Teodorescu, R. E., Bennhold, C., Feldman, G., & Medsker, L. (2013). New approach to analyzing physics problems: A Taxonomy of Introductory Physics Problems. *Physical Review Special Topics – Physics Education Research*, *9*(1). Available online at: doi:https://doi.org/10.1103/PhysRevSTPER.9.010103

Valcke, M., De Wever, B., Zhu, C., & Deed, C. (2009). Supporting active cognitive processing in collaborative groups: The potential of Bloom's taxonomy as a labeling tool. *The Internet and Higher Education*, *12*(3–4), 165–172. Available online at: doi:http://dx.doi.org/10.1016/j.iheduc.2009.08.003

van Hoeij, M. J. W., Hararhuis, J. C. M., Wierstra, R. F. A., & van Beukelen, P. (2004). Developing a Classification Tool Based on Bloom's Taxonomy to Assess the Cognitive Level of Short Essay Questions. *European Veterinary Education: Structuring Future Development*, *43*(3), 261–267. Available online at: doi:http://dx.doi.org/10.3138/jvme.31.3.261

Zheng, A. Y., Lawhorn, J. K., Lumley, T., & Freeman, S. (2008). Application of Bloom's Taxonomy Debunks the "MCAT Myth". *Science*, *319*(5862), 414–415. Available online at:doi:10.1126/science.1147852

How much do I need to write to get top marks?

Tom Benton Research Division

Introduction

'How much am I supposed to write?' must be one of the most frequent questions students ask themselves when faced with an essay task. I remember this question being asked by someone in the class nearly every time such a task was set for homework at school, and my own children invariably ask me the same question every time I am encouraging them to do their homework. Despite the ubiquity of the question, clear answers are hard to come by. Teachers at my school would reply (rather unhelpfully) "how long is a piece of string?" whilst my response to my own children is rather more determined by how much I know they will be able to write before they start seriously complaining of fatigue than by any strong educational evidence.

There are good reasons not to answer this question. First and foremost is the fact that the quality of a response is not determined by the quantity of writing. For example, no published mark scheme for GCSEs will specify the amount that candidates are supposed to write but rather will rightly point markers towards the skill the assessment is supposed to be measuring; for example, in the case of English Literature[1], the extent to which candidates have identified the key features of the text they are studying and are able to communicate effectively. With these points in mind it is understandable if teachers want to make sure the student's efforts are focussed on producing a high-quality answer to the question and not on meeting some arbitrary target in terms of how much to write.

However, whilst this article is in no way arguing against the overriding importance of high-quality content, it is reasonable for students to want some guide to how much is expected in terms of length. An older *BBC Bitesize guide to English Literature GCSE* suggested that for a 45-minute examination students might have a target of roughly 450 words[2] – whilst also providing some more specific advice around time management and practice in structuring an essay. This article will supplement this advice by showing the amount of writing produced on average by candidates awarded different grades.

The relationship between the length of responses and the marks awarded to them has long been established within the field of automatic essay scoring. To take one example, Murray and Orii (2012) describe their own attempts to build a statistical model to achieve accurate essay scoring as part of a machine-learning competition. As a baseline comparator to their own technique, they present the correlation between predictions from a model based on essay length alone (both word count and character count) and the marks awarded to students. Across 9 different essay tasks, these correlations were all strongly positive, ranging from 0.50 to 0.82. Indeed, the extent to which automatic essay scoring algorithms can rely upon essay length has been criticised in research literature. For example, Perelman (2014, p.104) stated that "Automated Essay Scoring engines grossly and consistently over-privilege essay length in computing student writing scores" showing that, for the essays in this same competition, estimated scores from seven commercial vendors of automatic essay scoring were far more strongly related to word counts than was the case when human marking was used. However, there is no existing research linking the length of handwritten responses in GCSE examinations to the grades achieved by students.

