002; Whithaus et al., 2008); the greater perceived length of handwritten essays; and the possibility that markers have higher expectations of typed responses.

Researchers have also investigated the potential role of specific demographic characteristics in students' performance in typed *versus* handwritten tests. Gender and ethnicity are ruled out as influencing factors in Bridgeman and Cooper's (1998), and Augustine-Adams et al.'s (2001) studies. ESOL (English as a second or other language) status is also dismissed as a factor by Augustine-Adams et al. (2001); however, in Kohler's (2015) very small-scale study of eight ESOL students, six participants performed better in handwritten exams.

It is possible that gender may have an influence on performance in dual-option situations. When students in Mogey and colleagues' studies were given the choice, more male students opted to type than females, although the difference was more pronounced in the study by Mogey and Hartley (2013) than in the one by Mogey et al. (2012). Given that boys may produce more text, and their essays may be of higher quality, when they type their work (Dixon, Cassady, Cross & Williams, 2005), the dual option may help male students to lift their marks.

The lack of consensus among study findings may be attributable in part to the settings in which the studies were conducted or to shortcomings in the methods adopted (Lee, 2004). Most of the studies surveyed refer to the assignments that students were required to complete as "essays" or "tests", since few were actually carried out in actual examination settings. For example, Mogey et al. (2010) observe that "mock" exams or artificial settings can influence participants' attitudes. Bridgeman and Cooper (1998) noticed practice effects where participants were tested in both modes, regardless of which mode they took the test in first. Mogey and Fluck (2015) admit to differences between the cohorts that they studied in Edinburgh and Tasmania.

Markers' performance

As well as investigating students' performance in typed *versus* handwritten exams, it is important to consider possible differences in the performance of the academics who mark the scripts. Indeed, Whithaus et al. (2008) specifically highlight "the need to analyse how the medium of reading an exam impacts the raters' ability to apply assessment criteria" (p. 14). Measures of markers' performance include severity and reliability.

Regarding severity, Whithaus et al. (2008) found that markers were no more severe on typed scripts than on handwritten scripts, despite the reported difference in expectations. Johnson et al. (2009) detected no significant difference between handwritten scripts marked on paper and scanned handwritten scripts marked on the screen: "Where an examiner was severe or lenient in one mode they were also similarly severe or lenient in the other mode" (p. 7). However, a subsequent study by Johnson, Hopkin, Shiell and Bell (2012) found that markers were slightly more lenient on screen than on paper.

There are empirical indications that reliability between markers may be improved by the marking of typed scripts. Bridgeman and Cooper tentatively ascribe the improvement to "the greater standardization in the word-processed essays in which raters cannot attend to differences in handwriting or overall neatness" (1998, p. 4). The difference may also depend in part on the marking scheme adopted: Lee (2004) found greater reliability in the marking of typed scripts when they were marked holistically, but not when they were marked analytically.

On the basis of a comprehensive literature review on reliability in the OSM of scanned handwritten scripts, Tisi, Whitehouse, Maughan and Burdett (2013) suggest that OSM tools may contribute to greater inter-rater

reliability in two ways. Individual exam papers can be split up so that different individuals mark different questions (item-level, as opposed to paper-level, marking); and the collection of analytics makes it possible to detect inconsistent or inaccurate marking throughout the marking period and to act where needed.

