MOVING TOWARDS ESSAY EXAMINATIONS WRITTEN ON COMPUTERS

Moving towards essay examinations written on computers

A study funded under the Principal's Teaching Award
The University of Edinburgh
September 2008

EXECUTIVE SUMMARY

For some years, staff at The University of Edinburgh have felt concerned that students do almost all their coursework on computers, but at the end of the semester they are examined using handwritten essays. One solution has been explored – whether students could, in a conventional exam setting, use a computer to author their essay-examinations rather than handwriting them. This has raised a number of questions and issues, some practical, some more psychological. From the outset the student body has spoken in favour of choice – let students choose whether to type or to handwrite. This is an intuitively appealing approach but leaves a lingering sense that perhaps that might not be "fair" – that the choice a student makes, to write or to use a computer, will affect their mark.

In the USA, law schools have been routinely running essay style exams on computers for many years. There is no choice, all students use computers. The leading software used in these examinations is Exam4, supplied by a company called Extegrity. Students download the software onto their own laptops, and the software ensures the examination is conducted securely Exam questions can still be distributed on paper in the exam hall and the submitted scripts can be printed and marked on paper, so the only part of the process that is different is that students do not handwrite their answers. The Exam 4 software was identified as suitable for this study.

The aim of this study was to identify and explore any systematic differences which may be introduced due to offering students the choice to handwrite or to type essay examinations. A class of 70 1st year divinity students were given the option of taking a mock exam, and the further option of using a computer or handwriting their answer. All the exam answer scripts were then faithfully transcribed into the opposite format so there was a printed and a handwritten copy of every script. These were then marked by 4 markers, such that every marker marked every script once, either in one format or the other.

No significant differences could be identified due to the format in which the students had written their answer. A variety of factors were explored including length of essay, overall score awarded, and some qualitative measures designed to indicate essay quality. The results indicated that, in line with existing research, markers showed a small tendency to mark the handwritten version of scripts very slightly higher than the typed version was noted, but it is suggested this could be minimised with some additional training for the markers.

Another finding was that there was a striking contrast in the variation in marks awarded between the (4 experienced) markers, and this variation was observed for most of the factors explored. More worryingly the variation between markers was not consistent. It was not that one marker consistently gave high marks and another low marks; the range of marks varied in different ways for different markers and differently for different factors being explored.

The study concludes that there is no reason why students could not be offered the choice to handwrite or to use a computer to type their essay examinations, this choice would not have an undue influence on the mark they would be awarded, and may help minimise the stress and anxiety felt at an already stressful time.

1.1 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*. "
>
> James Penny in Assessing Writing 8(3) 192-215

For some years, staff at the university of Edinburgh have felt concerned that students do almost all their coursework on computers, but at the end of the semester they are examined by handwritten essays. The issue was discussed in detail by a small group at the Change Academy in Sep 2004 (David Dewhurst, Jeff Haywood, Dai Hounsell, Nora Mogey, Sarah Nicholson, Craig Shearer, Simon van Heyningen) and this project is one eventual outcome from those discussions.

Others have asked questions about whether assessment methodologies are sufficiently aligned with teaching practices (Howell, 2003; Thomas, 2003) and possible solutions include some potentially major changes in the whole assessment procedure.

One option might be to offer students the choice to answer essay style exams using a computer or to handwrite. However this introduces additional complexities – the exam venue has to be suitable for both modes, or two venues are needed; variation in students keyboard skills is much greater than variations in handwriting speed; students with low IT confidence may be particularly disadvantaged and any marking variations between typed and handwritten scripts must be considered.

An easier approach is simply to require all students to sit their exams on a computer, and indeed this is already routine practice in many US law schools. After a small amount of exploratory work Exam 4 from Extegrity Inc was identified as the software used successfully in many US bar exams. Exam 4 has proved itself to be robust and reliable, and includes security features which were considered important, such as taking regular snapshots of the exam in progress, and the option to totally lock down the machine from

accessing any applications other than the exam software. (More details about the software are provided separately.)

It was then decided to undertake some early pilot studies with students. Initial concerns of the project team were that some students may wish to include a diagram or a table in their exam answer, and that this is difficult using a laptop keyboard. The solution which was proposed was to use tablet PCs thus giving the option to use the keyboard for text, or to use the tablet for inserting diagrams. The software authors were generous in their willingness to add functionality into their software to facilitate the inclusion of diagrams, charts and tables. Three different pilot studies have taken place since 2006, with subjects drawn from the student's association; 4th year biological sciences and a small MBA class who used Exam4 for a summative assessment – at the request of the students. These studies have 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 exam, and there has been a great deal of caution on the part of examination boards and boards of studies when course teams have sought permission to use this tool.

The idea of using tablet PCs to facilitate drawing diagrams is now not considered to be critical: the students' lack of familiarity with tablet PCs and the mechanism to rotate the screen was felt to be problematic, and in practice the university would provide paper to those students wishing to add diagrams to their answer. The diagrams can then easily be attached to the typed script, and in many disciplines is not a relevant consideration.

Not surprisingly all the MBA students (n=5) who requested to type their exams thought the idea of sitting essay exams using a computer was good, and said they would be happy to take an exam this way again. Of the 10 final year biology students, six were happy with the notion of essay exams on computer, two were not supportive and two were undecided. The main concerns were about typing ability and whether the software would crash, while the biggest perceived advantage was 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."*

So although support from the students has been consistent, it has also consistently been with a sense of reservation – it is an interesting idea to explore but is it really fundamentally fair ? Equally college and school exam boards have been reluctant to take the decision to move to keyboarded exams until they are confident that this will not result in a rush of student appeals.

Throughout this report the tendency is to use the terms handwriting and typing. "Typing" has generally been used rather than "word processing" or "using a computer" but in reality a hybrid meaning is usually intended. "Typing" is in no sense intended to mean simply pressing keys on a keyboard, it is taken here to include limited word processing functionality particularly cut and paste. It also encompasses an expectation that the technology a student uses to write their essay (either pen and paper or keyboard and computer) will not be independent of their cognitive processes.

This research seeks to answer some of the questions that are currently acting 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 ?
- Is there generally a difference in the amount of text that can be written or typed in an examination?
- Are students who type slowly any more or any less disadvantaged than students who handwrite slowly ?
- Do students perceive typing examinations to be as fair as handwriting responses to examinations?
- Do students report approaching the construction of an essay response differently when using a keyboard or handwriting ?

1.2 EXAM4 from Extegrity Inc

The idea of using computers to allow students to type responses to essay style questions is not new. US law schools particularly have extensive experience of exactly this, and have established procedures (e.g. Duke Law School http://www.law.duke.edu/curriculum/examonComputer.html, accessed 31 August 2006).

A small number of broadly similar commercial products were identified for potential testing. A key choice was whether to opt for a familiar software environment for the exams, such as Microsoft Word, or a very simple but less familiar word processing tool. It was felt that the use of a very simple tool would be more appropriate in that it was less likely to confound the mark for the academic content of the exam with a measure of the student's skill in using a particular word processor.

Exam4 marketed by Extegrity Inc. was selected and evaluation access was agreed with the vendor. Exam4 was attractive for a number of reasons. It has a long record of being used in high stakes exams across the USA, where it has proved to be stable and reliable (R. O'Hara, personal email). The software includes a range of excellent 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; 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 exams.

Exam4 needs to be pre-loaded onto the computers to be used via a familiar download-and-double-click operation that can be done easily and quickly by students themselves onto their own machines, or by IT staff onto University-owned machines. In the initial evaluation, the software was easily loaded and installed, but it immediately became obvious that its security mechanisms were so efficient they entirely locked down the computer, rendering the tablet's pen functionality inaccessible. This resulted in Extegrity agreeing to modify the software to provide drawing capability.

When launching the software the user follows a channelled, stepwise exam startup procedure, selecting from a series of simple menus or entering basic personal identification details which together configure all the pertinent administrative settings (e.g. saving) without the need for issuing any complicated instructions. Student selectable options include the choice of 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 exam startup sequence has been completed, the student clicks a button to begin the exam 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.

At its core Exam4 provides a simple word processor with basic formatting abilities: bold, italic, underline and left, centre or right justification. Certain

features and options are set in advance at the direction of the examiner and are part of the download. These may dictate whether features such as a spell-checker or cut, copy and paste are globally available, as well as the specific location where exams are submitted via the network. The interface is straight forward and intuitive, so users require only the most cursory orientation; no formal training is necessary.

A number of additional tools are available accessed via a 'Tools' menu as required. A basic spreadsheet is now incorporated along with some MCQ functionality. Exam4 also includes a simple figure drawing and editing interface. On tablet PCs sketches, diagrams and the like can be drawn with the tablet stylus or an external graphic s devices can be used, without compromising the security features of the system. These drawings become associated with the text portion of the exam file, and are printed automatically along with the essay.

An exam can be administered in different ways using Exam4. It was our intention to minimize 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 exam these are retained in encrypted format on their hard drives as well as transmitted to a specific nominated computer which can be located essentially anywhere. Transmitting functionality is built into the Exam4 student software, and exams are received by a small server application.(ExamReceiver) which is installed on a computer under the control of the exam administrator. A separate administrative tool, ExamOpener, is then used to print all the exam 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 changes significantly can be that students no longer handwrite their answers.

http://64.84.16.214/extegrity/MainFrame.asp

1.3 Literature

Very few relevant studies have been identified which provide empirical evidence relating to university students under examination conditions, however we do know that

Students with good IT skills do better at online tests than students with poor IT skills : Unsurprisingly there is indeed evidence that students with good IT skills perform better at online writing exercises. Horkay et al (2006), studying school pupils, found that hands on experience was significantly related to online writing assessment performance - computer familiarity added about 10% to the score achieved.

Students who are used to word processing do better when they type than when they handwrite : Russell & 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.

Students *may* write to a higher standard when using a word processor : In one of the few higher education examination studies Augustine-Adams et al (2001) concluded that on average a law student typing an exam could expect to perform slightly better than their colleague who handwrites. In other sectors there is substantial evidence that students who have written their essays using a computer write to a better standard (MacCann, Eastment & Pickering 2002; Russell & Plati 2001; Goldberg, Russell & Cook 2003; Hartley & Tynjala 2001) **However** Burke & Cizek found the opposite and demonstrated that irrespective of IT skills or confidence sixth graders produced better essays by hand than they did using a word processor !

Text composed using a word processor is subject to more revisions than text composed on 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. Similarly how sentences are constructed differs, and word processed sentences which can of course be easily revised achieve higher sentence structure scores (Burke & Cizek, 2006)

Students typically type more than they handwrite (Russell & Haney, 1997; Russell & Plati 2001; Wolfe Bolton Feltovich & Niday, 1996; Lee 2002).

Students will write more in a real exam than in a practice: Thomas (2003) This study with Open University computer science students demonstrated that typically each student submitted 30% less material in the mock exam than they eventually submitted in the final examination.

A type-written essay will be marked more harshly than an identical handwritten text : Several studies have demonstrated this effect, although the difference in scores is not always large. (Russell & Tao 2004, MacCann et

al 2002; Bridgeman & Cooper 1998) The reason for the difference is not known for certain but seems likely to be associated with an expectation that hand written work is essentially a first draft standard whereas typed text would normally have been more thoroughly revised. Markers seem to be more aware of surface errors in typed work than they are in handwritten scripts (Whithaus, Harrison & Midyette, 2008).

