ALT-J, Research in Learning Technology
Vol. 18, No. 1, March 2010, 29–47

Typing compared with handwriting for essay examinations at university: letting the students choose

Nora Mogey[a]*, Jessie Paterson[b], John Burk[b] and Michael Purcell[b]

[a]*Information Services, University of Edinburgh, UK;* [b]*School of Divinity, University of Edinburgh, UK*

(*Received 8 April 2009; final version received 25 January 2010*)

Students at the University of Edinburgh do almost all their work on computers, but at the end of the semester they are examined by handwritten essays. Intuitively it would be appealing to allow students the choice of handwriting or typing, but this raises a concern that perhaps this might not be 'fair' – that the choice a student makes, to write or to type, will affect their mark. The aim of this study was to identify and explore any systematic differences that may be introduced due to offering students the choice to write or type essay examinations. A class of 70 first-year divinity students were given the option of taking a mock examination, and the further option of handwriting or typing their answer. All the examination scripts were then faithfully transcribed into the opposite format so there was a printed copy and a handwritten copy of every script. These were then marked by four markers, such that every marker marked every script exactly once, in one format or the other. No significant differences could be identified due to the format in which the students had written their answer. Factors explored included length of essay, overall score awarded, and some qualitative measures designed to indicate essay quality. In contrast, the variation between the markers was striking.

Keywords: essay; examination; laptop; type; choice

Introduction

> I depend on a keyboard to write, and frankly that collection of ill-arranged keys has become an extension of my fingers into which I pour my thoughts. In addition, I depend heavily on spelling and somewhat on grammar checkers to fix my mistakes automatically, so I rarely slow down to correct the small errors. Moreover, I depend on the cut-and-paste facility to make up for my predilection to afterthoughts. Like most folks, I rarely write a paper from beginning to end; rather, I usually start with the 'results' and work backwards and forwards as the Muse inspires me. (Penny 2003)

For some years, staff at the University of Edinburgh have expressed concern that students do almost all their work on computers, but at the end of the semester they are examined by handwritten essays. Discussion with the students' association has been met with a supportive but slightly anxious reaction – it is an interesting idea to explore but is it really fundamentally fair? Equally college and school examination boards have been reluctant to take the decision to move to typed examinations until they are confident that this will not result in a rush of student appeals.

*Corresponding author. Email: Nora.Mogey@ed.ac.uk

ISSN 0968-7769 print/ISSN 1741-1629 online

DOI: 10.1080/09687761003657580
http://www.informaworld.com

This research initially aimed to answer some of the questions frequently cited as barriers to offering students the opportunity to type their responses to essay examinations. Questions such as: 'Is the mark awarded to an examination script influenced by the format of the script (typed or handwritten) rather than its content?' and 'Are students who type slowly any more or any less disadvantaged than students who handwrite slowly?' Data about students reactions and preconceptions to the idea of essay examinations on computer and some initial results from this study have been presented previously (Mogey et al. 2008a, 2008b, 2008c).

This paper completes and expands those results, particularly with respect to variation between markers and variation between questions.

Setting the scene

The idea of using computers to allow students to type responses to essay-style questions is not new (Howell 2003). US law schools routinely run high-stakes essay examinations on student-owned laptops (e.g. New York University, Supreme Court of Missouri, Kentucky Office of Bar Admissions),[1] and the software has proved to be stable and reliable (personal email). Despite this, very few relevant studies have been identified that provide empirical evidence relating to university students under examination conditions. In one of the few higher education examination studies, Augustine-Adams, Hendrix, and Rasband (2001) concluded that, on average, a law student typing an examination could expect to perform slightly better than their colleague who handwrites.

The process of composing an essay using a keyboard is different to using pen and paper. When writing by hand, planning what is to be written is a critical and important element, but use of a word processor makes editing text easy and therefore means the author can afford to spend less time in planning their work. In other sectors there is substantial evidence that students who have written their (non-examination) essays using a computer write to a better standard (MacCann, Eastment, and Pickering 2002; Russell and Plati 2001; Goldberg, Russell, and Cook 2003; Hartley and Tynjala 2001). Several studies have demonstrated both that text composed using a word processor is subject to more revisions than text composed on paper and that students typically type more than they handwrite (Russell and Haney 1997; Russell and Plati 2001; Wolfe et al. 1996; Lee 2002). However, Burke and Cizek (2006) found the opposite and demonstrated that, irrespective of information technology (IT) skills or confidence, sixth graders produced better essays by hand than they did using a word processor.

Overall familiarity with technology also seems to play a role in student performance. Horkay et al. (2006), studying school pupils, found that hands-on IT experience was significantly related to online writing assessment performance – computer familiarity added about 10% to the score achieved. Russell and Haney (1997) demonstrated that where (school) students were accustomed to writing on a computer their responses in tests were much better when they were allowed to type their answers – only 30% of students passed when handwriting, as opposed to 67% when they used a keyboard.

Care should be taken before extrapolating too far from a non-examination context into the stressful high-stakes summative examination setting. Thomas, Paine, and Price (2003), studying Open University computer science students, demonstrated that typically each student submitted 30% less material in the mock examination than they eventually submitted in the final examination. However it is worth noting that although a student typing can usually write more words, examination essays can sometimes be too long – slightly shorter essays score more highly than very long essays (Whithaus, Harrison, and Midyette 2008).

A further source of variability is the recorded behaviour of markers. Several studies have demonstrated that a type-written essay will be marked more harshly than an identical handwritten text, although the difference in scores is not always large (Russell and Tao 2004a; MacCann, Eastment, and Pickering 2002; Bridgeman and Cooper 1998). The reason for the difference is not known for certain but seems likely to be associated with an expectation that handwritten work is essentially a first draft standard whereas typed text would normally have been more thoroughly revised.

Background and preparation

A very simple tool that was less likely to confound the mark for the academic content of the examination with a measure of the student's skill in using a particular word processor was the preferred option. From the systems identified it was decided to opt for Exam4 (marketed by Extegrity Inc.).

Exam4 was attractive for many reasons in addition to its solid track record. The software includes a range of security measures: on launch it checks the local computer configuration for possible cheat mechanisms, such as running virtual computers; blocks access to all other materials on the hard drive and network; and makes regular backups of work in progress so that, in the unlikely event of a problem, all is not lost and all stored files are encrypted, thus controlling access to completed examinations.