Other research within the UK has investigated the average speed at which students can write under typical exam conditions. Such research is important for the purpose of determining the physical speed of writing below which a student may require further support by means of special considerations such as extra time or the facility to submit a typed (rather than a handwritten) essay as part of their examination. A review of this research is provided in Waine (2001). She reviewed 2 small-scale studies showing that in a free-writing task, where students had to decide what to write rather than simply copy it, students wrote on average between 14 and 18 words per minute. She also conducted her own study where, under examination conditions, 152 Year 10 (age 15) students were asked to write on the subject of 'My Life History' for a period of 30 minutes. Her results indicated that the mean writing speed of Year 10 students was 15 words per minute and that speeds between 10 and 20 words per minute were within the typical range. Similar research published by Patoss[3] (the professional association of teachers of students with specific learning difficulties) shows that, in a 20-minute free writing task, Year 10 students write at an average of 16 words per minute which rises to 17 words per minute for Year 11 students. Other research shows that when 16-year-old students are simply copying text they can write considerably even faster; at over 20 words per minute on average whilst writing neatly for 2 minutes, and at over 30 words per minute when writing as fast as possible (Barnett, Henderson, Scheib, & Schulz, 2009).

Overall, therefore, previous research has shown that the length of

1. See for example http://www.ocr.org.uk/Images/236719-mark-scheme-unit-a662-02-modern-drama-higher-tier-june.pdf (Retrieved 28 June 2017).
2. http://www.bbc.co.uk/schools/gcsebitesize/english_literature/prosejaneeyre/4prose_janeeyre_sprev1.shtml (Retrieved 28 June 2017).
3. https://www.patoss-dyslexia.org/SupportAdvice/InformationSheets/2012-09-02/Handwriting-Assessment/. (Retrieved 28 June 2017).

responses does have some association with achievement and also provided some norms around the possible writing speed of GCSE-taking-age children. However, none of these studies relate to performance in a real examination task. Thus, they do not provide any clue about how much writing is usually associated with achieving a high grade in a GCSE examination. The aim of this article is to fill this gap.

Basic method

As noted by Waine (2001), one of the main challenges with this type of research is the laborious task of manually counting the number of words written. To overcome this, building on work described in Benton (2017), the research for this article used computer processing of digital images of handwritten scripts to provide an estimate of how many words had been written. The basic process employed to count the number of words written on each page was as follows:

1. Use the dotted lines on the answer sheet to split the writing on the page into lines that can be processed separately (all essays included in this analysis were written on lined paper).
2. Remove any small objects (such as dots) from the image of each line. If, after this, there is no evidence of any ink remaining on the line then assume a word count of zero.
3. Within each line identify all clear horizontal gaps (i.e., horizontal spaces where there is no ink anywhere between the top and the bottom of the line being written on) and record the widths of these gaps.
4. Use cluster analysis to split these gaps between those that are likely to represent a break between words and those that are probably gaps between letters within the same word. In doing this it is assumed that any gaps wider than 5mm[4] must always represent a gap between words and that any gaps of less than 1mm must relate to a gap between letters within the same word.
5. The number of words on each line is now estimated as the number of between-word gaps on the line plus 1.
6. Add up the word counts across all lines on all pages within a candidate's examination booklet to produce a final estimated word count.

Further details on the processes involved in analysing images from examination scripts can be found in Benton (2017). The above approach was applied to a sample of 5,000 scripts from a 45-minute GCSE English Literature examination and the resulting word counts were linked with grades on the exam. However, before looking at the results of this analysis, it is first necessary to validate the word counting method itself.

Validation on a small scale example

In order to validate the word counting method above, the above process was applied to a sample of student responses to a short answer question from a GCSE Biology exam. The question itself asked "A supermarket is considering how they can make their shopping bags more sustainable. What is meant by sustainability?" and the answer space for students allowed them to write up to three lines of text in response. A random sample of 100 responses to this question was selected from amongst all

4. Actually 25 pixels within the resolution of images used for this analysis.

Word counts on question 'What is meant by sustainability?'