Essays can be too long – slightly shorter essays score more highly than very long essays this was research which explored university students exam essays (Whithaus Harrison & Midyette 2008)

2 METHODOLOGY

Christian Theology 1 is a class of about 70 first year students. Compared to other schools Divinity has an unusually high proportion of 'mature' students, this was reflected in this class.

The students were invited to sit a 'mock' examination during timetabled class time, during week 11 of a 12 week semester. The Exam4 software was demonstrated during a class in week 8 and students had the opportunity to try it out and become familiar with the software on their own laptops. Technical support was available on request and laptops were available for loan if needed. Students were allowed to sit the exam in the format of their choice: typing using a laptop (which could be their own or could be provided on loan) or handwriting onto paper, or they could decide not to sit the mock examination at all. Prompt feedback was promised to all students who opted into the study.

The mock exam was held in normal class time and was therefore necessarily shorter than a full exam would be. Students were allowed sight of the exam questions one week in advance in recognition that the mock was timed before the end of normal teaching. It was hoped this would motivate and assist students to prepare properly, but recognising that revision time would be limited. Students had a choice of one question out of three.
Q1. God reveals himself in history. Discuss
Q2. Choose any model of the church and explain its merits and criticisms
Q3. Sin disfigures the divine image in the human person

The mock examination was held in the regular class venue but under examination conditions. Students using laptops were mostly situated towards the front of the room, and all had access to power sockets. Students who were handwriting the exam were seated at the back of the room. All students were provided with scrap paper, which in this case was not collected in.

At the end of the exam 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, and these were both in turn duplicated and then blind marked. Four marks for each student script were generated, one from each of four markers, two for typed versions and two for handwritten versions. All the markers were experienced at marking first year divinity essays.

In addition to the scores for the scripts markers were asked to rate the scripts on 6 qualitative dimensions : 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

The total number of words written during the mock exam by each student and the number of words in any conclusion paragraph(s) were recorded.

Questionnaires were given to all participating students pre and post the mock exam seeking information about students' confidence with and attitude to IT, and about their preferred strategy when constructing examination essays. A small focus group was also held to explore student concerns about essay examinations on computer in more depth.

The collated data was compiled initially into Excel. Some basic cleaning was needed (typically where students had used something other than their matric number to identify themselves). The subsequent analyses were performed using Minitab.

3 RESULTS

3.1Description of the sample

Some students completed pre-questionnaires without sitting the mock exam, and some students who sat the mock exam failed to return the questionnaires resulting in some missing data.

There are mock exam marks for 37 students and there is Pre questionnaire data from 42 students These two items are probably the most important items for the purposes of the study.] Overall data of some sort was collected from 51 students as shown below :

Pre-questionnaire only from 14
Mock Mark only from 6
Pre Questionnaire plus mark only from 16
Post Questionnaire plus mark only from 3
Pre- questionnaire, mark and post Questionnaire from 12

69% (n=35) of the sample was female and 31% (n=16) male.

34 students were in the age range 18-21 years and 8 in the age range 28-54 years.

Comparison with the 'freshers survey' suggests the sample matches the undergraduate population as a whole very well for gender and is slightly biased in favour of more mature students.

37 (88%) students reported using a computer either daily or as frequently as possible, only two responded "less than daily" (both female). The students reported themselves as confident users of technology – 15 responding they were "very confident" and 25 responding "more or less OK". No one responded with anything lower than this (perhaps there is a reluctance to admit a low level of confidence ?).

A very even spread of typing ability was reported 12 saying they type faster, 14 saying they handwrite faster, and 12 students saying there was no difference in typing and writing speeds. Proportionately more females reported being faster handwriters and the males faster typists. Typing speed was not associated with IT confidence level.

Considering the group who took the mock exam a slight bias towards females and mature students has been but they do seem to represent a reasonable spread of academic ability.

3.2 Student attitudes and reactions to the idea of writing essay exams on computer

Do students report approaching the construction of an essay response differently when using a typewriter or handwriting ?

Students were asked about the preferences and approaches to writing essays – both for assignments and in examinations. 86% (n=36) said they would prefer to type an assignment. Reasons cited for the preference included legibility, ability to edit and rearrange text, and speed. Only 6 students said they would prefer to handwrite an assignment (one of whom reports typing faster than writing). Almost all students will use an essay plan (only 4 did not – 1 female and 3 males) but males are more likely to use a typed plan and females a handwritten plan. (This matches with the preference shown by females for writing over typing.) There was no obvious link between IT confidence and tendency to plan but those who write fast are more likely to handwrite a plan.

Moving on to examinations, many students report approaching them differently from assignments :

	n
I would not have any difference in approach	11
I would write a briefer plan	9
I would write a more detailed plan	8
Some other difference reported	9

Other differences about exam-essays included comments such as

- *"I try to conceptualise the whole thing in my head before starting. I spend more time on constructing sentences in advance before writing"* (male student who prefers to type, equal typing and writing speeds)
- *"Spider diagram the essay"* (female who prefers to handwrite but no difference in typing and writing speeds)
- *"My writing tends to flow more because I know that I can not easily alter what I have written"* (female who prefers to type assignments but handwrites faster)

Do students perceive typing examinations to be as fair as handwriting responses to examinations?

From the outset students have expressed concern about variability in typing speed a typical comment is *"It's not a level playing field as some people can type a lot faster than others."*

11 students thought they would write more using a computer and 11 thought they would write less, 10 though it wouldn't make a difference and 5 didn't know. As expected, those who reported themselves as fast typists thought they would write more (Table 1)

	Type faster	Write faster	No difference	Total
Type more	8	2	1	11
Write more	0	9	2	11
No difference	4	0	6	10
Don't know	0	0	5	5
Total	12	11	14	37

Table 1 : Students' expectations of whether they could type or handwrite more in an exam compared to reported writing speeds

When asked whether they would do better or worse as a result of typing an exam again the responses tended to reflect typing ability (Table 2)

	Type faster	Write faster	No difference	Total
Do better on PC	8	1	4	13
Worse on PC	1	9	2	12
No difference	2	0	1	3
Not sure	1	2	7	10

Table 2 : Students' expectations of whether they would do better or worse using a computer compared to reported writing speeds

Overall students did tend to think that using a computer would make a change to the quality of the work they produced in the exam, making comments in two broad areas : one to do with WP functionality (the ability to change layout and structure of the text) and one to do with fluency of thinking processes. But they were split as to whether these differences would improve or reduce from their score :

Impact on Quality of Essay	Count
no change	12
positive change, structure & layout	7
negative change, structure & layout	1
Positive change, thinking processes	6
negative change, thinking processes	6

In response to the direct question "Are essay exams a good idea?" , 10 students responded with broadly negative comments, 18 with broadly positive comments and 6 specifically mentioned the need to offer choice.

Positive Comments included

- *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., touchtyping) 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.*

Pro Choice comments included

- *Good idea to have a choice to make it fair on both those who type faster and those who write faster*
- *It's hard to tell till we try it. I definitely think it should never be obligatory.*
- *Not for me personally but I think it's a good idea to be an option.*

As expected fast typists want to type, and fast writers want to write. But there was strong support among the 'no difference in speed' group for typing exams (Table 3).

	Type faster	Write faster	No difference	Total
Positive reaction	7	1	10	18
Negative reaction	2	7	1	10
Give choice	0	4	2	6

Table 3 : Summary of open responses to "Do you think it is a good idea to use computers for essay exams? Why?" compared to reported writing speed

The two fast typists who would prefer to handwrite exams commented

- *"No. Too much stress"*
- *"Not really; I think it's a bit unnecessary & computers have a tendency to go wrong."*

And the fast handwriter who would like to type exams said

- *"Yes I do. The process of writing is different. Coming back to university after working I have had to relearn pen and paper. I think this is a backward step. "*

A number of students specifically suggested that it will be important to offer students choice, other comments included cautions that sufficient warning and

allowing plenty of practice time would be important. It may be of interest that it was only students who were faster handwriters who suggested the need for choice.

3.3 Data from the Mock Exam

37 students elected to take the mock examination, 28 female and 9 male. Proportionately more females opted to handwrite the mock, and the older students tended to be more likely than than the younger students to opt to take the mock exam. Using scores from two previous pieces of coursework allows us to ask if it was the students who had been more successful so far who elected to sit the mock – but there was no evidence this was the case.

	Handwrite	Type	Total
Female	11	17	28
Male	2	7	9
Total	13	24	37

10 students borrowed computers including 5 who had not requested prior access (and it is assumed therefore did not practice with the software). One student requested in advance that they should have a standard keyboard and mouse not just the laptop keyboard, and this was arranged.

24 typed and 11 handwritten scripts were collected at the end of the mock examination, 2 additional handwritten scripts had been generated one day earlier by students who were unable to attend the scheduled class time. The total number of words written was recorded for all scripts.

Females tend to write slightly more than males but there was no association between words written and the student's age.

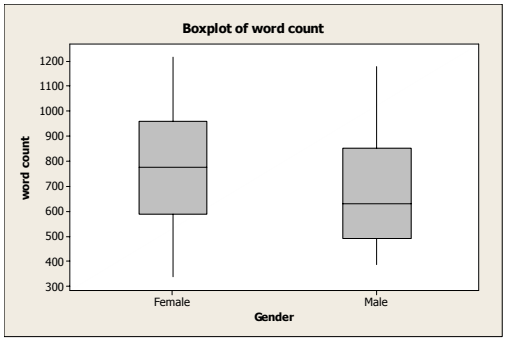


Figure 1 : Boxplot of word count by gender

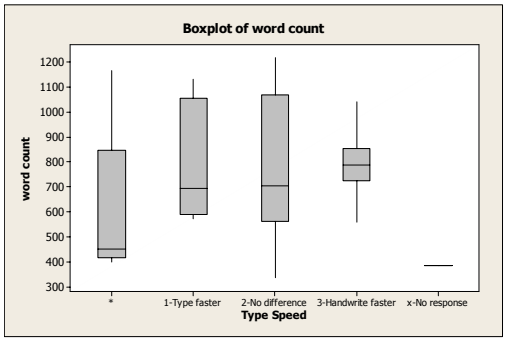
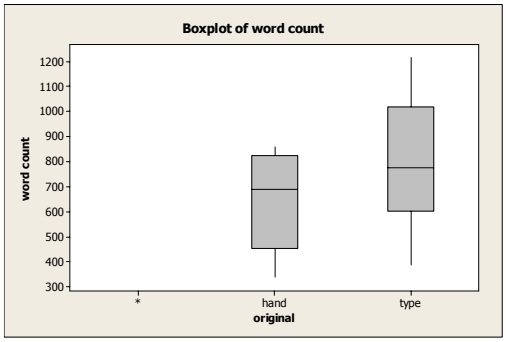


Figure 2 : Boxplot of word count by reported typing speed

Did the students who said they were good typists actually write more in the mock exam? No - there was practically no difference in the number of words written by those who reported themselves as fast typists and those who reported themselves equally fast at typing or writing. Fast handwriters wrote more on average and there was much less variation in the volume written by fast handwriters than the other groups. Those who didn't express comment about their relative speed of writing vs typing wrote noticeably fewer words than those who had responded to the question.

Is there generally a difference in the amount of text that can be written or typed in an examination?

Students who typed in the mock exam wrote more words than students who opted to handwrite.