When launching the software the user follows a channelled, stepwise examination start-up procedure, selecting from a series of simple menus or entering basic personal identification details that together configure all the pertinent administrative settings (e.g. saving) without the need for issuing complicated instructions. Students can choose a large or a small screen font, and whether to have a clock and/or a warning notice when time is running out. Once the examination start-up sequence has been completed, the student clicks a button to begin the examination itself. The software 'locks the computer down' so the student is unable to access the Internet, the hard disk or read information from an accessory device such as a USB stick or CD-ROM.

An examination can be administered in different ways using Exam4. It was our intention to minimise the changes from existing practice, so a physical (paper) question paper was still created, secured in staff offices until needed and distributed by hand in the examination venue. Students only use the computer to type their answers, and at the end of the examination these are retained in encrypted format on their hard drives as well as transmitted to a specific nominated computer that can be located essentially anywhere. A separate administrative tool is then used to print all of the examination files in a single batch. Printed scripts are distributed to staff for marking in the traditional manner. Thus the only part of the examination process that changes significantly is that students no longer handwrite their answers.

Three different pilot studies (Mogey and Sarab 2006) had established that although no students experienced difficulty in using the software, there was a general uncertainty (in the minds of both staff and students) about whether this was really fair and equivalent to a handwritten examination. There has been a great deal of caution on the part of examination boards and boards of studies when they have been asked to support the use of laptops for essay examinations.

The main concern expressed by students has been about typing ability and whether the software would crash, while the biggest perceived advantage is the ability to edit text: "it is easy to skip back and forward, rereading and changing areas as new ideas spring to mind. This is a vast improvement. In addition towards the end, handwriting does not deteriorate". The notion that the whole cognitive process of writing on a computer differs significantly from the process of handwriting an essay has not been raised by the students themselves.

The pilot studies also provided responses to the direct question 'Are essay exams a good idea?' About one-third of students responded with broadly negative comments, and about two-thirds with broadly positive comments.

Positive comments included the following:

> Yes, as the world is becoming more and more computerised, we must embrace this in all parts of academic life.
>
> Yes, because the nature of exams are changing and revision styles are changing because of computers.
>
> Yes. People are using computers more in the workplace, so it would be beneficial.

Negative comments included:

> No, because it would put people on different starting points (e.g., touch typing) Also exam conditions are different, we have always done exams on paper.
>
> No. Computers can crash & break down. This would not be good if we had a time limit. They are not efficient and safe compared to pen and paper.
>
> No. I would write less; it would interrupt my thought process.

This study was designed to provide examination boards, boards of studies, teaching teams and students with some evidence on which to base their judgements about whether using laptop computers for essay examinations should be considered. Of particular interest was whether the use of computers could be offered as a choice – was there any fundamental and systematic difference in the score achieved using a computer rather than pen and paper to compose their essay? At the time of setting up the study, the behaviour of markers and marking variation was (unfortunately) not specifically considered, but the data gathered did allow some *post-hoc* exploration of these effects.

Methodology

Christian Theology 1 is a class of about 70 first-year students. The students were invited to sit a 'mock' examination during timetabled class time, during Week 11 of a 12-week semester. Previously software had been demonstrated and technical assistance was available (although not needed), laptops were available for loan if required. Students were allowed to sit the examination in the format of their choice: typing using a laptop or handwriting onto paper, or they could decide not to sit the mock examination at all.

The mock examination lasted one hour, and was held in the regular class venue but under examination conditions. Students were allowed sight of the examination questions one week in advance and had a choice of one question out of three (Q1–Q3). Students using laptops were mostly situated towards the front of the room, and all had access to power sockets. Those students handwriting were seated at the back of the room. All students were provided with scrap paper, which was left in the venue.

At the end of the examination, typed submissions were collected on a USB stick prior to decryption and printing. All originals were marked swiftly in order to provide formative feedback to the students well in advance of the real examination. Meanwhile a professional typist was employed to produce faithful typed scripts from the handwritten originals, replicating any spelling and grammatical errors, and similarly the typed originals were distributed amongst 'volunteers' who each created a handwritten version. Thus a typed and a handwritten version of each script was generated, each of which was duplicated and then blind marked.

Four marks were generated from each student script, one from each of four markers. Two of the marks were for typed versions and two for handwritten versions. Each marker graded each student essay exactly once, creating a balanced design. All of the markers were experienced at marking first-year divinity essays. The total number of words written during the mock examination by each student and the number of words in any conclusion paragraph(s) were also recorded.

In addition to producing scores for the scripts, markers were asked to rate the scripts on six qualitative dimensions. This information is normally used in the feedback provided to students, but were used in this study to give an indication of the quality of the script. These dimensions are engagement with the topic; knowledge of the subject matter; demonstration of critical thinking skills and abilities; evidence of wider reading, beyond the core recommended texts and articles; structure and presentation of the essay; references and bibliography. In each case items were recorded, on an ordinal scale, as one of: unsatisfactory; OK; good; very good or excellent.

Results

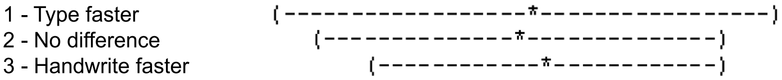
Thirty-seven students chose to sit the mock examination (28 female and nine male). Twenty-four typed scripts and 11 handwritten scripts were collected at the end of the mock (with proportionately more females opting to handwrite than to type) and with two additional handwritten scripts from students who were unable to attend at the scheduled time. The group had a slight bias towards females and towards mature students but represented a reasonable spread of academic ability, based on tutorial marks to that point.

Twelve students chose to do Q1, 15 chose Q2 and 10 chose Q3. There was no difference in question choice made by male or by female students, and the spread of marks achieved suggest all three questions had identical difficulty. Students who elected to answer Q3 and to type scored more highly than the students who elected to answer Q3 but to handwrite. Curiously, 11 of the 12 students who chose Q1 also chose to type their responses, the split was more even for the other questions.

Students who typed in the mock examination generally wrote more words than students who opted to handwrite. Using a two-sample *t*-test of the null hypothesis (H_0: There is no difference in the mean number of words which will be handwritten or typed) results in $t = -2.15$, $p = 0.041$ (25 degrees of freedom), suggesting that this is statistically significant. However the number of words written was not associated with students' reported typing speed (see Table 1 and Figure 1).

Table 1. Number of words written by students handwriting or typing.

	n	Mean	SD
Type faster	6	785.5	238.6
No difference	9	780.3	295.2
Handwrite faster	12	802.0	127.0



```
1 - Type faster         (-----------------*-----------------)
2 - No difference          (-------------*-------------)
3 - Handwrite faster           (------------*------------)
```

Figure 1. Individual 95 confidence intervals for means based on pooled standard deviations.