The question of equivalence and its implications for implementing e-exams

Among the authors whose papers are surveyed in this report, Whithaus and colleagues (2008) stand out in considering that the processes of handwriting and typing do not differ significantly. For them, it appears more important that students are given the dual option so that they can choose the mode in which they feel the most competent. In contrast, Lee's (2002) findings suggest that:

> … the constructs measured in computer and paper modes are not the same. That is, the incorporation of computers into writing assessments involves a new way of thinking about composing processes, which introduces a source of variability in the original constructs. Inevitable sources of non-equivalence of the construct between them might lead to differences in test performance to some extent. (p. 152)

In relation to research questions 1 and 2, the preceding survey of the research literature leads us to conclude likewise: the move to e-exams does lead to changes in the processes and outputs of sitting and marking typed and handwritten exams, so they are not equivalent. This view rules out the dual option as a solution for students whose typing is less proficient, or who prefer to handwrite their exams for other reasons. An alternative solution is to make e-exams compulsory and offer students opportunities to develop their overall proficiency in general and IT proficiency for assessment in particular, so that they can concentrate on demonstrating their knowledge of the topic in the e-exam rather than expend cognitive resources grappling with the mechanics of production. Indeed, Weigelt-Marom and Weintraub (2018) report that learning to touch-type can narrow the gap between typing and handwriting speeds, and, with time and practice, students may touch-type faster than they handwrite. The positive effect is greater among students with special needs; even so, the option to handwrite must remain open to certain students in this category.

The marking of e-exams is another area for careful consideration. The largely insignificant differences between the marks achieved in typed *versus* handwritten exams suggest that the risk of grade inflation (conversely, deflation) resulting from the change of tool is negligible. So the key question becomes: Do markers treat typed responses akin to coursework essays and, therefore, expect higher standards? If yes, then there may be a requirement to recalibrate marking schemes. Also, if marking typed scripts on the computer improves reliability, a case could be made for compulsory OSM, accompanied by training in IT proficiency for marking. However, such a move would have implications for academics' freedom to make their own choices. Finally, since most of the existing evidence of academics' OSM strategies so far is derived from studies involving scanned handwritten scripts, we advocate a closer investigation into the OSM of typed scripts and its impact on academics' marking strategies.

Regarding research question 3, although the differences in academic outcomes may be statistically insignificant, they matter to students whose marks hover on the boundaries between grades. Some will benefit from a move to e-exams, but others will be disadvantaged. However, we should balance this concern against the inequity that has historically existed in handwritten exams, where some students can write more fluently than others, thereby achieving higher marks. Indeed, we may need to reframe our approach to the question. Rather than puzzle over the conflicting evidence of differences, or design further studies that attempt to replicate the conditions of high-stakes exams in low-stakes exams or "mock" settings in the search for a definitive answer, it may be more productive to adopt e-exams across the board and use the analytical tools in the e-exam software to understand better students' writing behaviors and academics' marking strategies in a digital medium. For example, data on the number of characters that a student types, their typing activity over the course of time and the length of their script may yield insights that can be used to help them adopt appropriate writing strategies for e-exams. The potential for analytics in checking, and improving, reliability between markers (Tisi et al., 2013) was referred to earlier in this paper. Research into the contribution of analytics to our knowledge and understanding could, together with training in IT proficiency for assessment, go some way towards minimizing the inequities in students' performance that inevitably arise when they have no choice over the mode in which they sit their exams. Arguably, with such digital tools at our disposal, e-exams may ultimately prove more equitable – or at least less inequitable – than handwritten exams.

As the preceding paragraphs imply, changing the tool used for a particular activity may (and arguably should) prompt a reassessment of the activity itself. Indeed, moving from handwritten to typed exams raises deeper

questions about the purpose, nature and validity of essay-based exams as a form of summative assessment:

> … when the stress in the course work has been on word-processed output, then handwriting extended prose under exam conditions could be regarded as a poor alignment of assessment practices with intended learning outcomes (Biggs 1999) and further, it may not be an accurate reflection of the quality of work the student is capable of producing. (Mogey et al., 2008, p. 39)

However, a wholesale transformation of summative assessment cannot be achieved overnight. Essay-based exams, conducted increasingly on the computer, are likely to persist as a feature of assessment in higher education for a number of years, and so the research reviewed in this paper should be of ongoing relevance.