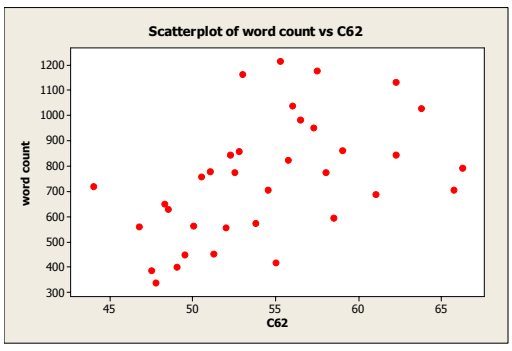


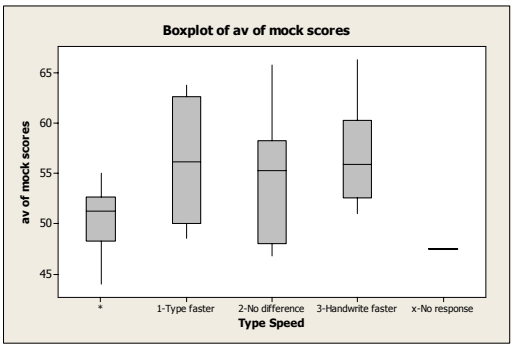
Using a Two-Sample T-Test of the null hypothesis Ho : There is no difference in the mean number of words which will be handwritten or typed, results in T= -2.15, p= 0.041 (25 df) statistically significant at 5% level.

Hence there does appear to be evidence that in general students will type more than they will handwrite, but the amount written is not strongly associated with students reported typing speed. This may indicate that the amount written in an exam is only partially dependent on the speed of writing – it must also depend on the fluency of thought.

There is only data from 8 students to contrast how much they thought they would write with how much they actually managed to write. 4 out of 8 said they had done as they had expected, of those where reality differed from expectation: 2 thought they would type more,1 reported typing less and 1 said they couldn't judge; 1 said they would type less but thought it hadn't actually made a difference and 1 said it wouldn't make a difference but they reported actually typing less.

Are students who type slowly any more or any less disadvantaged than students who handwrite slowly ?





Longer scripts tended to score more than short scripts (correlation = 0.484) (illustrated by the scatterplot). The boxplot suggests that there is no systematic difference in the scores obtained by fast or slow typists. There is a large amount of variability in the data but the highest scores were achieved by students who did not report themselves as being fast typists.

Is the mark awarded to an examination script influenced by the format of the script (typed or handwritten) rather than its content ?

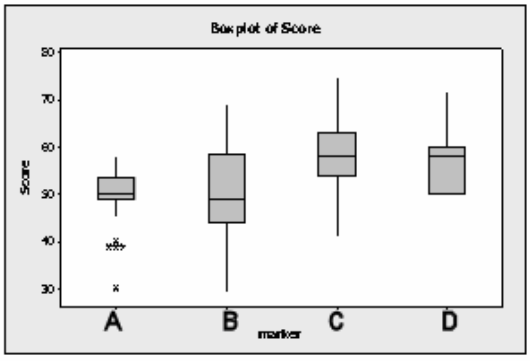
Generally where originals were typed then scripts scored more highly than where originals were handwritten scripts.

Scripts marked in their original formats
Mean Score Awarded handwritten scripts = 52.79 St Dev = 7.13 (n=26)
Mean Score Awarded typed scripts = 54.90 St Dev = 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 more.

All Scripts (including transcriptions) (n=74)
Mean Score Awarded handwritten scripts = 55.12 St Dev = 8.25
Mean Score Awarded typed scripts = 53.19 St Dev = 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 there is evidence that the handwritten scripts are being marked up slightly and the typed scripts marked down slightly. However the format of the scripts are not the only source of variation and the differences between markers are likely to be important.



marker	N	Mean	StDev	Min	Max
A	37	49.54	6.03	30.00	58.0
B	37	50.46	9.32	29.00	69.0
C	37	58.68	7.33	41.00	75.0
D	37	57.95	6.39	50.00	72.0

Using a general linear model to analyse the contribution to variability in scores confirms that the variability due to differences between the markers is the most important effect, and the contribution due to differences in the format of the script are not statistically significant

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				

S = 7.35341 R-Sq = 25.80% R-Sq(adj) = 23.73%

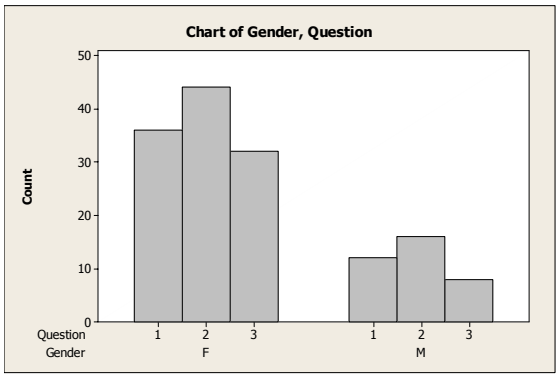
It is of interest to explore the differences between the markers and between the questions in more detail.

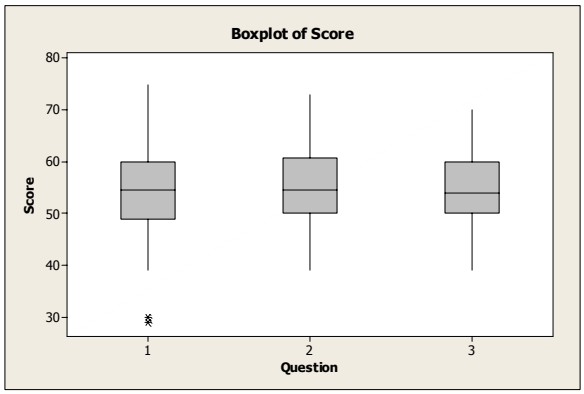
Exploring differences due to the question chosen by the student

Students chose one of three pre-seen questions

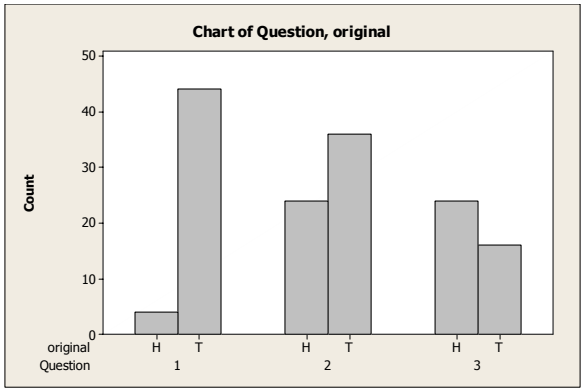
Q1. God reveals himself in history. Discuss
Q2. Choose any model of the church and explain its merits and criticisms
Q3. Sin disfigures the divine image in the human person

12 Students chose to do Q1; 15 chose to do Q2 and 10 chose to do 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 identical difficulty

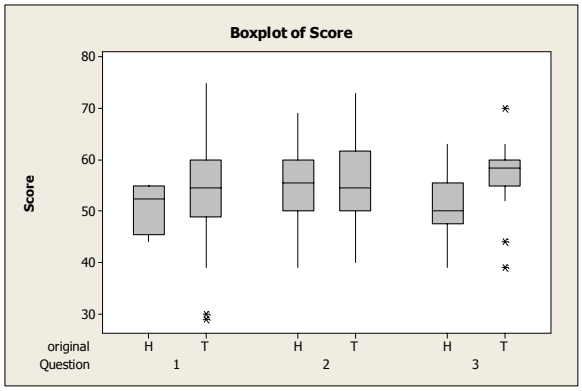




But there is evidence for a strong association between format preference (handwriting vs typing) and question choice; nearly 92% of all the students who chose Q1 also chose to type their responses.

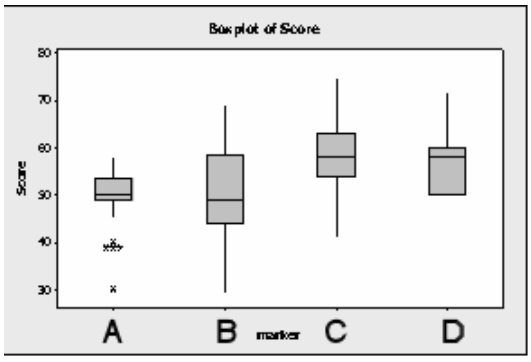


Also noticeable is that of the students who chose to do Q3 the 40% who typed achieved significantly higher scores than the 60% who handwrote their responses. This suggests we might wish to know about why students chose one question rather than another.



Did the markers mark all the questions evenly ?

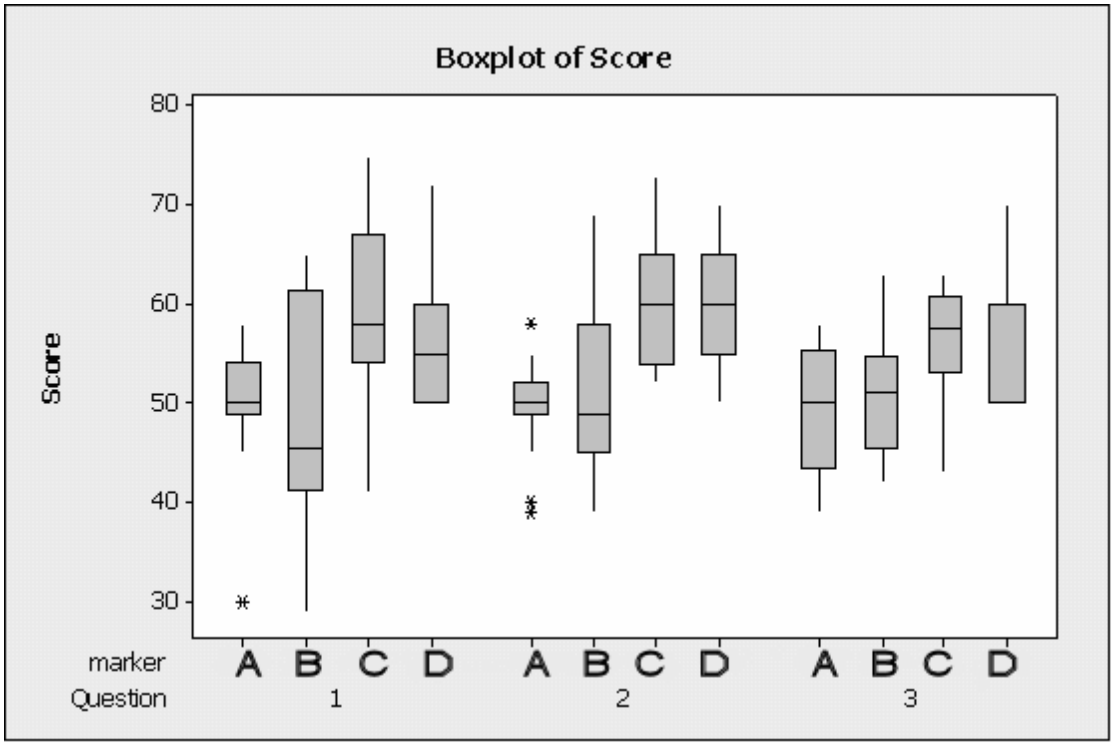
The distribution of marks for each of the four markers is shown in the boxplot below:



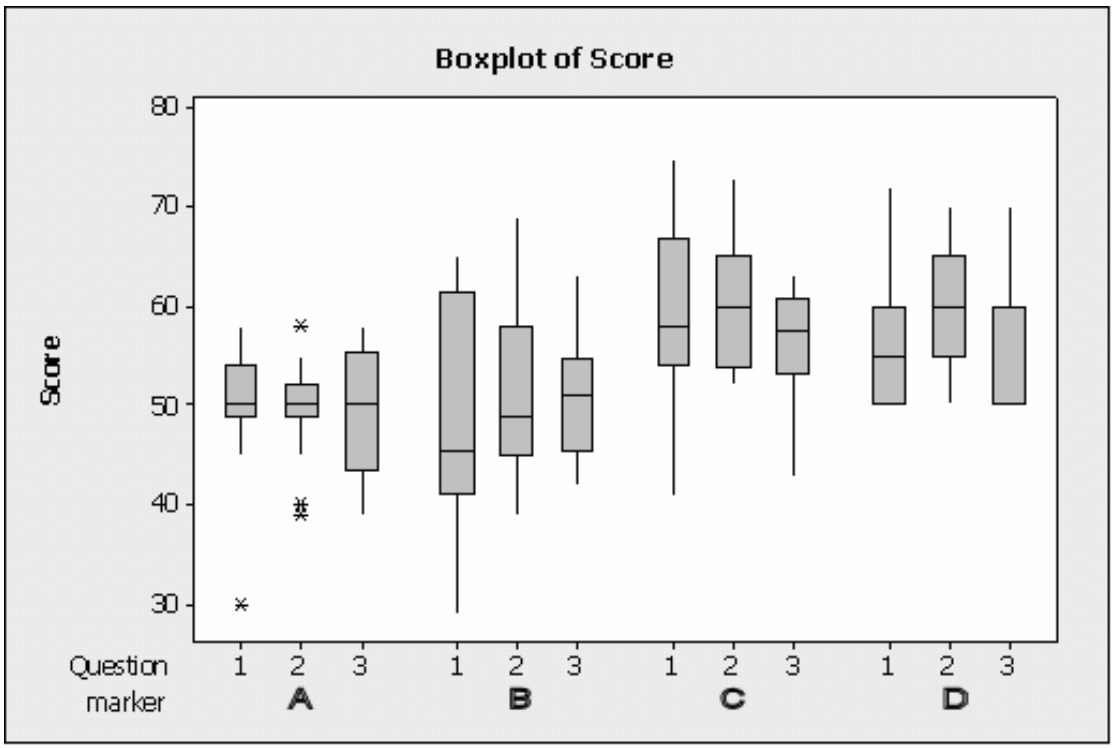
It can be seen that in general marks awarded by marker A are more bunched than those awarded by other markers, and that this marker has a tendency to be unusually harsh on some candidates (indicated by the lowly scored outliers).

Drilling down to examine marking at the individual question level in the boxplots below we do start to see some differences: Marker A has marked question 3 differently to Qs 1 & 2 – there is a much greater spread of scores with a roughly symmetrical distribution and no outliers. Similarly marker D

who's marks tend to show a positive skew, demonstrates a different scoring pattern for question 2 where the marks show a symmetrical distribution.



The graph above suggests that any one question was not marked in the same way by all the markers.

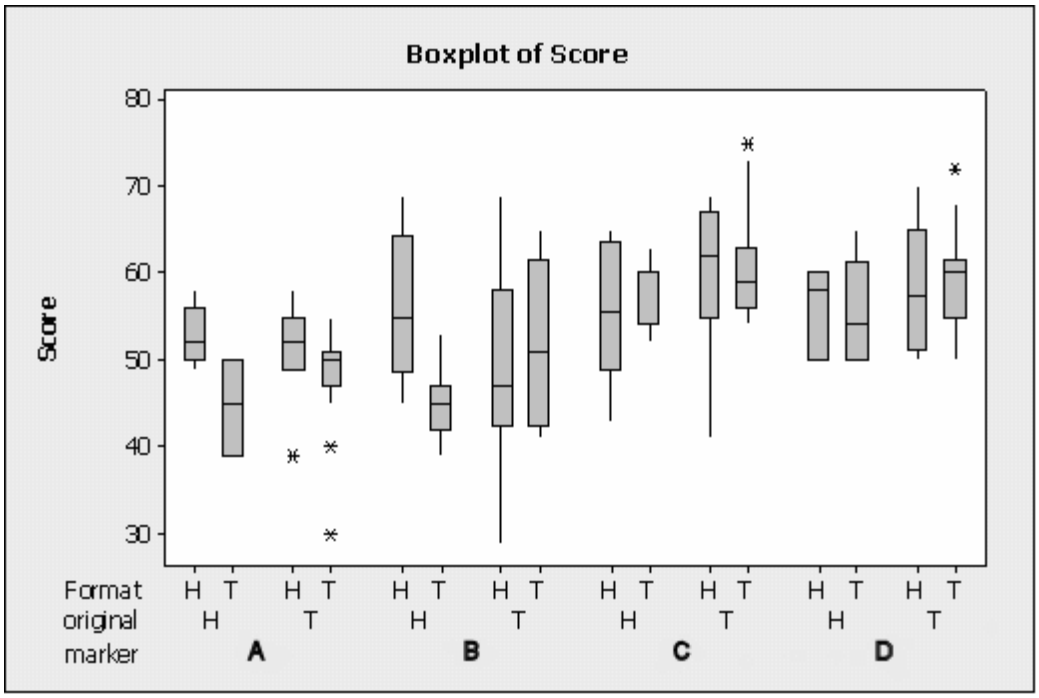


Further, this graph suggests that there is also variation in how any one marker marks the different questions.

So is there a difference in how the markers approach typed or handwritten scripts?

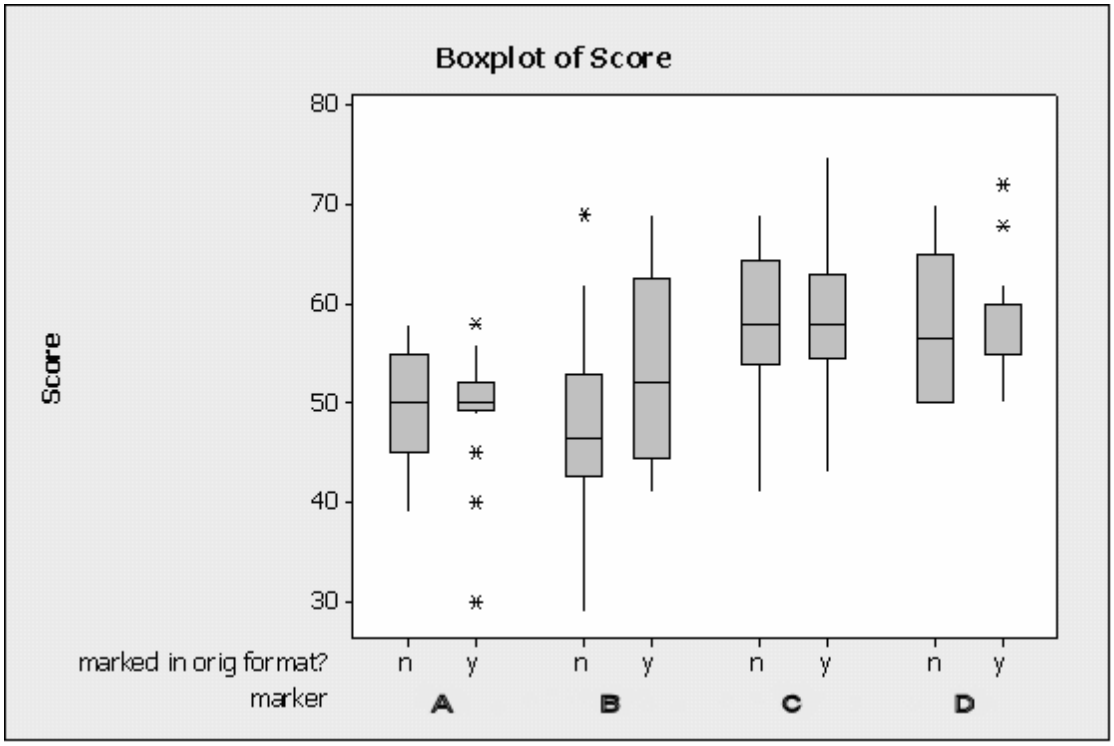
The boxplot below suggests yes there are differences for example marker A appears to mark typed scripts consistently and significantly lower than handwritten scripts. With the other markers there are also differences but they appear to be less consistent or systematic.

It seems reasonable to suggest that with practice marker A's consistent difference in approach could easily be overcome, for the other markers this may be harder to achieve.

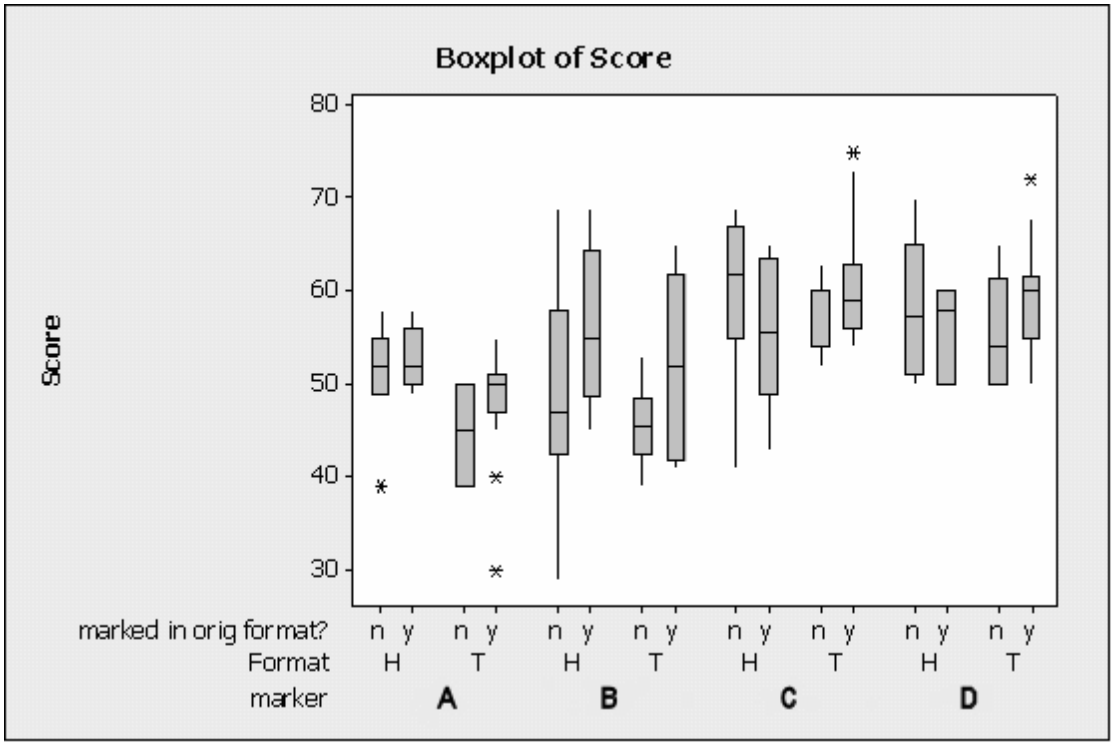


Do markers seem to mark differently if the script was a transcribed version? Are they in some way put off if the script has been duplicated?

The following graph suggests that for marker B there may be an effect – scripts marked in the original format were marked significantly higher than the transcribed versions. For the other markers the originals and the transcripts show broadly similar marking distributions and there is no real evidence of any systematic differences.