This may indicate that the amount written in an examination is only partially dependent on the speed of writing/typing – one could speculate perhaps that it depends also on fluency of thought.

Thomas has suggested that the conclusion is critical for giving impression of a well constructed essay (Thomas, Paine, and Price 2003). This study found no evidence of correlation between conclusion length and overall score for the essay (*r* = 0.009). There was also no correlation between overall length of essay and number of words in conclusion – actually there was a small negative correlation (*r* = –0.017). Typed conclusions were in general slightly longer than handwritten conclusions (mean number of words in conclusion: 43 handwritten, 57 for typed), although these data should be interpreted cautiously given the restricted time available for the mock examination and students lack of examination practice with keyboarding essays.

Generally, where originals were typed then scripts scored more highly than where originals were handwritten scripts. For scripts marked in their original formats: mean score awarded handwritten scripts = 52.79, standard deviation (SD) = 7.13 (*n* = 26); and mean score awarded typed scripts = 54.90, SD = 9.0 (*n* = 48).

However, when looking at the marks awarded to the all scripts (ignoring their original format), then the handwritten scripts generally score slightly higher. For all scripts (including transcriptions, *n* = 74): mean score awarded handwritten scripts = 55.12, SD = 8.25; and mean score awarded typed scripts = 53.19, SD = 8.53.

This gives weak evidence in support of a format effect. (These are 72 handwritten scripts and 72 otherwise identical scripts but in a typed format.)

Without a format effect the mean scores should be identical, the predicted value would be 54.16. So, in line with the research cited earlier, there is some evidence that the handwritten scripts are being marked up slightly and the typed scripts marked down slightly.

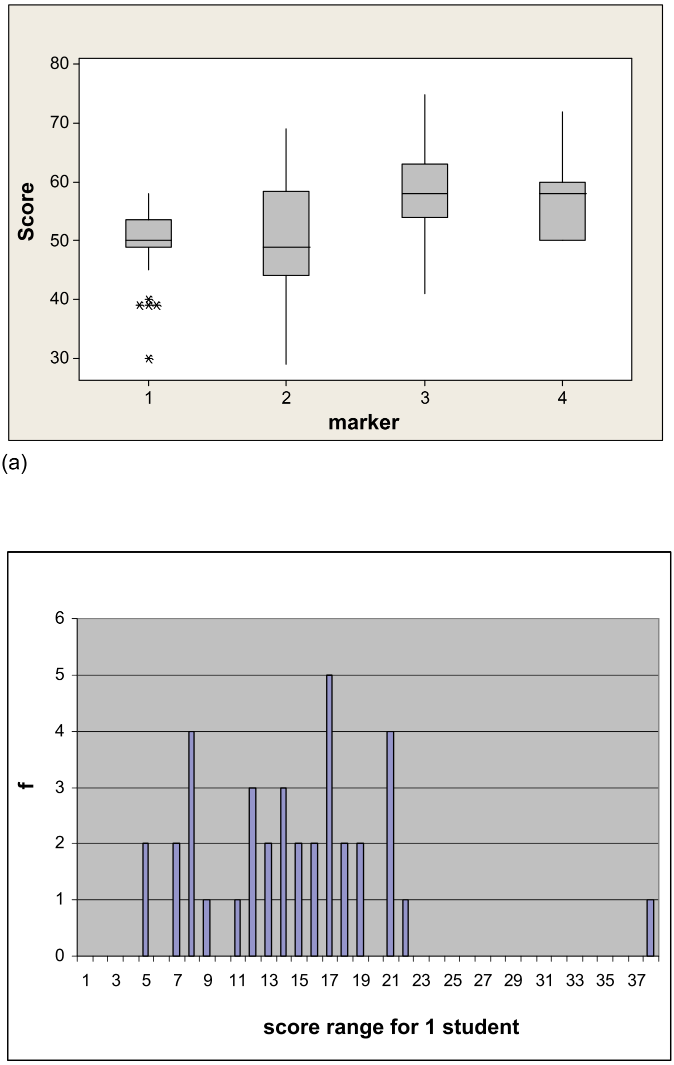
Using a general linear model (see Table 2) to analyse the contribution to variability in scores makes it clear that the variability due to differences between the markers is a far more important effect, and the contribution due to differences in the format of the script is not statistically significant. However it is recognised that the fit of the model is weak and leaves much variability unexplained.

In analysing the data further, to explore differences between markers and between questions, it was decided to use boxplots to form a general impression of the data. The limited numbers of experimental observations, especially once grouped by any or all

Table 2. Analysis of covariance.

Source	DF	Seq. SS	Adj. SS	Adj. MS	*F*	*P*
Format	1	138.17	107.99	107.99	2.00	0.160
Marker	3	2550.87	2550.87	850.29	15.72	0.000
Error	143	7732.38	7732.38	54.07		
Total	147	10421.43				

Note: $S = 7.35341$, $R^2 = 25.80\%$, $R^2(\text{adj.}) = 23.73\%$.



(a)

(b)

Figure 2. (a) Boxplot of score by marker alone. (b) Variation in scores between markers.

of question, marker, scripts and format, were felt to make a full quantitative analysis questionable. The results presented are only intended to suggest likely trends, based on the evidence available.

The distribution of scores for each of the four markers is shown in Figure 2a. It can be seen that, in general, scores awarded by Marker 1 are more clustered than those awarded by other markers; and that in addition to this central tendency, Marker 1 appears to have been unusually harsh on some candidates (indicated by the lowly scored outliers) and Marker 2, and to a lesser extent Marker 3, make much better use of the full range of possible scores. Comparing the four marks awarded to each individual student by the four graders gives marks for any one student with a minimum range of five and an outlying maximum range of 38 (see Figure 2b).

Further when the four marks for each student are placed in rank order, we see that Marker 1 tends to grade lower than their colleagues, and Markers 3 and 4 tend to grade highly (see Table 3).

Drilling down to examine marking at the individual question level, some differences do start to emerge. Figure 3 suggests Marker 1 has marked Q3 differently to Q1 and Q2 – there is a much greater spread of scores with a roughly symmetrical distribution and no outliers. Similarly Marker 4, whose marks tend to show a positive skew, demonstrates a different scoring pattern for Q2 where the marks show a symmetrical distribution.

Table 3. Marker trends when four marks for each student are placed in rank order.