Future work

Two possible directions for future research into e-exams were indicated in the previous section; namely:
1. A comparative study of the marking of typed scripts on the screen vs on paper, in order to determine whether the outcomes are equivalent: hence, whether academics should be able to exercise choice regarding the medium in which they mark.
2. An investigation into the potential role of analytics, and specifically keystroke metrics, in uncovering students' behaviours in planning, writing and revising typed exam answers, with a view to supporting them to develop appropriate strategies.

Regarding the role of analytics, studies have already been conducted on coursework essays by, among others, Conijn, van der Loo and van Zaanen (2018) and Türkay, Seaton and Ang (2018). Interestingly, Conijn *et al.* found discrepancies between students' self-reported behaviours and their actual behaviours as tracked by the software. Given that some of the research reported in this paper has relied on self-reports, Conijn *et al.*'s finding opens up an additional avenue for future investigation.

Conclusion

The literature survey reported in this paper has focused on the academic implications of moving from handwritten exams to e-exams. This by no means downplays the practical and policy implications of the move, addressed in the other part of our "landscape" work (Masterman & Fresen, 2017). Institutions will wish to take all factors into account when planning the implementation of e-exams.

The paper set out to explore whether handwritten and typed exams can be treated as equivalent, by addressing two research questions relating to the students who sit the exams and the academics who mark them, and a third asking whether differences in students' performance matter. Our conclusions suggest that there are indeed changes in moving from handwritten to typed mode: that is, we can answer research questions 1 and 2 in the affirmative. The answer to question 3, whether differences in performance matter, is more nuanced. True, existing gauges in the form of marking schemes may need to be checked, and recalibrated, for exams conducted on the computer. However, by additionally capitalizing on the analytics available in e-exam tools, we should have new and powerful gauges by which to understand the strategies adopted by students and markers and, thus, to navigate the high seas of assessment in a digital age with more equitable outcomes for students than hitherto.

Acknowledgements

References

Augustine-Adams, K., Hendrix, S.B., & Rasband, J.R. (2001). Pen or printer: Can students afford to handwrite their exams? *Journal of Legal Education, 51*, 118–129.

Bridgeman, B., & Cooper, P. (1998). Comparability of scores on word-processed and handwritten essays on the Graduate Management Admissions Test. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA, 13–17 Apr. https://eric.ed.gov/?id=ED421528

Brunel University (n.d.). *About Digital Assessment @Brunel* https://www.brunel.ac.uk/about/education-innovation/Digital-Assessment-Brunel/About-Digital-Assessment-Brunel [viewed 24 Jul 2018].

Charman, M. (2014). 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), 834–843.

Conijn, R., van der Loo, J., & van Zaanen, M. (2018). What's (not) in a keystroke? Automatic discovery of students' writing processes using keystroke logging. Paper presented at the 8th International Conference on Learning Analytics & Knowledge, Sydney, 5–9 Mar. https://rianneconijn.com/wp-content/uploads/2018/05/LAK_metrics18_fullpaper_2.pdf [viewed 2 Oct 2018].

Connelly, V., Dockrell, J.E., & Barnett, J. (2005). The slow handwriting of undergraduate students constrains overall performance in exam essays. *Educational Psychology, 25*(1), 99–107.

Dixon, F., Cassady, J., & Cross, T. (2005). Effects of technology on critical thinking and essay writing among gifted adolescents. *The Journal of Secondary Gifted Education, 16*(4), 180–189.

Fluck, A., & Hillier, M. (2017). eExams: Strength in diversity. Paper presented at the 11th IFIP TC3 World Conference of Computers in Education, Dublin, Ireland, 3–6 Jul. http://transformingexams.com/files/fluck_hillier_2017_eExams_strength_diversity_IFIP_WCCE_paper_45.pdf [pre-review copy; viewed 24 Jul 2018].

Fluck, A., Pullen, D., & Harper, C. (2009). Case study of a computer based examination system. *Australasian Journal of Educational Technology, 25*(4), 509–523.