Exploring further (graph below) suggests that for marker B the effect is consistent for hand > typed scripts as well as for typed > hand scripts.



3.4 Exploring the Quality of essays

In section 3.2 we saw that the range of scores suggest the three questions were of comparable difficulty, but that there is considerable variation between the markers. A trend associating length of essay with score was also noted.

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 all the markers were required to grade each essay on six characteristics

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

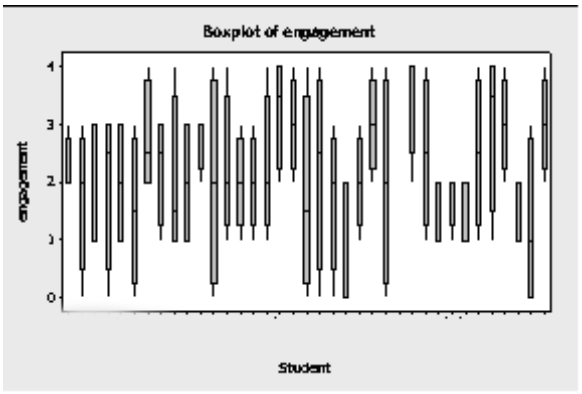
In each case items were recorded, on an ordinal scale, as one of unsatisfactory; OK; good; very good or excellent. However so many of the papers failed to score references and bibliography that this characteristic has been omitted from the analysis.

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 A marked no items as excellent and marked a total of 46 as unsatisfactory while marker D marked none as unsatisfactory and marked 20 as excellent (remember all markers were marking the same essays). See table below

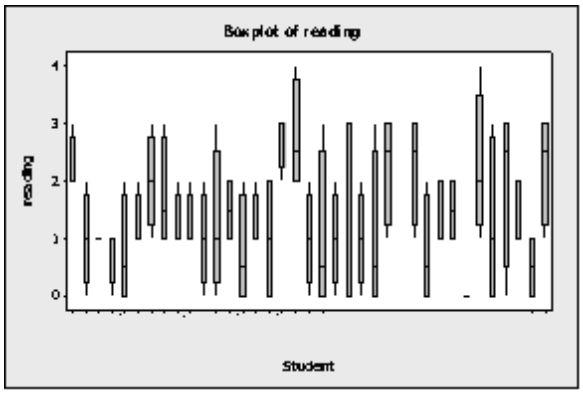
Marker	unsatisfactory	OK	good	Very good	excellent
A	46	80	47	2	0
B	18	38	65	41	18
C	15	66	45	40	8
D	0	4	78	61	20

To make analysis easier have transcribed data to a numerical scale where 0= unsatisfactory, 1= ok, 2=good, 3=very good, 4=excellent.

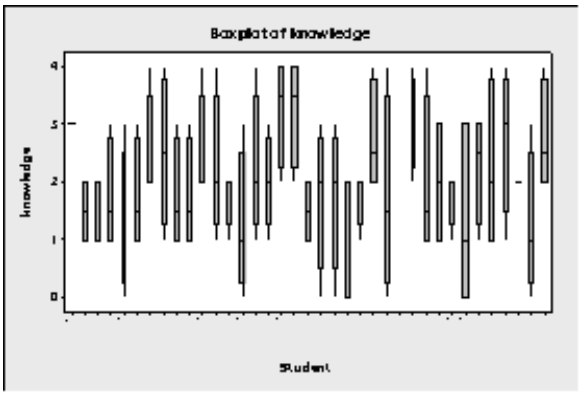
The diagrams below were used to explore the data graphically to investigate the variation between the markers. In these diagrams each vertical line represents the range of marks awarded by the four markers for the script of one individual students. The set of vertical lines in one graph represent the 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.



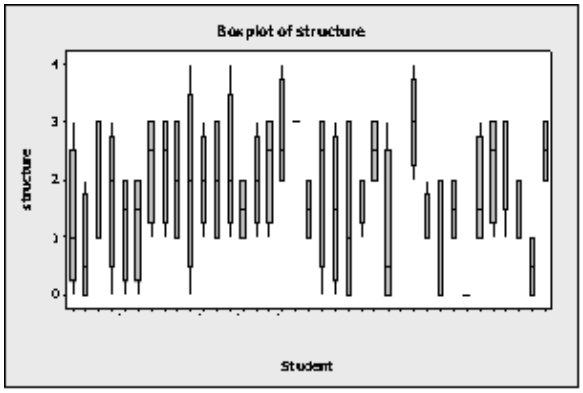
Engagement



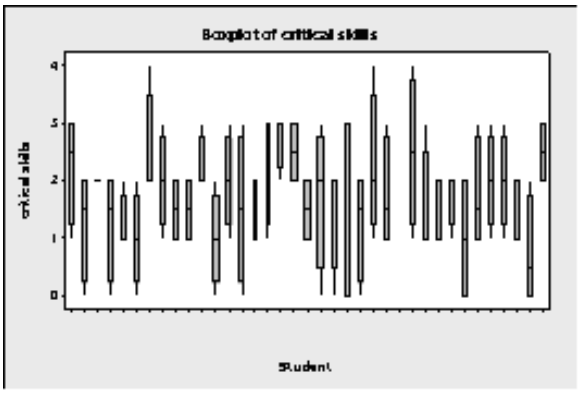
Wider Reading



Knowledge



Structure & Presentation

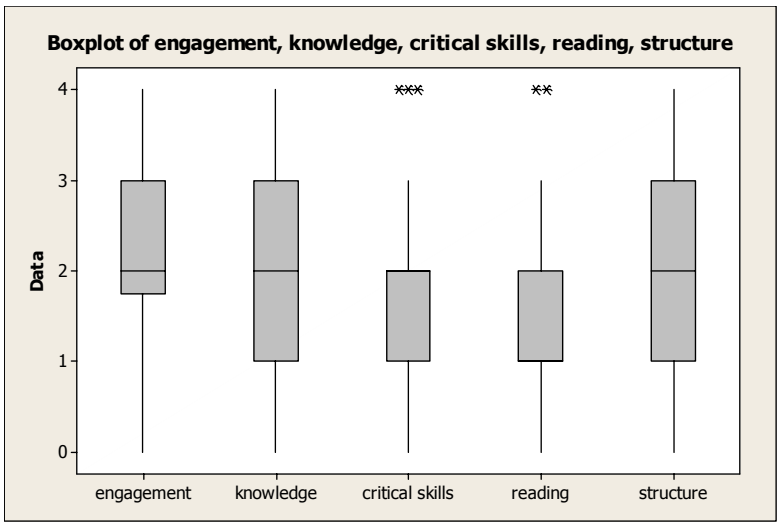


Critical Skills

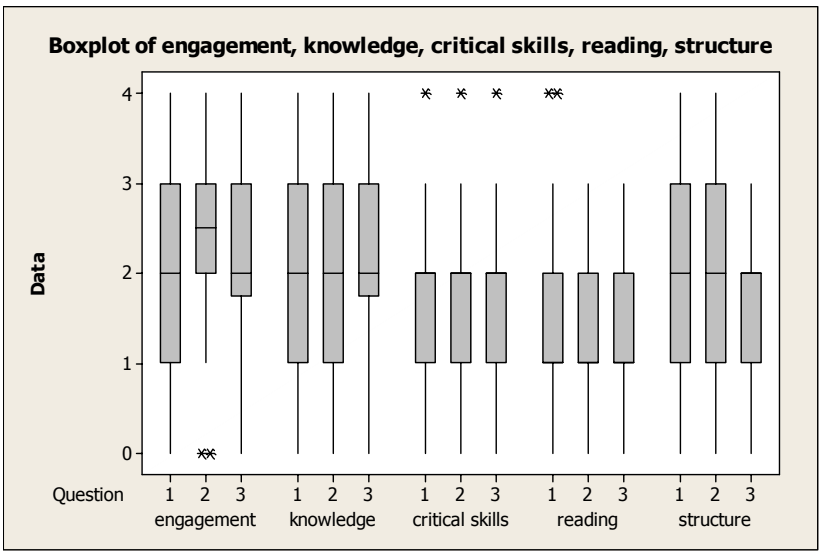
We can see that there is considerable variation – it is not simply a case of markers differing by a single category – there are many examples of total disagreement with one marker rating an item as excellent while another rates it as unsatisfactory.

Critical Skills and Wider reading arguably show slightly more consistency than the other items, but they are also the two items which 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.

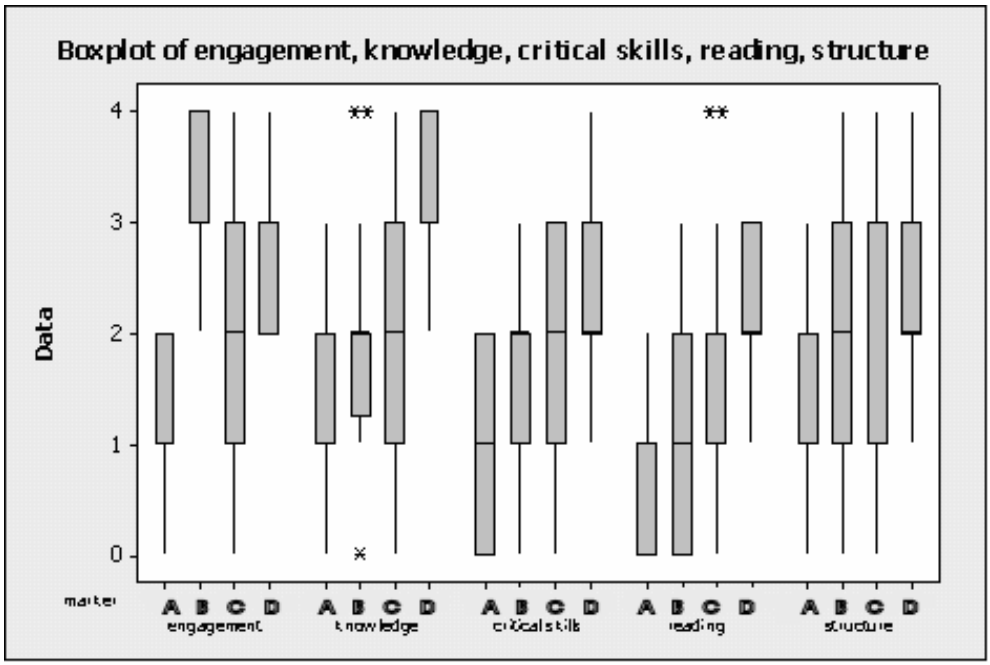
Overall the engagement characteristic was marginally the most highly graded factor.