Marker	First	Second	Third	Fourth
1	1	3	16	17
2	4	10	8	15
3	20	10	5	2
4	17	14	6	0

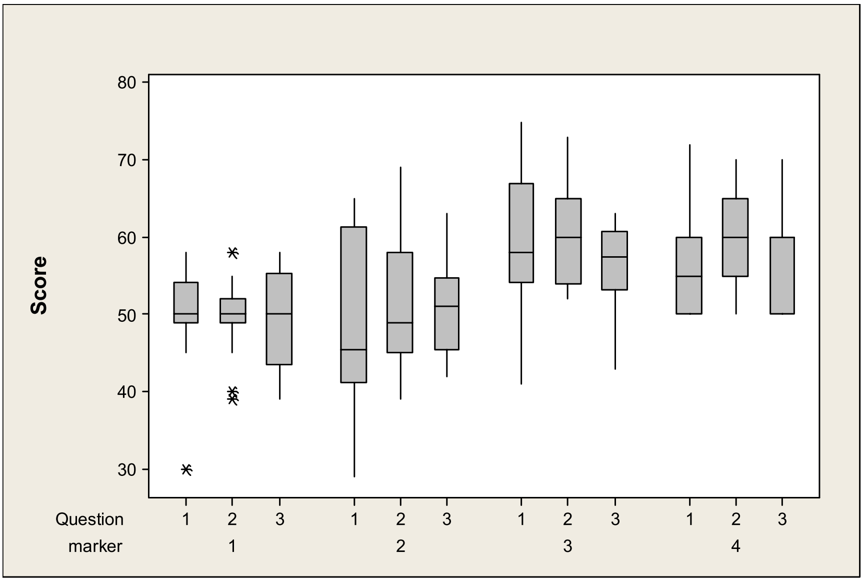


Figure 3. Boxplot of score by marker then question.

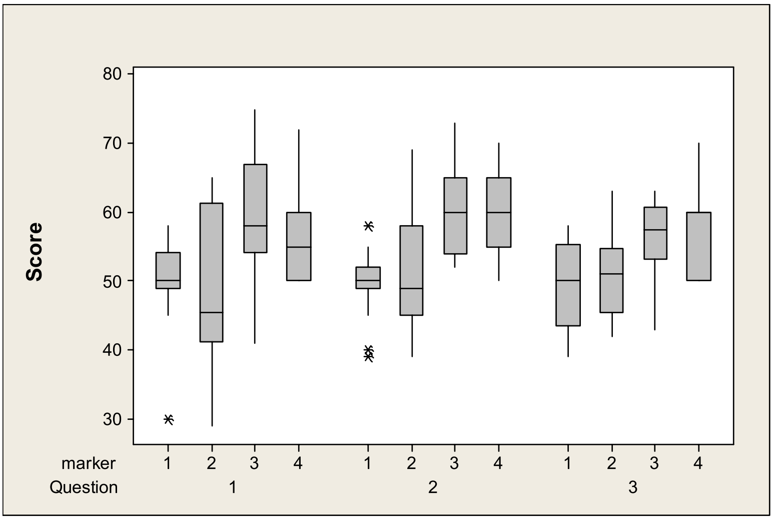


Figure 4. Boxplot of score by question then marker.

Reusing the same data but presenting it question by question (Figure 4) suggests that no one question was marked in the same way by all markers.

So is there a difference in how the markers approach typed or handwritten scripts? Figure 5 suggests there may be. For example, Markers 1 and 2 appear to have graded typed scripts systematically lower than handwritten scripts. This may of course be due to chance – these markers may have been given scripts in a typed format that were genuinely poorer essays.

The mean of the four scores awarded by the different markers was taken to achieve a comparable score for each script. For each marker it is then possible to calculate the difference between the mean for any student and the score that marker awarded. This

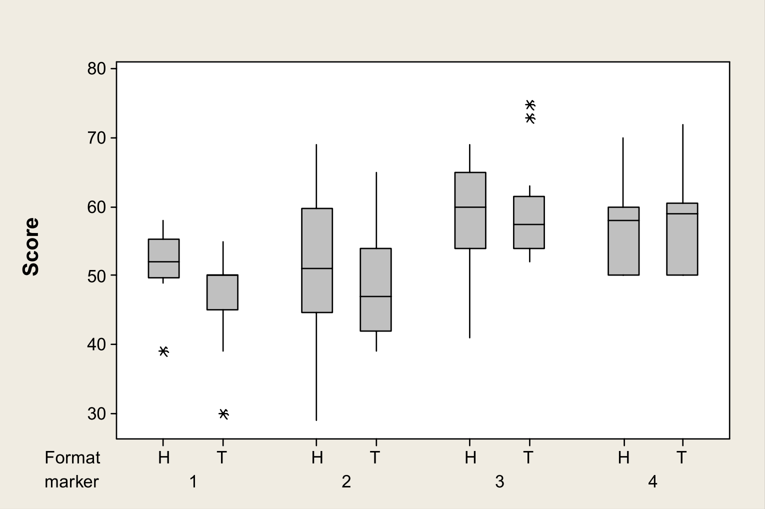


Figure 5. Boxplot of score by marker then format.

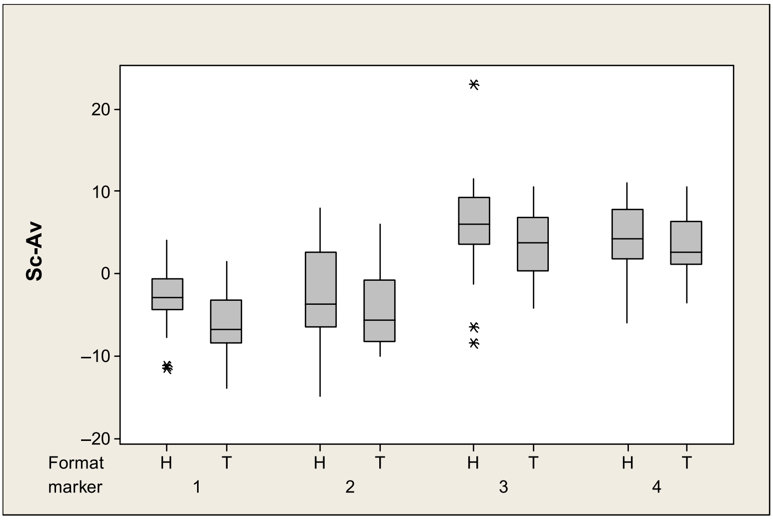


Figure 6. Boxplot of variations in score from the mean by marker then format.

is illustrated in Figure 6 and shows that Marker 1 tends to score slightly low and Markers 3 and 4 tend to score high in comparison with their peers. Figure 6 also shows this difference for handwritten and typed scripts, and demonstrates that for all four markers there was a small but consistent tendency for typed scripts to be awarded a lower mark than handwritten scripts.

