Stylistic differences between typed and handwritten essay examinations

Nora Mogey, Jessie Paterson
The University of Edinburgh
Nora.Mogey@ed.ac.uk

Abstract

This paper stylistically analyses students’ answers for two mock examination questions, one of which was written using pen and paper, and the other using a basic word processor on student-owned laptops. The results provide some evidence that there are linguistic or stylistic differences between handwritten and typewritten examination essays. The data support other research in suggesting that students can and will type more words than they will handwrite. This study also presents evidence that the typed scripts tend to be more complex using significantly more words of three or more syllables and having a higher fog index.

Introduction

While there are many interesting examples of computer assisted assessment, the traditional examination essay remains a familiar assessment instrument in many UK Universities. Discussions about various ways technology can be exploited in an examination context include open web exams (Williams, 2004) and computer adaptive testing (Zenisky, 2010), in addition to objective test style examinations, already widely implemented. However there are only limited examples of allowing students to word process their essay exams, with the longest established examples being from law schools in USA, where this is already the norm.

Technically, allowing students to use laptops to author their essay exams can be quite straightforward. Security issues, which may be an initial cause for concern in some contexts, can be robustly managed, and do not need to be a barrier in practice (Mogey et al, 2007).

Recognising that exams are already stressful, allowing students the choice of typing or handwriting, based entirely on individual comfort and preference, has some attractions. Earlier research (Mogey, Purcell, Paterson & Burk, 2010) demonstrated that allowing students the choice was not unfair particularly when one considers the range of factors which often impact on their exam performance and score. Almost all of these factors are beyond the control of the student including, but not only, marking variability (Barrett, 1999).
However, students have demonstrated a reluctance to take up the option (Mogey, Cowan, Paterson & Purcell 2012). Although perhaps initially surprising, their caution is perfectly rational. Most university students have never been asked to sit timed essay-type examinations using anything other than a pen and paper. To write an examination essay using a keyboard is outwith their experience. Practice is clearly an important factor. The issue is not simply about comfort, familiarity and how to use the software, or how much you can type in a fixed time. It seems likely that there are also real differences in the planning, editing & revising behaviours associated with constructing the essay and perhaps stylistic differences in the final product which may in turn influence the marker when they read and grade the finished script. This paper focuses on exploring those stylistic differences.

Will a handwritten exam essay and a typewritten exam essay show linguistic or stylistic differences? If so what might cause those differences, and is either one likely to be regarded by the marker as superior?

In a fixed time period students will usually generate more words type-written than they will in the same time when handwriting and the amount written correlates with essay mark (Powers, 2005). However in a timed situation, such as an examination, Lovett et al (2010) demonstrated that length of text correlates with quality (scores) for typed text but not for handwritten text. They concluded that “the relationship between writing speed and writing quality is sensitive to whether compositions are typed or hand written”. Mogey, Cowan, Paterson & Purcell (2012) found that students composing examination essays using a word-processor do report making use of use editing tools such as cut & paste. Further Mogey & Hartley (2012) comparing university student mock-exam essays, identified some stylistic differences between hand-written and typed scripts, finding that typists tended to write more words and more sentences; but the sentences tend to be shorter and arranged in fewer paragraphs and overall in slightly more readable style.

Marker reactions to typed and handwritten scripts have also been recorded: hand written and type written versions of same script will not usually be marked equally - handwritten scores are slightly lower (Powers, Fowles, Farnum & Ramsay 1994, Mogey, Purcell, Paterson & Burk, 2008) although the strength of this effect can be moderated through additional marker training and awareness.

Giving students the choice to handwrite or to type essay-style examinations may be a reasonable thing to do, but how will students make the choice that is actually best for each individual? Until the potential implications of any differences between the two modes of composition are understood, any choice is essentially just a gut reaction, a lucky (or perhaps unlucky) guess. This study was designed to supplement our limited knowledge about the differences between composing exam essays using pen & paper compared with keyboard and word processor. In particular it sought to identify any stylistic or linguistic differences such as differences in sentence length or variation in vocabulary.

**Research Design**

An unseen practice examination was scheduled in class time for a first year, first semester, Divinity course. The exam consisted of two question prompts, and all the students were asked to respond to the prompts in the same order (A first and then B). About half the class, seated towards the back of the room, were directed to type their first response, and the other half to handwrite their first answer. Then halfway through the class, students were directed to switch to the second question, and to use the other mode of composition.
Handwritten answers were collected on paper; typewritten responses were collected on USB sticks and printed. Both questions were marked by a single experienced marker, and feedback provided to the students within two weeks. A total of 55 students attended the practice exam. 25 students typed their first answer, and 30 students handwrote their first answer.

Students were informally observed during the practice examination, making a note of approximately how many were actively writing at different times, and additional information was collected from students before and after the practice via two surveys and a focus group. This is written up in detail in Mogey, Cowan, Paterson & Purcell 2012.