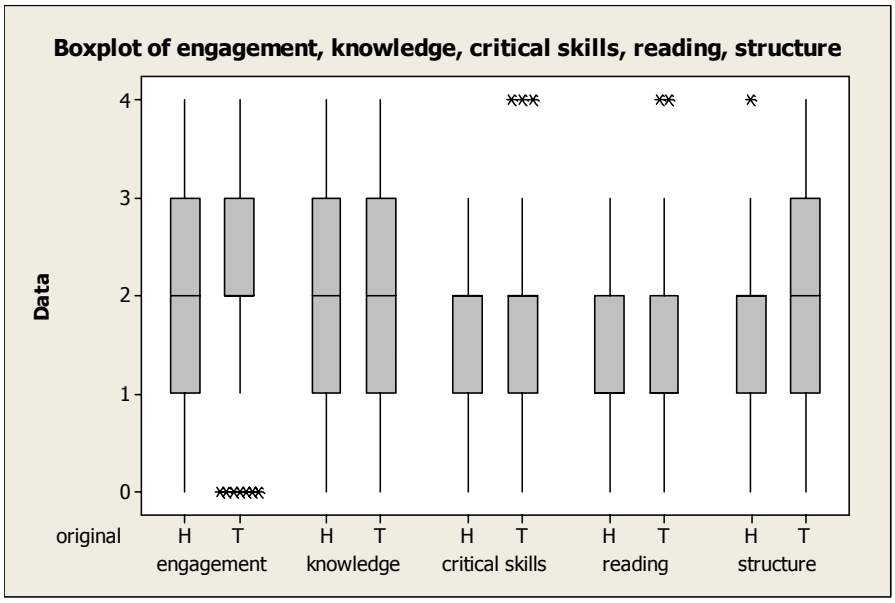
There is no evidence of a difference between questions



However there is clear evidence of differences between markers, as already seen in the opening table (remember all markers marked all the texts, although in different formats.)



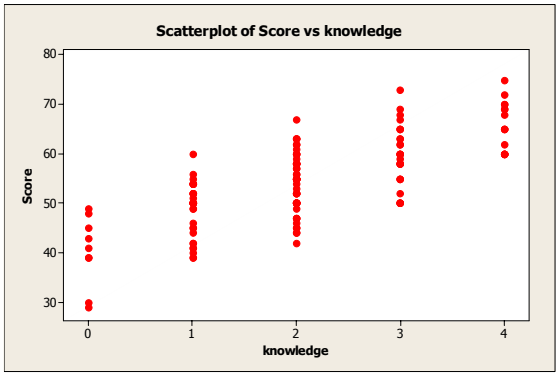
Is there evidence that hand written scripts are rated differently on these quality characteristics than typed scripts ?



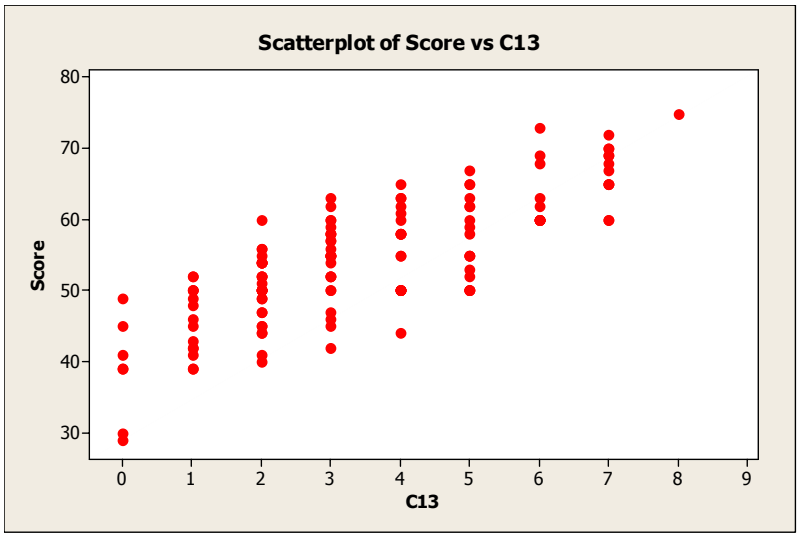
Clearly there are no huge differences indicated here, although there is some suggestion that essays which were typed showed better subject engagement and handwritten essays were less well structured/presented.

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 given below:

Engagement	0.47
Knowledge	0.743
Critical skills	0.719
Reading	0.735
Structure	0.589



Scattergraphs showing the rating for the knowledge factor (left) show a much stronger correlation with final score than the engagement factor (right). The effect can be strengthened by creating a new factor (knowledge + reading), bringing the correlation with final score up to 0.788



The students who typed the mock exam were asked if they thought using a computer to type their responses made a difference to WHAT they wrote. Their responses were:

No
Yes.
Not really, but it was better formatted
No, but I was able to go back and embellish what I wrote
I think my eloquence was improved
Doubt it. Maybe?

Yes, I think I was more to the point and clearer in my expressions
Yes
No
No, not really. If anything, I think the quality may have been more.
No
More logical - in order which may not have happened if handwritten

Pete Thomas from the OU has suggested that the conclusion is critical for giving impression of a well constructed essay. This study did not ask about the conclusion specifically but did measure (count) the length of the conclusion. There was no correlation between conclusion length and overall score (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. Typed conclusions were in general slightly longer than handwritten conclusions (mean no of words in conclusion 43 handwritten, 57 for typed).

Also no correlation between overall score and age. More variation in scores achieved by males than females.

3.5 Report from student Focus Group discussion 15 April 2008

The focus group was led by Gill Chetty.

A focus group discussion was set up to find out how the students found the experience of writing an essay type exam in this way. To help pinpoint what problems they encountered, what they perceived as advantages or disadvantages of writing an exam in this way.

All students who took part in the mock exam were invited to take part in the focus group discussion, four students responded to the invitation, two students actually took part. Both these students had done the exam on a computer, one using her own laptop and the other using a laptop provided by the School. These students will be referred to as Student A and Student B. Student A was a 19 year old first year undergraduate and student B was a mature student. Student A had experience of doing course work on a computer from school, Student B said she did not have a lot of experience of writing using a computer, apart from email and course work for her current degree.

Question 1: Think back to the mock exam. Describe your thoughts/feelings/anxieties prior to the exam.

The students were asked to think about how they felt prior to the exam, about their thoughts, feelings and anxieties.

Student A:

- Worried that she couldn't type fast enough.
- Concerned that the writing process would be different from how she normally wrote via a computer.
- Concerned that it would be more of a linear process as she wouldn't have time to restructure the essay.

  "I don't type very fast, I think I can probably type at a similar speed to writing, but maybe a bit slower."

  "But I think that is a slight issue that everyone types at different speeds. I was worried if I was going to be able to get enough down and I'm just not used to it. So it did make me quite nervous and I didn't know if I was going to be able to think in the same way while I was typing. I was worried that I wouldn't be able to think in a linear way because when I write essays, and stuff I change the sections round and I edit the essay as I go along. I wouldn't have time for that in the exam so I was worried that that would distract me too much."

Student B:

- Worried because she couldn't touch type
- Worried that she would be able to input enough in the time allocated.
- Worried that she couldn't type fast enough.

"I was really anxious about it because I can't touch type, I'm very slow, it's a two finger job and I was quite sure I wouldn't be able to get enough stuff down and apart from knowing what to say that was a real added anxiety. I just couldn't type fast enough I was quite keen to do it just to see how wrong or how right I was in that, I was quite interested."

Question 2: How did you go about writing the essay in the mock exam?

Student A:

- Wrote a plan on paper (was concerned that if she wrote it on the computer, she would forget to take it out in the end).
- Was worried that if she deleted a paragraph, she wouldn't be able to get it back later if she needed it.

"I think it was alright, I wrote a plan on paper actually, because that's how, because I can't.. because writing a plan on the computer, I'm worried that I'd forget to take it out. Not that it would be a big problem leaving it in but, it's nice to have the paper there, and then like to have something separate, rather than all on the screen. So I think if you are going to have typed exams, you have to have a piece of paper for people so that they can write."

Student B:

- Wrote down section headings on the computer and then went back to fill in the sections. She found this strategy worked well.

"I found it actually amazingly easy because I just put down everything I could think of in a row, just got headings down, then as I thought of it added bits to it. Where as normally I would have had pen and paper and jotted a few things down, it was on the top of my brain, and I actually found it very good because you could just kind of, do one bit and then leave it and come back to it later and work your way round, so I thought that was actually very good. But I didn't have time to leave anything or check it or anything but it was actually very useful.".

Question 3: What were the advantages and disadvantages of doing an essay exam on the computer

When asked about the advantages and disadvantages of doing an essay exam in the computer, they mentioned the following points:

Advantages:

- Easier to structure the essay.
- Didn't have to write postscripts at the end.
- *"If one could type fast enough it would be perfect!".*

Disadvantages:

- Starting an exam facing a blank screen was more daunting than facing a blank sheet of paper.
- The set up of the laptops to do the exam took a long time, this meant you had less time to write the exam and also increased your anxiety.
- No concept of how much you had written, with a hand written exam you aim to write about three sides of A4.
- A lot of noise from the keyboards, they thought that those who were hand writing would have been distracted by all the noise. (The two participants were not distracted by the noise of the keyboards).
- You couldn't type as much as you would have hand written.
- Can't touch type.
- Felt intimated at the sound of other people typing a lot faster than you could.
- Didn't think the programme was user friendly.

Question 4: What did you like most about the experience?

Student A liked being able to use her own laptop, as she was familiar with the keyboard.

Student B liked the fact of being able to start by writing our the main headings, and the ability to leave sections and go back to them later,

Question 5: General issues

When asked about general issues/concerns about doing an exam in this way, the following points were made:

- They both thought cheating might be a problem as it was easy to read other people's screens.
- Not a level playing field as some people can type a lot faster than others.
- Student A had problems with her wrists; she thought she wouldn't be able to type for a two-hour exam.
- Issues with downloading the programme onto your computer, Student A had downloaded the programme to her computer; Student B had opted to use a School laptop, as she was concerned it might "trash" her computer.
- When you write an exam using a word processor, spelling isn't assessed as they was a spell checker, they both thought it was important that spelling and grammar were part of the assessment.

- Wouldn't like to be staring at a computer screen for 2 hours, they thought there were health implications, that it would be bad for your eyesight.
- They thought that even if you are doing the exam on a computer, you also need access to pen and paper for writing notes etc.
- Water for drinking needs to be made available to candidates in an exam, they wondered there would then be a risk of spilling on the laptops.
- How to you reference, add footnotes?
- Setting up the computers for the exam needs to be done quickly and smoothly.
- Desks needed to be big enough to hold laptop, exam paper, pen and paper for writing notes.
- How would you provide enough power supply points for 100 students?

Question 6: How could this type of exam experience be improved?

When asked how this type of exam experience could be improved, the following were mentioned:

- Opportunity to play around with the programme before the exam, so that you were familiar with the screen and knew what features were available.
- Taught to touch type.
- Have the opportunity to do a lot of mocks before doing it for real.
- Programme includes a word count.

  "When I write, I work with, the amount of paper that I've used, whether I've written enough for a question, I know it doesn't work quite like that, but when it was on the screen because it wasn't separated into pages, or anything I had no concept of how much I had written, really. So either a word count or page breaks, or something, I think word count, would be quite useful".

Question 7: If you were now given the option of doing an exam, an essay type exam, using a computer or hand writing it, what would you opt for?

- Student A said she would choose to hand write as this is what she had experience of. On this occasion, she only chose to do it on a computer because it was a mock exam and she was interested in experiencing what it was like to do an exam in this way, but for a real exam she would prefer to hand write.

  "I would choose hand writing because I don't feel confident enough about typing one, even though I've tried it, for a mock, if not for an actual exam. I did it for the mock, because it didn't matter, and if I had

done really, really well on that mock, I would have felt slightly different. Obviously I didn't revise very much for it but I would have felt slightly more confident about it, but it's just too much of an alien idea, because I've used to .. I've written all my exams, so the idea of starting typing now, is just crazy. "

- Student B, said if she was given more time she would choose to use the computer.