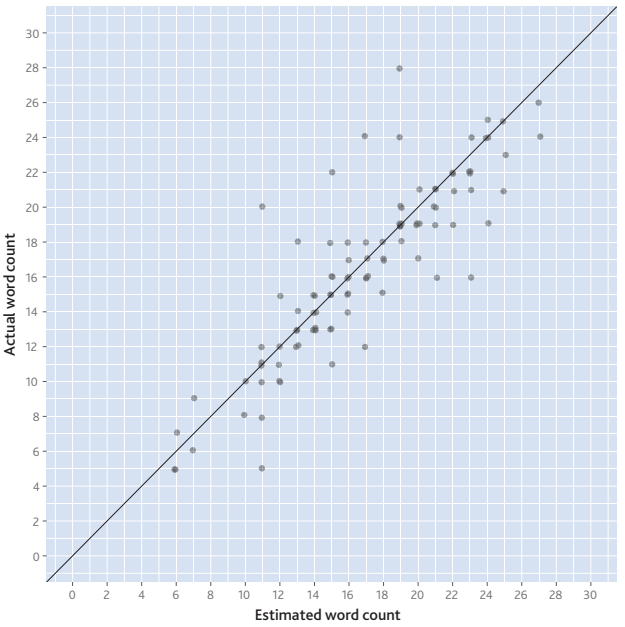

Figure 1: A comparison of estimated and actual word counts for the short Biology question

students taking the examination and a manual word count of each student's response was recorded. Then, the automatic process described above was applied to the same set of images and the estimated word counts were compared to the actual ones.

The results of the analysis are shown in Figure 1. As can be seen, the automated word counting mechanism was far from perfect but, nonetheless, did give a reasonable idea of the length of responses. The overall correlation between automated word counts and the actual length of responses was 0.87. Also, importantly, there was no evidence that the automated method was biased towards either over-counting or under-counting the true number of words. The actual mean number of words spent answering this question was 16.5 compared to a mean of 16.8 for the automatically estimated number. Similarly, the median of both the actual and the estimated word counts was 17. Further scrutiny of individual cases revealed that the automated word counts might be too low if candidates crossed out work and then rewrote sections of their answer over the top – thus obscuring any clear gaps between words. On the other hand, if a candidate's writing was too close to the line they were writing on (and perhaps dipped underneath this line) the algorithm may fail to include all of their writing within the image being analysed. This could lead to large horizontal gaps within words and, thus, the total number of words being over-estimated.

Notwithstanding these weaknesses, the analysis indicates that the automated method of generating word counts provides a reasonable basis for calculating how much candidates are writing in longer essays.

Word counts and grades for an English Literature examination essay

The analysis in this section examines GCSE English Literature essay responses from June 2014. In this particular examination, candidates were required to supply just a single essay response and were allowed a total of 45 minutes to complete their work. A random sample of 5,000 essays was selected for further analysis. The answer booklet was

restricted to six pages and a small minority of candidates where the number of archived scanned pages associated with their response differed from this was excluded from the analysis[5].

To begin with, for further validation, the automatic word counting process was applied to three pages from three different candidates taking this test. The average number of words per page from the automated process was found to match the actual average number of words showing that the process was generally suitable to be applied to full page responses.

Next, the automated word count process was applied to all essays. A total of 14 essays were removed from the analysis because the estimated word count was zero (this might be because the candidate's response was typed so was not within the standard answer booklet). A further two responses where the estimated word count exceeded 1,500 (which would imply the candidate wrote more than 30-words per minute throughout the entire exam) were also excluded from the analysis. The association between the estimated number of words written by each candidate and the grade they were awarded on this particular examination component is shown as a boxplot in Figure 2. The boxes in this plot indicate the inter-quartile range for the estimated word count within each grade with the central line denoting the median. The extra lines and dots show the full range of estimated word counts with the dots indicating outliers. Some summary statistics from this plot are provided in Table 1.