In order to encourage students to take the practice seriously, and to try to simulate in some way the pressure of a high stakes examination, two £50 prizes were available one for the best handwritten, and one for the best type written script. One additional £50 prize was to be allocated randomly. In order to be eligible for the prizes students did have to complete both surveys.

The typed originals were also retained in a pdf format, and the handwritten scripts were carefully typed out giving a digital version for analysis. This gave 55 digital versions for each of question A and B. Basic descriptive information (such as word count, sentence length) was obtained for each digital script using the "usingenglish" website.

**Findings**

The use of the keyboard started and finished at different rates for the two exam questions. It was noticeable that students took some time to start typing for the first question, but started almost immediately for the second. For the first question students appeared to be writing/typing fairly steadily, whereas for the second question the number of students not actively writing/typing varied quite erratically.

**Will the essays show linguistic or stylistic differences?**

Over the two questions, students typically typed more words than they handwrote (Table1). The mean difference between their responses was 49 words (p<0.00) (Table 2).

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>min</th>
<th>max</th>
<th>Mean</th>
<th>Std dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typed</td>
<td>55</td>
<td>119</td>
<td>526</td>
<td>306.42</td>
<td>95.19</td>
</tr>
<tr>
<td>Handwritten</td>
<td>55</td>
<td>146</td>
<td>451</td>
<td>257.33</td>
<td>68.72</td>
</tr>
</tbody>
</table>

Table 1: number of words typed and handwritten - basic descriptive analysis

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
<th>Mean Difference</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type - hand</td>
<td>4.410</td>
<td>54</td>
<td>.000</td>
<td>49.09091</td>
<td>26.7734 – 71.4084</td>
</tr>
</tbody>
</table>

Table 2: One sample t-test for difference between no. of words typed & handwritten

All students attempted question A first and question B second. Fig 1 below plots the number of words handwritten vs the number of words typed for individual students, but shows the two groups of students separately. The graph suggests strong correlation (r=0.53) between number of words typed and the number of words handwritten – so students who handwrite more also tend to type more, independent of the question they were tackling.
Due to the arrangements in class, it happened that rather more students typed their response to QB. Considering responses to each question separately show that the amount written was very broadly consistent, except for the students who handwrote QB (recall that all the question Bs were answered second) who appeared to write less and who all appeared to write much the same amount (See fig 2).

Fig 2 : Boxplots to show number of words written or typed by question

A quick exploration of the data suggests that there is no big difference in the number of sentences written or typed – again the handwritten QBs tend to have fewer sentences, consistent with their being generally shorter. The broad pattern is the same for the two modes of composition and the two questions.
But considering average sentence length, there is some evidence that there may be differences between the typed and the handwritten scripts, the data for the typed responses looks different from the two handwritten responses (fig 4).

Lexical density and fog index are both readability measures. The Using English website, which was used to generate some of the stylistic data, rather than using the standard formula for lexical density, actually calculates the "type-token ratio"—the number different words used as a percentage of the entire text. Here a dense text might have a score of about 60-70%. The Gunning Fog index is a readability index, where high values indicate a more complex text. For reference, The Times and The Guardian have a fog index of about 11 or 12, and technical documentation has a Fog Index between 10 and 15. Table 3 shows the mean and the range for the "lexical density" and the fog index for the typed and handwritten scripts; both appear broadly comparable, and both suggest the student essays are fairly complex (mean fog index handwritten =12.5; typed=13.3).
For this study the interest is in whether there is a difference in style when students compose by hand or using a keyboard, so measures for typed and handwritten variables have been matched by individual and analysed using one sample t tests. Results are shown in Table 4.

The Quality of the Essays

Generally students scored about 5 marks lower on QB (the second question) than they did on QA (Table 5).

Comparing scripts by mode suggests that the score for typed scripts is slightly lower than for handwritten scripts (Table 6), but the mode effect is almost certainly confounded with a question effect.
There is a weak (but statistically significant) correlation between score and words \((r=0.288, \ p=0.002)\) suggesting that longer answers have generally received slightly higher scores. Looking at typed and handwritten scripts separately identifies that this correlation is stronger for handwritten scripts \((r=0.36)\) than for typed scripts \((r=0.28)\), both results significant at 1% and 5% levels respectively.

However looking at this data graphically identifies a clear outlier in the bottom right, and if this single observation is removed the correlation for typed scripts between score and length rises to 0.41 (sig at 1%). There is no evidence of a different pattern for the two writing modes.

There is a (statistically significant) relationship between the use of hardwords (more than 3 syllables) and score (Fig 7) \((r=0.256, \ p=0.007)\); this appears to be stronger for handwritten responses than for typed scripts.
Score does not appear to be closely related to either readability measure. Most of the essays have used relatively complex language regardless of the composition mode. However Fig 8 does perhaps illustrate a tendency towards an inverse relationship between lexical density and score, suggesting that answers written in easy language tended to achieve a lower score.