The raw score for an essay ideally gives an indication of its quality, but it is recognised that there are different characteristics and components that contribute to the overall quality and overall score awarded. In this study, and in line with standard practice in the School of Divinity, all the markers were required to grade each essay on six characteristics, already identified in the Methodology section. In each case, items were recorded, on an ordinal scale, as one of: unsatisfactory; OK; good; very good or excellent. However, so many of the papers failed to score references and bibliography that this characteristic has been omitted from the analysis.

Each marker marked every essay so ideally the comments and grades should have been well aligned. However, large and systematic differences in the opinions of the markers were evident as with the overall scores. Marker 1 did not regard any items as excellent and judged a total of 46 as unsatisfactory, while Marker 4 judged none as unsatisfactory and 20 as excellent (remember all markers were marking the same essays). See Table 4. This reinforces marking trends for the overall score demonstrated in Figure 2a.

Table 4. Marker trends scoring essay quality characteristics.

Marker	Unsatisfactory	OK	Good	Very good	Excellent
1	46	80	47	2	0
2	18	38	65	41	18
3	15	66	45	40	8
4	0	4	78	61	20

The data were explored further (mapping unsatisfactory as zero, OK as one, etc., up to excellent as four, simply for convenience of analysis).

In the five diagrams in Figure 7 (one for each characteristic), each vertical line represents the range awarded by all four markers for the script of one individual student. The set of vertical lines in one graph represents the variation within the class. Poorly rated students have lines towards the base of each chart, highly rated students have lines towards the top of each chart. Short lines indicate good agreement between markers, long lines indicate variation.

Hence it is clear that there is considerable variation – it is not simply a case of markers differing by a single category, there are many examples of total disagreement with one marker rating an item as excellent while another rates it as unsatisfactory.

Critical skills and wider reading arguably show slightly more consistency than the other items, but they are also the two items that received the overall worst scores. It is possible that the consistency effect is increased due to some markers not wishing to mark first-year students too harshly and risk discouraging them. It is possible that some markers avoided using the unsatisfactory category, thus artificially compressing the data.

Overall the engagement characteristic was marginally the most highly graded factor. There was no evidence of a difference between the questions and, as with overall scores, there was evidence of differences between markers in how they rated these quality characteristics.

However, the main interest for this study is whether there is evidence that handwritten scripts are rated differently on these quality characteristics to typed scripts?

In Figure 8 the data have been grouped according to the original format of that script. No huge differences are indicated, although there is some suggestion that essays which were typed generally showed better subject engagement and handwritten essays were less well structured/presented.

The grades assigned to the factors knowledge, reading and critical skills are all more closely associated with the final score for the essay than the factors engagement and structure. Correlations to final scores and correlations between factors are as follows: engagement, 0.47; knowledge, 0.74; critical skills, 0.72; reading, 0.74; and structure, 0.59 (see Table 5).

Discussion

Limitations of this study

This study has only considered a single group of students from one discipline area, where essay writing may be approached differently from other subjects. However, it is the format of the examination not the discipline that is the key interest, so this is not considered to be too problematic. It is also noted that the students volunteered to take the mock examination, thus forming a self-selected sample. It is unfortunate that because of the voluntary nature of the study some potentially useful data were not available, and in some cases the numbers in the study are on the low side to permit rigorous statistical analysis.

Further, although this study attempted to simulate some of the stresses experienced in an examination hall, they can never really be replicated in a mock examination. The mock examination was held in normal class time and was therefore necessarily shorter than a full examination would be. Students also had only limited revision time and did not really have the chance to be as prepared as they would normally be for an

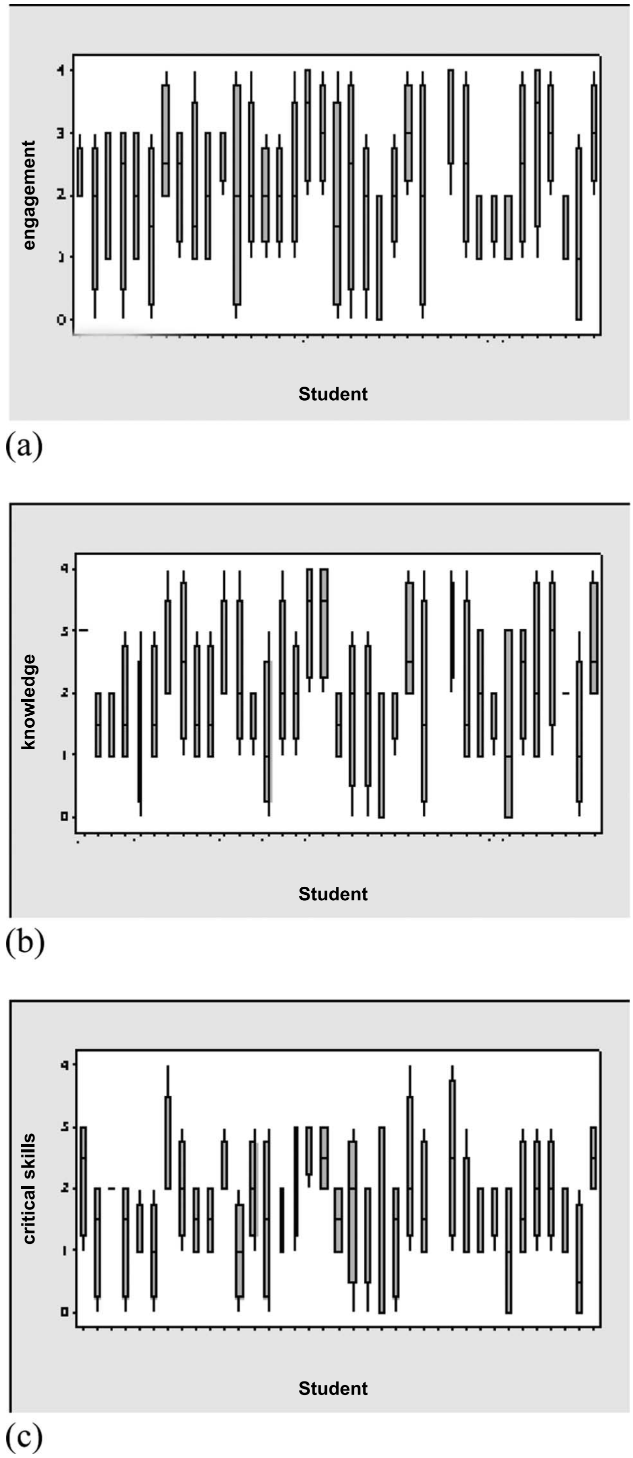


Figure 7. Boxplots of (a) engagement, (b) knowledge, (c) critical skills.