Table 1: Summary statistics for the relationship between estimated word counts and English Literature grades

Grade	*Number of candidates*	*Median estimated word count*	*Mean estimated word count*
A*	605	694	705.8
A	1008	637	652.0
B	1565	582	597.3
C	1009	517	538.8
D	493	492	500.7
E	142	450	460.4
U	162	370	383.5

Figure 2 shows a clear relationship between how much candidates wrote and the grades they were awarded. The correlation between estimated word counts and the marks awarded on the test (out of 49) was 0.46. The median number of words written by a grade A* candidate was 694 implying that they wrote around 15-words per minute in the exam, though, of course, they may not have used the entire time available in the exam for writing. Inspection of a few A*-graded essays of this length indicated that this relates to around five pages of writing. In contrast, the median number of words in a grade E essay was only 450 indicating 10 words were written per minute of the exam. In interpreting these numbers, it is important to remember that some of these candidates may have given up writing before the end of the available time.

Figure 3 shows the relationship the other way around, displaying the association between estimated word count and the number of marks awarded. The blue line shows how the mean number of marks awarded varied with the amount of writing. The dotted lines indicate the grade boundaries on the exam. Crucially, this shows that whilst candidates

5. A total of 27,351 candidate scripts for the exam were available within our digital script archive. The number of scanned pages in the archive matched the length of the answer booklet provided for 26,338 of these.

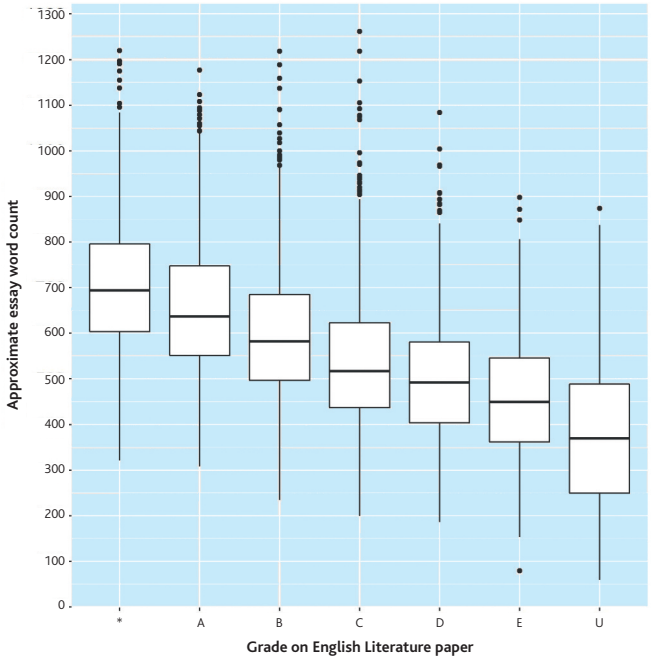

Figure 2: The relationship between word counts and achievement on the English Literature GCSE paper (The * represents the A* grade)

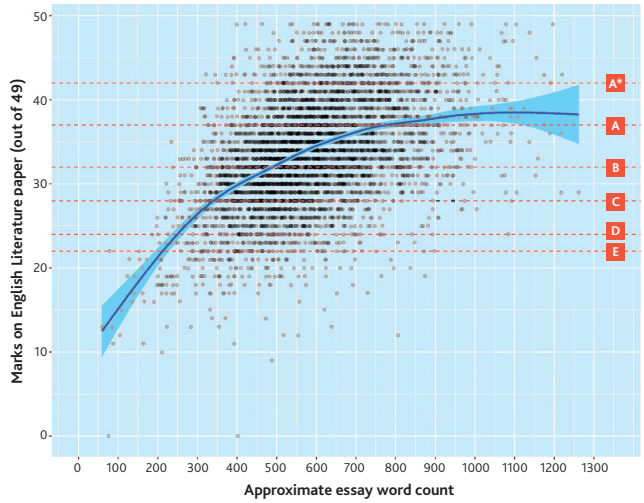

Figure 3: The relationship between estimated word count and mean mark

achieving the highest grades did tend to write more, a longer answer was by no means a guaranteed path to high marks. As can be seen, even for the longest essays, the average mark achieved by candidates never reached the top (grade A*) boundary. Indeed, the relationship between essay length and the mark awarded flattens off beyond 700 words indicating that there was no benefit in candidates writing extremely long responses.