  "*If I could have longer, I'd have the computer, but I'd need a lot more time".*

- They could see advantages of doing exams on a computer but would like the opportunity for a lot of mock exams so they had more experience of doing an exam in this way.

Question 8: anything else?

- They were unhappy that the mock exam had not been given a mark. They thoug t it would have been useful for formative assessment if the mock exams had been given a mark, not just comments. They judged the merit of an essay from the mark it had been given; they knew the significance of marks.
- Thought it would be a shame to do Greek exams in this way, as there was something very satisfying about writing Greek.

Summary

Their main concern was not being able to type quickly and accurately; they thought this would unfairly disadvantage them. They also thought it important to be given the opportunity to familiarise themselves with the programme's user interface prior to the exam. Although they both were familiar and comfortable with writing essays on a computer, they were apprehensive of writing an essay on a computer under exam conditions. They had developed strategies for hand writing essay based exams but not for word processing them.

3.6 Feedback from Markers

The markers were asked for any feedback and comments. Specific questions were posed as well as seeking general comments. Only two of the four markers supplied the requested feedback.

When you were marking were you aware of whether you were marking the original or a transcribed version of the answer? If so did it make a difference to you?

- It made little or no difference. I could tell the transcriptions were accurate because of the quality of the work and the spelling and grammatical errors!
- No difference but I could tell sometimes, as transcriptions in handwritten form had to reduplicate mistakes I associate with typos.

Did you feel you were influenced by whether the answer was typed or handwritten?

- In general the handwritten answers were very poorly written, often illegible. I much prefer marking typed answers.
- No

Did you feel you marked the different questions differently?

- I didn't think I marked the different questions differently. I always have the marking scheme at hand.
- No

How did you do about marking the scripts? For example did you mark all questions 1 first etc, etc

- I grouped the questions together.
- I marked them entirely randomly, in the exact order I received them

Did you think there was any difference in the quality between the typed and hand written answers?

- There was not much difference in quality between the handwritten and typed scripts.
- No

Comment

“I suspect students didn't do a lot of revision for these exams, given that they were mock ones - there wasn't evidence of a lot of secondary reading. The same problems of huge irrelevancies cropped up as usually do in weekly seminar sheets.”

4 Discussion

4.1 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 at first year level most courses will be looking for broadly similar skills, so this is not considered to be too problematic. It is also noted that the students volunteered to take the mock exam thus forming a self selected sample. It is unfortunate that because of the voluntary nature of the study some potentially useful data was 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 exam hall, they can never really be replicated in a mock exam. *" I did it* (used the laptop) *for the mock because it didn't matter, and if I had done really really well on that mock I would have felt slightly different. Obviously I didn't revise very much for it so I would feel more confident in a real exam, but its just too much of an alien idea because I've written all my exams."* Clearly revision, confidence and pressure must all have some impact on what and on how students write. However, exam boards are quite reasonably most reluctant to allow experiments in high stakes exam situations, hence 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.

Some may feel it is unreasonable to expect students to provide their own laptops. One argument is that the institution should provide the machines - thus gives a level of security that cannot be guaranteed with student owned machines, and another argument is that this makes it less likely that a machine will fail during an exam. Many of the issues raised can best be negated by demonstrating successful use cases. One student elected to borrow a machine because they were concerned the software would "trash" their machine – observing peers should help offer reassurance that this is not in practice a reported issue.

It is established that students have different understandings of what is expected in an essay in order to achieve a high score (Hounsell, 1997) 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 likely to be a different process to constructing an essay for an assignment, even if provided with identical tools.

Typing speed is frequently presented as a major concern and cited as a source of inequity. This study has 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 exam 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.

As a final general comment it may be useful to stress that it is not suggested that essay exams on computer should be the only assessment method used in any course, nor will it be appropriate for every discipline. Even if The School of Divinity does decides to take this forward as is expected, it may in the future be necessary to devise mechanisms for handling Hebrew, Greek or Arabic. This study is only intended as a starting point.

4.2 Looking at the fundamental questions

Clearly students type at different speeds and this does have an impact on how much they are able to write in an examination, but the variation between markers was much more significant than the variation between fast and slow writers and typists. If students are given sufficient warning of any requirement or opportunity to type their examinations they will not be unduly penalised. Poor typists can be supported to learn this skill if desired. There are many typing tutors readily available to suit all different learning preferences. 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.

Security and reliability issues have been raised generally 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.

The biggest concern highlighted by this study is the variation due to marker differences 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 & Tao 2004). Russell & Tao (2004 b) concluded that computer printed scripts would score on average 1.3 points less than the same words in a handwritten script. This study agrees that markers may indeed be influenced by format – and that difference might be worth almost 2 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 & Tao (2004 b) 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. Otherwise variation between markers is controlled by school and college quality assurance processes, there is no reason why marking of scripts which have been generated digitally should cause any alteration to establish procedures and guarantees of fairness.

This study has caused the project team to reflect more generally about assessment and assessment strategies. Pedagogically teaching, learning and assessment all consort. However we recognise that some students are mainly strategic learners focussed on passing exams, whereas others are seeking a deep understanding of their chosen subject. In principle intended learning outcomes and assessment methods should align, but in reality it is often the case that teaching happens and then exams are crafted. If teaching is geared too heavily towards the assessment then there is real risk that for many students the teaching is mearly a means to an end, the process of learning becomes secondary to passing the exams.

It may be that as an institution we need to think about and challenge our assessment strategies further. It is a relatively simple and complex matter to set exam questions. Or is it? Setting exam 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 exams 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?

4.3 School Marking Procedures and pressures on the markers

The level 8 course 'Christian Theology 1' was the basis of this pilot. This course has an assessment regime consistent with other level 8 courses within the school of Divinity, namely, 60% in the final exam, and 40% in-course assessment. In Christian theology 1, the in-course assessment is distributed equally between 1 essay (20%) and 10 tutorials (20%), the latter providing both summative and formative assessment. Students must pass both the in-course and exam elements of the assessment with at least 35% in each component.

In-course assessment is undertaken by tutors, but is moderated by the Course manager to pick up any severities in marking. Level 8 tutors in Divinity go through the normal University training scheme. The papers were marked as normal. Usually tutors mark first year essays and the tutorial element of assessment. While tutorial sheets (10 x 2%) are marked by tutors in tutorial groups, essays (anonymous) are distributed randomly among 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. The ECMS is brought to the attention of the tutors. If there is a disparity, marks may be lowered or raised by the course manager to ensure parity. This involves sampling by the course manager.

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. Tutors would also have been aware that this study was artificial and so it was less vital that their marking was absolutely accurate. Generally tutors have a limited life-span, which typically extends for about two years. Christian Theology 1 has used adopted a model of identifying one key tutor with prior experience of that course, and this has proved to be a helpful approach. Nevertheless it is observed that tutors tend to mark more severely than the course organizer would.

Formal exam assessment is undertaken by 3 examiners (3 questions, with one examiner for each).

If a student is in danger of failing, the course manager looks at border-line cases and reviews these, both in the elements of in-course assessment and final exam assessment prior to these being sent to the external examiner.

4.4 Venues and Practicalities – Some comments and some recommendations

Provision of a robust and reliable software solution is not sufficient to allow the routine holding of computer based essay examinations. Computer based exams have a number of requirements over and above standard hand written essay examinations including :

- Access to power for all machines without creating a hazard from trailing cables
- Access to network (either wireless or wired) if required
- Desks large enough to accommodate exam papers, scrap paper and a computer
- Controlling the opportunity to view neighbouring screens
- Minimising disturbance due to keyboard noise
- Scheduling and invigilating examinations appropriately
- Loan computers for students who may require them
- Spare hardware in the case of failure
- Emergency procedures in the case of power or other technical failure

It has proved difficult to find a room which can easily be converted into a computer-examination hall. Even disregarding the very important desire not to close computer labs for open access use, the design of most computer labs if far from ideal for examinations. Very few other rooms have sufficient power to be considered as a possible venue. Some raked lecture theatres do have power, but the desk width is small and would be hard to use for an examination context. In some cases it has been possible to run taped extension cables but this is really only feasible for say 20-30 exam seats and is clearly not a sustainable (or sensible) approach.

However it has been possible to secure funding to add a raised floor in one floor of Adam House, and this was fitted over Summer 2008. After some negotiation (with Exams office and Festivals Office principally) it was agreed to use the ground floor which has now been fitted with a floor which offers 63 powered examination seats. The hall can still be used for traditional examinations with minimal loss of seats. A ramp had to be included to enable any disabled users to access the space, and a separate electric circuit was used so that it can be entirely disconnected during the festival when the space is used as a bar. A plan of the new space is included in this report. There is still space at the back of the hall which can be used for bags and coats.

Exams office have been consulted about the changes, and have expressed concern about possible loss of traditional exam seats. It was recognised early on that computer exam desks will need to be slightly larger than the traditional exam desks used, and experiments also suggested that the spacing of computer exam desks would need to be slightly wider than for traditional exams (vertical screens are more easily read from a distance than a horizontal handwritten script). Therefore the layout of the hall will need to be different for computer exams than for traditional exams. In order to minimise work for servitors and other support staff it is suggested that

- Exams in this hall should be scheduled as far as possible in blocks with all the computer exams together and all the traditional exams together
- Expandable desks should be sourced which can be used in a small format for a traditional exams and a larger format for computer exams. When not in use exam desks are currently stored at the back of the exam venue, the logistics of having two separate sets of desks and moving them in and out of the exam hall as needed is not straightforward. Furniture Office are still working on identifying suitable desks.