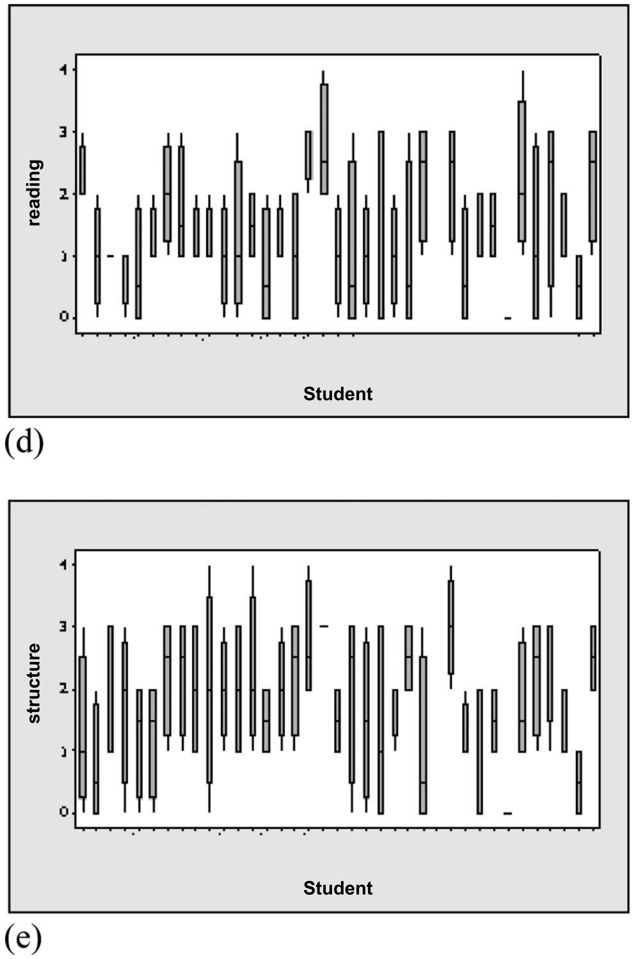


Figure 7. continued. Boxplots of (d) wider reading, and (e) structure and presentation.

examination; this was felt to be of particular concern because these were first-year students, and was compensated for by allowing prior sight of the questions. One student reported: “I did it (used the laptop) for the mock because it didn’t matter”. Clearly revision, confidence and pressure must all have some impact on what and how students write. Similarly, because this was a mock and did not ‘count’, students may have taken it less seriously. One marker said “I suspect students didn’t do a lot of revision for these exams … there wasn’t a lot of evidence of secondary reading. The problems of huge irrelevancies cropped up as they usually do with weekly seminar sheets”. However, since examination boards are quite reasonably reluctant to allow experiments in high-stakes examination situations, it is probably necessary to use mock examinations as this study has done, until staff and students feel properly informed about the implications of using laptops for essay examinations. It is established that students have different understandings of what is expected in an essay in order to achieve a high score (Hounsell 2005) and understanding about what is expected in an examination may be different again. It is also recognised that constructing an essay in an examination is

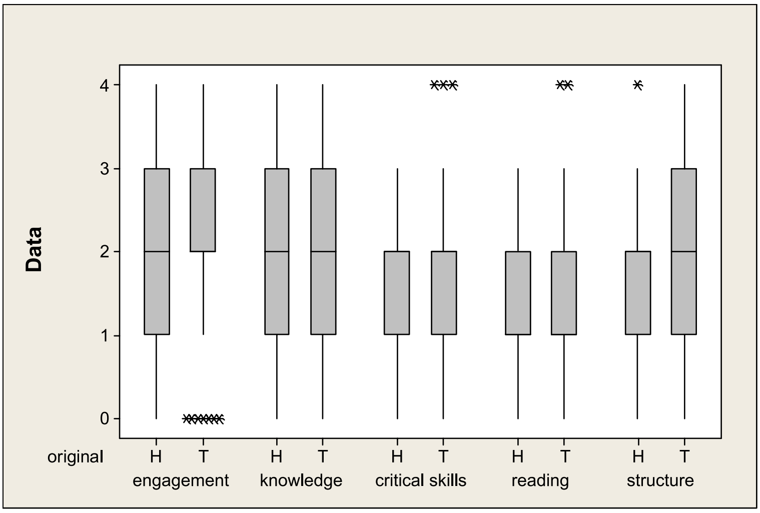


Figure 8. Boxplot of engagement, knowledge, critical skills, wider reading and structure.

Table 5. Correlations for essay quality characteristics.

	Engagement	Subject knowledge	Critical thinking	Wider reading
Subject	0.77			
Critical	0.81	0.72		
Wider	0.69	0.66	0.62	
Structure	0.76	0.60	0.57	0.67

likely to be a different process to constructing an essay for an assignment, even if provided with identical tools.

Similarly the markers would also have been aware that this study was artificial and so it was less vital that their marking was absolutely accurate. The four tutors in this study were all Divinity PhD students (one completed, one sitting viva, one in the final stages of writing up and one mid-way through). All of them had to juggle marking with the tight deadlines the project demanded for turnaround, and this study asked them to mark more papers than would usually have been the case. This may have contributed to the variability observed. Generally, postgraduate students typically tutor for only two years, and it has been found useful to identify one lead tutor with prior experience of the course. Nevertheless, it is observed that tutors generally tend to mark more severely than the course organiser would.

In conversations with staff and students, typing speed is frequently presented as a major concern and cited as a source of inequity. In line with other research, this study has also demonstrated that students who typed have, in general, written more than students who wrote by hand. However, a further issue here is familiarity with how much you should write for an examination answer and what that looks like in handwritten or typed format “I had no concept of how much I had written, with a hand written exam you aim to write about three sides of A4”. This is as much an issue for markers as for candidates – it is possible that a typed response is perceived by the marker as being too brief and therefore marked down. It would have been interesting to explore marker perceptions in more detail.