Graham, S., Weintraub, N., & Berninger, V.W. (1998). The relationship between handwriting style and speed and legibility, *The Journal of Educational Research, 91*(5), 290–297.

Hillier, M. (2014). The very idea of e-Exams: student (pre)conceptions. In *Rhetoric and reality: Critical perspectives on educational technology. Proceedings ascilite, Dunedin 2014*. http://ascilite.org/conferences/dunedin2014/files/fullpapers/91-Hillier.pdf

Hillier, M. (2015a). E-exams with student owned devices: Student voices. In *Proceedings of the International Mobile Learning Festival: Mobile learning, MOOCs and 21st century learning, 22–23 May 2014, Hong Kong SAR, China* (pp. 582–608). https://www.researchgate.net/publication/286869830_e-Exams_with_student_owned_devices_Student_voices.

Hillier, M. (2015b). To type or handwrite: student's experience across six e-Exam trials. In *Proceedings ascilite*, 30 Nov–3 Dec, Perth, Australia (pp. 143–154). http://www.2015conference.ascilite.org/wp-content/uploads/2015/11/ascilite-2015-proceedings.pdf

Hillier, M., & Fluck, A. (2013). Arguing again for e-exams in high stakes examinations. In *Electric dreams. Proceedings ascilite, Sydney 2013*. http://www.ascilite.org/conferences/sydney13/program/papers/Hillier.pdf

Hillier, M., & Lyon, N. (2018). Writing e-Exams in pre-university college. Paper presented at the Open Conference on Computers in Education (IFIP TC3), 25-28 June, Linz, Austria. http://transformingexams.com/files/Hillier_Lyon_OCCE2018_student_experiences_eexams_preuni_FP.pdf [pre-review copy; viewed 24 Jul 2018].

Johnson, M., & Nádas, R. (2009). Marginalised behaviour: digital annotations, spatial encoding and the implications for reading comprehension. *Learning, Media and Technology, 34*(4), 323–336.

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.

Johnson, M., Nádas, R., & Shiell, H. (2009). An investigation into marker reliability and some qualitative aspects of on-screen essay marking. Paper presented at the British Educational Research Association Annual Conference, University of Manchester, 2–5 Sep. http://www.leeds.ac.uk/educol/documents/187452.pdf [viewed 24 Jul 2018].

Kim, H.R., Bowles, M., Yan, X., & Chung, S. J. (2018). Examining the comparability between paper- and computer-based versions of an integrated writing placement test. Article in press. *Assessing Writing*. https://doi.org/10.1016/j.asw.2018.03.006

Kohler, B. (2015). Paper-based or computer-based essay writing: differences in performance and perception. *Linguistic Portfolios, 4*(1), Article 13, 129–146. http://repository.stcloudstate.edu/stcloud_ling/vol4/iss1/13 [viewed 24 Jul 2018].

Lee, H.K. (2004). A comparative study of ESL writers' performance in a paper-based and a computer-delivered writing test. *Assessing Writing, 9*, 4–26.

Lee, Y.-J. (2002). A comparison of composing processes and written products in timed-essay tests across paper-and-pencil and computer modes. *Assessing Writing, 8*, 135–157.

MacCann, R., Eastment, B., & Pickering, S. (2002). Responding to free response examination questions: computer versus pen and paper. *British Journal of Educational Technology, 33*(2), 173–188.

Masterman, L. & Fresen, J. (2017). *Landscape Report: Research and Practice in E-exams*. Unpublished report, University of Oxford. Available on request.

Mogey, N., & Fluck, A. (2015). Factors influencing student preference when comparing handwriting and typing for essay style examinations. *British Journal of Educational Technology, 46*(4), 793–802.

Mogey, N., & Hartley, J. (2013). To write or to type? The effects of handwriting and word-processing on the written style of examination essays. *Innovations in Education and Teaching International, 50*(1), 85–93.