Fig 8: scatterplot showing relationship between score and lexical density (left) and score and fog index (right).
Conclusions

The results provide some evidence that there are linguistic or stylistic differences between handwritten and typewritten examination essays. The results support previous research which suggests that typically students can and will type more words than they will handwrite. However in contrast to earlier work (Mogey & Hartley 2012) (which also considered essays written in mock examination conditions in the same discipline but with rather fewer scripts available), different conclusions are reached about other aspects of the texts. In that study, typed sentences tended to be shorter than handwritten; fewer hardwords were used in the typed scripts and the fog indices of the typed scripts were lower than those of handwritten scripts. The earlier study thus suggested that typed scripts were overall more readable. This study finds the opposite, the typed scripts tend to be more complex. In this study, typed scripts used significantly more hardwords (words of three or more syllables). Typed scripts in this study had a higher fog index (suggesting the use of more complex language) but a slightly lower type token ratio, that is, a somewhat less diverse vocabulary.

Perhaps the issue here is how long a script appears to be – not how long it actually is – how long it “feels” to both the author and the marker. Both student and teacher have a broad understanding of how long an exam essay should be, and handwritten words flow and expand to appear longer. Norton & Hartley (1986) showed strong correlation between exam essay length & grade. Typed text is generally more compact than handwritten script – an A4 page of type written text will normally need between one and two A4 pages for a handwritten transcription. It is feasible that students are actively padding their text out so that it appears longer and more in line with their expectation of how long their answer should be. For both typed and handwritten scripts there was a positive correlation between the number of words written and the number of hardwords. However it is possible that the questions posed demand the use of specialist vocabulary including a higher proportion of long (hard) words than would be used in a normal discussion. It is also possible that students are using the same words repeatedly, rather than the high proportion of hardwords necessarily indicating a sophisticated and broad vocabulary.

Care should therefore be taken about extrapolating too far from this study. Theology is an academic discipline with its own language of discourse. The question posed required both subject knowledge and a command of specialist vocabulary. Although these students in some ways are in the very early stages of their studies of this discipline they will have already picked up some of the language of discourse that is appropriate. A layman would, most probably, have had difficulty constructing a plausible response to either question. The results indicated an inverse relation between lexical density and score suggesting that markers favour a level of sophistication or specialism in language and vocabulary and tend to penalise things written in to colloquial a voice. This result matches with the incidental observations of the transcriber.

For this class throughout the semester a deliberate focus of tutorials had been about helping students to appreciate the difference between different question prompts, so they answer the question posed, as opposed to writing down everything they can recall on the topic of the question. For this reason the two questions in the mock exam were essentially different takes on the same topic. Students all answered question A first and question B second, regardless of their writing mode. This order effect may in part be responsible for the observed differences in thinking, pausing and writing behaviours for the two questions, as well as the typical difference in score of about 5 marks. Another possible effect would be
that the marker graded all the question A answers, and then went on to grade all the question b answers, which could have contributed to the observed overall difference between the scores for the two questions.

Extrapolating from behaviour in a practice examination setting into a real high stakes exam can only be done with limited confidence. The nature of this practice was necessarily different from an end of semester exam: the room, the invigilation arrangements and not least the time available, all combine to make the context artificial. It was clear that some students took the practice very seriously, while others treated it with less gravitas.

The practice exam was held in a timetabled class 1 hour slot, so time was particularly short compared to regular examinations. By the time students had arrived and settled and been given some instructions valuable minutes had been lost. Further, on launch, the software conducts a security check which takes between 1-4 minutes depending on the processor speed and technical set up of each individual machine. In this time candidates are advised to read the questions, make their selections and start planning their response(s). In this case we directed all students to answer question A first and question B second regardless of composition mode, so they won’t have needed the same time to read and select questions as they would in most examination contexts.

This study has looked at stylistic differences between handwritten and typewritten examination scripts, and touched on possible implications for the marks awarded to the scripts. It suggests that there are some differences particularly in length of script generated and the complexity of vocabulary chosen. Why those differences might arise has not yet been explored. Data collected after exams means we know students do use the word processing features (Mogey, Cowan, Paterson & Purcell 2012). How and whether students plan exam essay answers, and whether they plan differently when writing or typing would be an interesting follow up study, and may directly impact on the stylistic qualities observed in scripts.

The motivation for this study was to help guide and support students when they are offered the choice to handwrite or to type essay-type examinations. While recognising that practice and familiarity with typing under timed conditions are important factors in informing student choice, it is argued that it is equally and perhaps more important to understand if there are real differences in the strategies and processes of composition leading to real differences in the quality of examination essays.

Acknowledgements

We are most grateful to colleagues and students in the School of Divinity, The University of Edinburgh, who have been co-researchers in this ongoing work, especially Mike Purcell. Additional advice and conversations with John Cowan from Edinburgh Napier University have been invaluable; Jim Hartley at University of Keele suggested the idea of a stylistic analysis. Support has also been received from our institutional ‘Principal’s teaching award scheme’ and from the Higher Education Academy Philosophy & Religious Studies Centre.
References


Mogey N, Hartley J, (2012) To write or to type? The effects of handwriting and word processing on the written style of examination essays, Accepted for publication Innovations in Education and Teaching International, details tbc