Lessons learned

Clearly students type at different speeds and inevitably this does have an impact on how much they are able to write in an examination. If students are given sufficient warning of any requirement or opportunity to type their examinations they need not be unduly penalised. Poor typists can be supported to learn this skill if desired. Students with special needs will continue to be considered individually so no student needs to be unfairly disadvantaged. Connelly, Dockrell, and Barnett (2005) demonstrated that first-year undergraduates had a handwriting fluency similar to that which would be expected in 11-year-old children. They found most students have little requirement to handwrite and their handwriting fluency is therefore limited. Hence, the assumption that all students are equally able to handwrite examinations is fundamentally flawed.

One student elected to borrow a machine because they were concerned the software would 'trash' their machine. Although security and reliability issues have been raised frequently, it is almost always by those with only limited knowledge or experience of the software or the procedures proposed, and can be countered by pointing to successful examples of implementation. We have taken the view that because student laptop ownership is known to be above 90% (from Freshers' survey data), and because some laptop keyboards feel quite different from others that most students will be most comfortable using their own machine. Power will be provided to all desks so battery life should not be a concern, beyond that students are expected to provide a machine in an exam-worthy state, or to request a loan machine.

Thus we have arguments to counter concerns about the technical reliability of the solution, and concerns about variation in typing speed, but a concern about variation due to marker differences remains and whether markers might consciously or subconsciously influenced by the appearance of a script. Previous studies have shown a small but consistent effect when marking handwritten originals and their typed transcripts (Powers et al. 1994 ; Russell and Tao 2004a). Russell and Tao (2004b) concluded that computer-printed scripts would score on average 1.3 points less than the same words in a handwritten script. Our study agrees that markers may indeed be influenced by format – and that difference might be worth almost two marks to the average student (55.12 – 53.19 = 1.93). Such variability could of course be removed by ensuring all markers were only given scripts in one format, but the cost of transcribing large numbers of scripts almost certainly render this impractical. Russell and Tao (2004b), however, demonstrated that giving the markers typed scripts printed in cursive font, and alerting the markers to the format effect, both had the effect of reducing the difference in the score; both approaches may be practical to implement. Variation between markers is normally controlled by school and college quality assurance processes; there is no reason why marking of digitally generated scripts should alter established procedures and guarantees of fairness.

The original aim of this study was to explore differences between handwritten and typed answers, not to explore consistency, or lack of it, between markers. However, it has become apparent that the variation between markers was much more significant than the variation between fast and slow writers and typists. There is time here only to acknowledge that variation between markers is of course not a new phenomenon.

Bloxham (2007) presents an interesting summary of current marking practices and research in the area. Even the use of clear assessment criteria combined with careful briefing of markers does not eradicate the variability (Hanlon, Jefferson, and Molan 2005). Brown, Bull, and Pendlebury (1997) argue that the lack of consistency for an individual marker is even more critical than variation between markers because of the difficulty in making fair adjustments when a marker's standard is not constant.

Standard practice for the School of Divinity is for in-course assessment to be marked by tutors, but moderated by the course manager in order to pick up any severities in marking. Tutors in Divinity go through the normal university training scheme. While tutorial sheets (10 x 2%) are marked by tutors for their own tutorial groups, essays (anonymous) are distributed randomly among all the course tutors. Thus coursework from any one student is unlikely to be marked only by a single tutor. Any disparities of marking tend to be picked up at in-course assessment level; marks are returned regularly to the course manager. From the marking spread, it is possible to identify tutors who may be more severe in marking and those who may be more generous, and where necessary marks may be lowered or raised by the course manager to ensure parity. Where a student may be in danger of failing, then the course manager looks at border-line cases and reviews these, both in the elements of in-course assessment and final examination assessment prior to these being sent to the external examiner.

The University of Edinburgh is a traditional, research-led university. Change is generally easier to implement successfully in a series of small steps. Thus from the outset the object has been only to investigate the possibility of allowing students to type rather than to handwrite essay examinations. At the moment there is no intention to move towards either digital or automatic marking of essays. Perhaps once staff become accustomed to having access to digital scripts it will be a small and logical extension to mark them digitally. This could then offer benefits in terms of marking consistency and speed (Sargeant, Wood, and Anderson 2004) and also offer an opportunity to increase the quality of feedback students receive on their examinations.

In tandem with this study, work has been undertaken to equip a large examination hall with sufficient power and network access to make it suitable as a venue for essay examinations on laptops. Desks will be spaced more widely than in paper-based examinations because experiments suggest that at standard examination desk spacing a laptop screen could possibly be read by an adjacent student. Similarly it has been decided to use a slightly larger desk to allow students space for rough paper (or paper for diagrams) and the question paper. In time, further venues can be adapted in line with demand. Clear examination processes and procedures are being developed and it is hoped to give some specialist training and support to invigilators.

Some wider implications

This study has caused the project team to reflect more generally about assessment and assessment strategies. Although aware of the need for aligning teaching, learning and assessment, we recognise that some students are mainly strategic learners focused on passing examinations, whereas others are seeking a deep understanding of their chosen subject. Setting examination questions may need more consideration and direction – for example, in Christian Theology 1: Rather than 'Discuss the ways in which God reveals himself in history', there may need to be a further instruction: 'Your answer should indicate a map of your response, substantive engagement with the question, and a final concluding paragraph of your own'. We need to teach/learn students how to sit examinations and to clarify how they differ from coursework essays.

There are other assumptions made about assessments that should perhaps be challenged. As the Internet and other digital information sources become more accessible, more mobile and more ubiquitous, is it not appropriate to challenge (in some disciplines) the need to remember facts in preference to the ability to synthesise and present a coherent and convincing argument? Similarly, as students expect more choice and a more personalised learning environment, would it be appropriate to think about how assessment can also be personalised and tailored, both in content and in timing.

Conclusions

The study was designed to explore any systematic differences in mark awarded due to the format in which the examination script was created. Only small differences have been found in the mark awarded to an examination script depending on the format of the script (typed or handwritten) but the difference is not significant and is trivial compared with variation between markers. It has been shown that students are able to type more than they can handwrite, but this was not associated with their reported typing speed. Students who wrote more tended to get slightly more marks, but again this was not associated with reported typing speed. No evidence has been found of systematic differences in essay quality due to the format, and data about how students report approaching a handwritten essay versus a typed essay from this study is inconclusive; differences may or may not exist. What is clear is that many students find the possibility of electing to type their essay examinations attractive, indeed some express surprise that this is not standard practice.

The problem of students routinely doing coursework on computer but being assessed by a written essay can be tackled in two main ways – change the type of assessment being used or make sure that the practice and the final assessment use the same medium. Discussion about the merits or demerits of the essay as an assessment tool and what is a correct balance between coursework and examinations are not likely to be concluded quickly, hence it is considered essential to correct the mismatch between how students write coursework and how students write examinations.