Mogey, N., & Paterson, J. (2013). Stylistic differences between typed and handwritten essay examinations. *International Journal of e-Assessment, 3*(1), 1–14.

Mogey, N., Cowan, J., Paterson, J., & Purcell, M. (2012). Students' choices between typing and handwriting examinations. *Active Learning in Higher Education, 13*(2), 117–128.

Mogey, N., Paterson, J., Burk, J., & Purcell, M. (2010). Typing compared with handwriting for essay examinations at university: letting the students choose. *ALT-J, 18*(1) 29–47. http://dx.doi.org/10.1080/09687761003657580

Mogey, N., Sarab, G., Haywood, J., van Heyningen, S., Dewhurst, D., Hounsell, D., & Neilson, R. (2008). The end of handwriting? Using computers in traditional essay examinations. *Journal of Computer Assisted Learning, 24*, 39–46.

Moore, C.P. (2018). Adding authenticity to controlled conditions assessment: introduction of an online, open book, essay based exam. *International Journal of Technology in Higher Education, 15*(26). https://doi.org/10.1186/s41239-018-0108-z

Newland, B., & Martin, L. (2016). *Electronic Management of Assessment in UK HE 2016.* Heads of e-Learning Forum. https://helfuk.blogspot.com/p/projects.html [viewed 25 Jul 2018].

Noyes, J.M., & Garland, K.J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics, 51*(9), 1352–1375.

Peverly, S.T. (2006). The importance of handwriting speed in adult writing. *Developmental Neuropsychology, 29*(1), 197–216.

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.

Shaw, S. (2008). Essay marking on-screen: implications for assessment validity. *E-Learning, 5*(3), 256–274.

Sindre, G., & Vegendla, A. (2015). E-exams and exam process improvement. Paper presented at the Norwegian Informatikkonferanse – NIK), Ålesund, Norway 23–25 Nov. https://www.researchgate.net/publication/299497855_E-exams_and_exam_process_improvement

Sülzenbrück, S., Hegele, M., Rinkenauer, G., & Heuer, H. (2011). The death of handwriting: secondary effects of frequent computer use on basic motor skills. *Journal of Motor Behavior, 43*(3), 247–251.

Tisi, J., Whitehouse, G., Maughan, S., & Burdett, N. (2013). *A review of literature on marking reliability research. Report for the Office of Qualifications and Examinations Regulation (Ofqual).* Slough, UK: NFER. https://www.nfer.ac.uk/publications/MARK01/MARK01.pdf [viewed 24 Jul 2018].

Truell, A.D., Alexander, M.W., & Davis, R.E. (2004). Comparing postsecondary marketing student performance on computer-based and handwritten essay tests. *Journal of Career and Technical Education, 20*(2). No page numbers. https://ejournals.lib.vt.edu/JCTE/article/view/637/922 [viewed 24 Jul 2018].

Türkay, S., Seaton, D.T., & Ang, A.M. (2018). Itero: a revision history analytics tool for exploring writing behavior and reflection. Paper presented at the CHI Conference on Human Factors in Computing Systems, CHI 2018, Montreal, Canada, 21–26 Apr. https://dblp.org/rec/conf/chi/TurkaySA18 [viewed 2 Oct 2018].

Walker, R., & Handley, Z. (2016). Designing for learner engagement with computer-based testing, *Research in Learning Technology, 24*(1), 30083. http://dx.doi.org/10.3402/rlt.v24.30083

Weigelt-Marom, H., & Weintraub, N. (2018). Keyboarding versus handwriting speed of higher education students with and without learning disabilities: Does touch-typing assist in narrowing the gap? *Computers and Education, 117*, 132–140.

Whithaus, C., Harrison, S.B., & Midyette, J. (2008). Keyboarding compared with handwriting on a high-stakes writing assessment: Student choice of composing medium, raters' perceptions, and text quality. *Assessing Writing, 13*, 4–25.