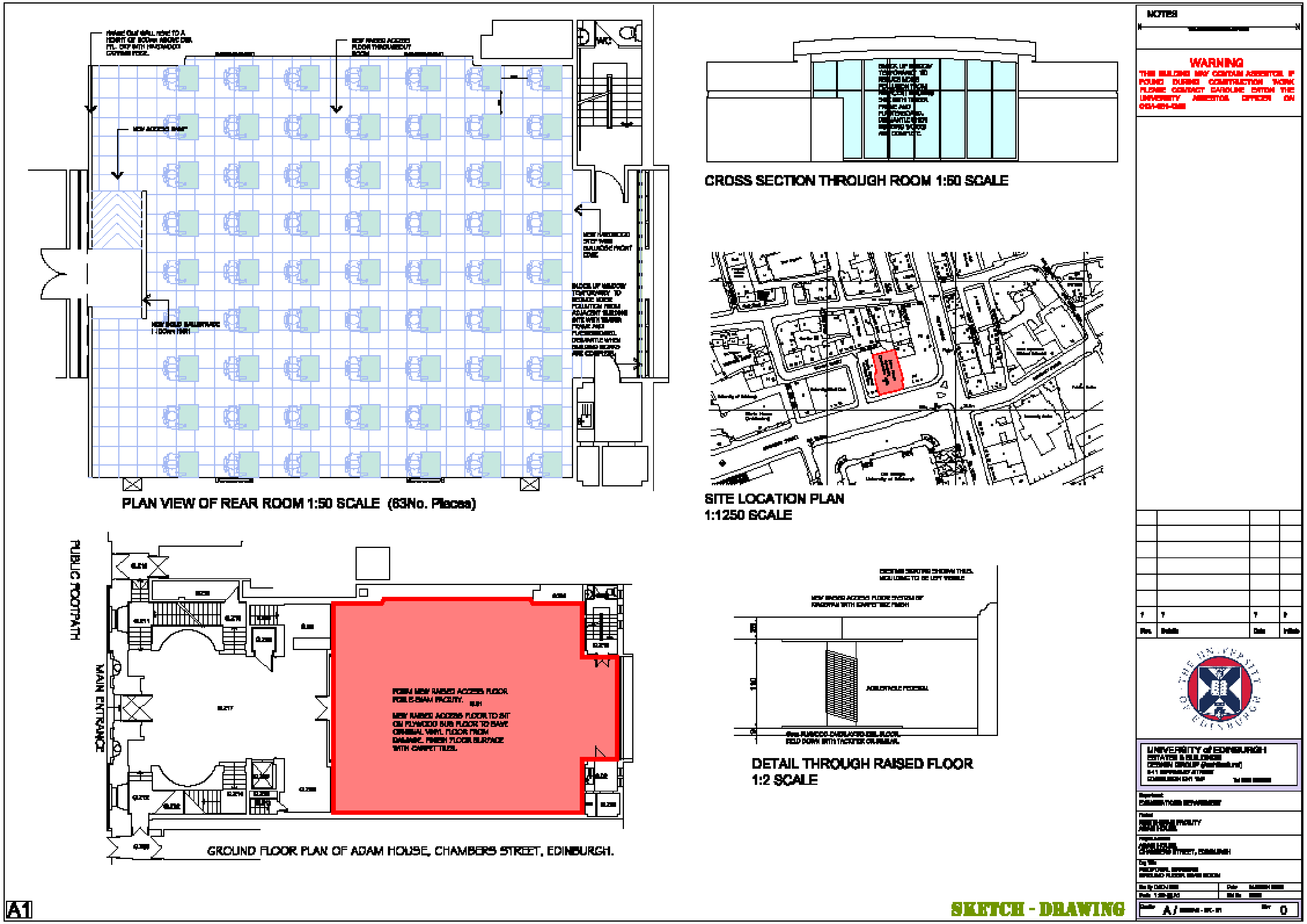
Holding a synchronous computer based examination for more than 63 students will remain problematic and at the moment it is suggested that solutions for each such event are decided individually.

Further liaison with exams office about the provision of training and information for invigilators still has to be arranged. Although the basic invigilation arrangements need not change it is important that all invigilators feel confident that they know what they are expected to do. Experience with computer based objective examinations suggests that most invigilators are apprehensive or even anxious about computer based examinations. They can be uncertain how to deal with student questions, and some have shown a tendency to panic and assume a technical problem when most commonly students have failed to understand or follow an instruction. It is recommended that training and support for invigilators is considered urgently, so that all invigilators used for computer based examinations understand their role, understand the nature of the examination and feel confident in their ability to apply appropriate procedures.

Emergency procedures must be established in line with assessment regulations and other exams.

With the relatively small numbers trialled to date noise has not proved to be a significant problem. Laptop keyboards are generally considerably quieter than full size keyboards. It is recommended to have a small supply of external keyboards in case of any last minute problems developing. Similarly it is recommended that a USB stick could be used to collect scripts in the case of any last minute network problems.

It is clearly important to allow students to practice using the software before an examination, preferably in exam like conditions. Students who own their own laptops can be allowed to download and use the software without restriction because the exam questions are provided on paper and are not bundled into the software. Students who do not own a laptop or who do not wish to use their own laptop must be permitted access to a loan machine in sufficient time to allow them to practice and become as comfortable as possible with it. The impact this might have on availability of loan machines will need to be carefully monitored.


NOTES
WARNING
CROSS SECTION THROUGH ROOM 1:50 SCALE
PLAN VIEW OF REAR ROOM 1:50 SCALE (63No. Places)
SITE LOCATION PLAN
1:1250 SCALE
PUBLIC FOOTPATH
MAIN ENTRANCE
GROUND FLOOR PLAN OF ADAM HOUSE, CHAMBERS STREET, EDINBURGH.
DETAIL THROUGH RAISED FLOOR
1:2 SCALE
UNIVERSITY of EDINBURGH
A1
SKETCH - DRAWING

4.5 Conclusions

Initially five questions were posed and the answers to these were explored in section 3

- **Is the mark awarded to an examination script influenced by the format of the script (typed or handwritten) rather than its content ?**

Yes but this is not a significant difference and is small compared to differences between markers.

- **Is there generally a difference in the amount of text that can be written or typed in an examination?**
- **Are students who type slowly any more or any less disadvantaged than students who handwrite slowly ?**

Yes, students are able to type more, but this was not strongly 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.

- **Do students perceive typing examinations to be as fair as handwriting responses to examinations?**

Students are keen to be offered the choice of handwriting or typing.

- **Do students report approaching the construction of an essay response differently when using a keyboard or handwriting ?**

The data here is not really sufficient to be conclusive.

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 has been considered essential to correct the mismatch between how students write coursework and how students write exams.

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 pre-warning 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 exams.

Boards of Studies 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 to 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.

5. RECOMMENDATIONS

For the school

Consider whether current marking processes are as robust and equitable as possible.

Identify a small number of pilot courses which agree to examine their students using essay examinations on computer or by offering the choice of handwriting or typing:

- It is hoped that the findings of the project will be brought before the Divinity Board of Studies (of which the manager of Christian Theology 1 is also convenor of Board of Studies). A proposal will be brought before Board of Studies, following discussion at subject area (Theology and Ethics) with a view to introducing pc exams in Christian theology in session 2009-2010. This gives enough time for details to be considered at CUGSC in the 2008-2009 academic session, and for details to be published as necessary for the 2009-2010 academic session.

Ensure students in any pilot courses are fully informed from the outset and provided with ample opportunities to practice, including access to loan laptops if necessary. Embedding familiarity with the software into formative assessment within courses seems important. This could be achieved via 'class exams' or take home exams using the exam software, rather than always submitting course essays using word.

Consider whether adjustments to the current assessment strategy would be appropriate. For example

- The School of Divinity at level 8 distributes final course marks 60:40 (Final exam : coursework), which makes it one of the schools in HSS with a standard pattern of assessment. Consideration may be given to varying this, perhaps to 50/50, to reduce any perceived risk in participating in the pilot.
- Occasional in-course assessment which contributes to an eventual mark could be part of the answer. History of Medicine and History of Science are wholly assessed by essays. Similar happens in Education.

For the university

Course teams appear nervous to commit to the use of computers for essay exams. A clear statement of strategic support from University officers may be a helpful catalyst for change. If there is a desire to move to essay exams on computer, it would also be helpful to indicate whether there is a strong strategic preference to offer students the *choice* of handwriting or typing.

Other courses in other schools should be approached with a view to examining their students using essay examinations on computer. Taught postgraduate courses may be an appropriate initial target group. Again the students should be informed of this from the outset and provided with ample opportunities to practice, including access to loan laptops if necessary.

If both handwritten and typed scripts are to be considered, markers should be given training to alert them to the likelihood of scoring variation due to format. Markers should be familiarised with how long a 2000 word essay looks in handwriting and in typed print.

A decision should be taken about software requirements and appropriate procurement processes initiated if necessary.

Invigilation arrangements must be clarified. Emergency procedures must be established in line with assessment regulations and other exams. Training and awareness sessions for invigilators should be arranged and any invigilators for essay exams on computer should be familiar with the procedures, and the differences between procedures for these exams and traditional essay exams should be highlighted for them.

It is recommended to have a small supply of external keyboards in case of any last minute problems developing. Similarly it is recommended that a USB stick could be used to collect scripts in the case of any last minute network problems.

Consideration should be given to the most appropriate and effective use of the space now created in Adam House. It may be desirable to use this facility for other computer based examinations such as online objective tests and open-internet exams. The purchase of a set of university owned laptops with an agreed configuration and software may also increase the ways in which the space can be used.

Any decision that all examination scripts are to be typed immediately leads to two further opportunities : it will be easy to provide students with a copy of their examination script (either printed or digital); and it may be feasible to move to marking on computer, which (with practice) can offer considerable benefits in terms of speed and consistency.

For further study

It would be of interest to consider any specific differences in the approach to essay construction for students in different years of their study, or in different discipline areas. Results from first year divinity courses may not be readily generalisable.

The study was artificial in that it considered data gathered in a mock exam. The process of transcribing the mock exam scripts has generated some interesting data; and it would be interesting to replicate the study with a real examination

This investigation has assumed that a secure environment which locks the computer down and which uses a simple word processor is preferable to an open context , where security is not seen as a significant problem and where students might use word or another word processor of their choice and might have open access to the internet or other digital resources. Such a situation clearly moves the assessment away from the traditional closed essay examination, but surely merits investigation as a feasible alternative style of assessment.

It would have been interesting to explore marker perceptions in more detail.

6. References

Augustine-Adams K, Hendrix B, & Rasband J. (2001) Pen or printer : can students afford to handwrite their exams ? *Journal of legal education* 51, 118-129

Burke J, Cizek G (2006) Effects of composition mode and self perceived computer skills on essay scores of sixth graders. *Assessing Writing* 11 148-166

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

Goldberg A, Russell M, & Cook A. (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

Horkay N, Bennett RE, Allen N Kaplan B & Yan F (2006) Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP *Journal of technology learning and assessment* 5(2)

Howell S (2003) e-Learning and paper testing: why the gap? Educause Quarterly Number 4 2003

Hounsell D. (1997) Contrasting conceptions of essay-writing, *The Experience of Learning*, Scottish Academic Press 2nd ed 106-125

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(2) 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**, 173-188.

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(1) 39-46

Mogey N, & Sarab G, (2006) Essay exams and tablet computer – trying to make the pill more palatable, *Proceedings for 10th CAA Conference 2006*

Powers D, Fowles M, Farnum M, Ramsay P (1992) 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, 220-233

Russell M & Haney W. (1997) Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via pencil and paper. *Education Policy Analysis Archives* URL http://epaa.asu.edu/epaa/v5n3.html (last accessed 10 May 2008)

Russell M & Plati T (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., & Tao W. (2004) Effects of handwriting and computer print on composition scores: a follow up to Powers, Fowles, Farnum & Ramsay. *Practical Assessment, Research and Evaluation* **9** (accessed 25 Feb 2008 at http://pareonline.net/getvn.asp?v=9&n=1)

Russell M., & Tao W. (2004 b)The influence of computer print on rater scores. *Practical Assessment, Research and Evaluation* **9(10)** (accessed 11 May 2008 at http://pareonline.net/getvn.asp?v=9&n=10)

Thomas P, Paine C & Price B (2003) Student experiences of remote computer based examinations. Proceedings from the 7th CAA conference Available at http://www.caaconference.co.uk/pastConferences/index.asp

Whithaus C, Harrison S, Midyette J (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, Bolton S, Feltovich B & Niday Dm (1996) The influence of student experience with word processors on the quality of essays written for a direct writing assessment. *Assessing writing* 3(2) 123-147

7. Project team

From The School of Divinity

Dr Michael Purcell
Dr John Burk
Dr Jessie Paterson

From Information Services

Nora Mogey

With support from

Dr Peter Wright, formerly School of Psychology
Thomas Graham, EUSA

And with thanks to

Greg Sarab from Extegrity Inc for access to the software
All the participating students
The four postgraduate markers from the School of Divinity
Frances Provan for statistical advice
Many friends and colleagues who transcribed scripts into handwriting
Susan Watson for administrative and typing assistance
Gill Chetty for hosting focus groups

And to TLA and the Principal's Teaching Award Scheme