- Choice 1: All students in a class will type their examinations.
  This is not substantially different from the current position where all students (with the exception of some with special requirements perhaps) are forced to handwrite their responses. It is anticipated that the variation in typing speeds will be greater than the variation in handwriting speeds, but we believe this can be addressed relatively simply by ensuring students have enough prewarning that their examination will be typed – and by providing opportunities to increase individual typing skills. Essentially it would be feasible to assume that typing proficiency is expected of a modern student, just as fluency in reading is currently assumed, even 'though student reading speeds vary greatly'.

- Choice 2: Offer students the choice of handwriting or typing their examinations.
  Boards of studies (who are responsible for quality in courses) have been reluctant to consider this suggestion because it means students are not all doing the

> same thing – and because of a risk that the choice to write or to type might unfairly or unknowingly influence the grade achieved. This study has sought to examine those concerns and where possible to offer some answers.

We have demonstrated that the variation due to difference in format is negligible compared with variation due to differences between markers, and we therefore conclude that although there is evidence of a format effect that we can nevertheless justify giving students the choice of whether to type or to handwrite their essay examinations.

Acknowledgements

This study was funded with support from The Principal's Teaching Award. http://www.tla.ed.ac.uk/centre/PrincipalsTeachingAward/PrincipalsTeachingAward.htm.

Notes

1. Websites of US universities offering essay examinations on computer. http://www.law.nyu.edu/llmjsd/graduateadmissions/laptopcomputerrequirement/index.htm. http://www.courts.mo.gov/page.asp?id=1724. http://www.kyoba.org/forms/exam%20instruction%20manual.pdf.

References

Augustine-Adams, K., B. Hendrix, and J. Rasband. 2001. Pen or printer: Can students afford to handwrite their exams? *Journal of Legal Education* 51: 118–29.

Bloxham, S. 2007. Marking. In *Developing effective assessment in higher education: A practical guide,* chapter 6. Maidenhead, UK: Open University Press.

Bridgeman, B., and P. Cooper. 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, April 13–17, in San Diego, CA, USA.

Brown, G., J. Bull, and G. Pendlebury. 1997. *Assessing student learning in higher education.* London: Routledge.

Burke, J., and G. Cizek. 2006. Effects of composition mode and self perceived computer skills on essay scores of sixth graders. *Assessing Writing* 11: 148–66.

Connelly, V., J. Dockrell, and J. Barnett. 2005. The slow handwriting of undergraduate students constrains overall performance in exam essays. *Educational Psychology* 25, no. 1: 99–107.

Goldberg, A., M. Russell, and A. Cook. 2003. The effect of computers on student writing: A meta analysis of studies from 1992 to 2002. *Journal of Technology Learning and Assessment* 2: 1–51.

Hanlon J., M. Jefferson, and M. Molan. 2005. An examination of the incidence of 'error variation' in the grading of law assessments. http://www.ukcle.ac.uk/research/projects/mitchell.html.

Hartley, J., and P. Tynjälä. 2001. New technology, writing and learning. In *Writing as a learning tool: Integrating theory and practice,* ed. P. Tynjälä, L. Mason, and K. Louka. Dordrecht, The Netherlands: Kluwer.

Horkay N., R.E. Bennett, N. Allen, B. Kaplan, and F. Yan. 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, no. 2: 4–49.

Hounsell, D. 2005. Contrasting conceptions of essay-writing. In *The experience of learning: Implications for teaching and studying in higher education,* ed. F. Marton, D. Hounsell, and N. Entwistle, 106–25. 3rd ed. (Internet). Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment.

Howell S. 2003. e-Learning and paper testing: Why the gap? *Educause Quarterly* 4: 8–10.

Lee Y. 2002. A comparison of composing processes and written products in timed essay tests across paper and pencil and computer modes. *Assessing Writing* 8, no. 2: 135–57.

MacCann R., B. Eastment, and S. Pickering. 2002. Responding to free response examination questions: Computer versus pen and paper. *British Journal of Educational Technology* 33: 173–88.

Mogey N., M. Purcell, J. Paterson, and J. Burk. 2008a. Handwriting or typing exams – can we give students the choice? Paper presented at the 12th CAA Conference, Loughborough, UK.

Mogey N., M. Purcell, J. Paterson, and J. Burk. 2008b. Essay exams on computer: Fair or unfair? Paper presented at ALT-C, Leeds, UK.

Mogey N., and G. Sarab. 2006. Essay exams and tablet computer – trying to make the pill more palatable. Paper presented at the 10th CAA Conference, Loughborough, UK.

Mogey N., G. Sarab, J. Haywood, S. van Heyningen, D. Dewhurst, D. Hounsell, and R. Neilson. 2008c. The end of handwriting? Using computers in traditional essay examinations. *Journal of Computer Assisted Learning* 24, no. 1: 39–46.

Penny, J. 2003. Reading high stakes writing samples: My life as a reader. *Assessing Writing* 8, no. 3: 192–215.

Powers, D., M. Fowles, M. Farnum, and P. Ramsay. 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: 220–33.

Russell, M., and W. Haney. 1997. Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via pencil and paper. *Education Policy Analysis Archives.* http://epaa.asu.edu/epaa/v5n3.html.

Russell, M., and T. Plati. 2001. Effects of computer versus paper administration of a state-mandated writing assessment. TC Record. http://www.tcrecord.org/Content.asp?ContentID=10709.

Russell, M., and W. Tao. 2004a. Effects of handwriting and computer print on composition scores: A follow up to Powers, Fowles, Farnum & Ramsay. *Practical Assessment, Research and Evaluation* 9. http://pareonline.net/getvn.asp?v=9&n=1.

Russell, M., and W. Tao. 2004b. The influence of computer print on rater scores. *Practical Assessment, Research and Evaluation* 9, no. 10. http://pareonline.net/getvn.asp?v=9&n=10.

Sargeant, J., M.M. Wood, and S. Anderson. 2004. A human–computer collaborative approach to the marking of free text answers. In *Proceedings for 8th CAA Conference 2004,* 361–70. Loughborough, UK: Loughborough University.

Thomas, P., C. Paine, and B. Price. 2003. Student experiences of remote computer based examinations. Paper presented at the the 7th CAA conference, July, Loughborough, UK.

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

Wolfe, E.W., S. Bolton, B. Feltovich, and D. Niday. 1996. The influence of student experience with word processors on the quality of essays written for a direct writing assessment. *Assessing Writing* 3, no. 2: 123–47.