At the other end of the spectrum the relationship is clearer. Nearly all responses of fewer than 200 words resulted in a grade U, suggesting that whilst very long answers are not necessary for a good mark, candidates must write enough to make sure that the examiner can recognise their knowledge at all. With this in mind it would be good advice for all candidates, even those who are not expecting to achieve the highest grades, to ensure that they produce at least a full page of writing in response to an English Literature exam question allowing 45 minutes to write an essay. It might also be noted that no candidate in the sample was awarded a grade better than a grade B without writing at least 300 words.

To complete the analysis, a brief manual inspection of some of the outliers in Figure 3 was conducted. Specifically:

- An inspection of the script awarded a grade E but where the estimated word count was below 100 revealed that a single candidate really did achieve a grade E with around half a page of writing.
- Inspecting the scripts for the two candidates apparently writing more than 1,200 words but only awarded a grade C revealed that both of them submitted 6 complete pages of writing. This reinforces the point that very long answers do not guarantee that a candidate will be awarded the highest marks.
- Inspecting the scripts for the 9 candidates awarded a grade A* but where the estimated word count was below 400 showed that in 8 out of 9 cases the candidates wrote less than 2.5 pages and in some cases less than 2 pages. For the other case (actually the grade A* candidate with the lowest estimated word count), the candidate had an unusually slanted writing style that probably obscured the gaps between words. Nonetheless, the other eight cases clearly show that it is possible to achieve the highest grades with fairly short answers.

Conclusion

This article has provided some fairly detailed information on the link between the amount candidates wrote for an English Literature essay and the marks they were awarded. As might be expected, there was a clear link, particularly at the lower end of achievement. This is no surprise as it is clearly impossible for candidates to be awarded the highest grades unless they provide enough material to demonstrate their skills to the examiner. With this in mind, if candidates are asked to spend 45 minutes answering an exam question they should aim to provide at least a page of writing in response and at least two pages (or thereabouts) if they want to have a chance of achieving any of the higher grades.

However, it is also very clear that the length of the response alone is insufficient to achieve a high mark. Beyond a certain essay length, the relationship between writing more words and achieving more marks flattened off. Thus, there is no evidence that writing extremely long answers makes a substantial difference to grade outcome, showing that quantity certainly does not trump quality. To reinforce this, we can note that inspection of individual essays revealed instances where, with well organised responses, students achieved all of the marks available on the exam with relatively short answers.

References

Barnett, A. L., Henderson, S. E., Scheib, B., & Schulz, J. (2009). Development and standardization of a new handwriting speed test: The Detailed Assessment of Speed of Handwriting. In *BJEP Monograph Series II, Number 6-Teaching and Learning*, 137–157. British Psychological Society. Available online at: https://doi.org/10.1348/000709909X421937

Benton. T. (2017). The clue in the dot of the "i": Experiments in quick methods for verifying identity via handwriting. *Research Matters: A Cambridge Assessment publication*, *23*, 10–16. Available online at: http://www.cambridgeassessment.org.uk/Images/375445-the-clue-in-the-dot-of-the-i-experiments-in-quick-methods-for-verifying-identity-via-handwriting.pdf

Murray, K.W., & Orii, N. (2012). *Automatic Essay Scoring*. Available online at: http://www.cs.cmu.edu/afs/cs.cmu.edu/Web/People/norii/pub/aes.pdf

Perelman, L. (2014). When the "state of the art" is counting words, *Assessing Writing*, *21*, 104–111. Available online at: https://doi.org/10.1016/j.asw.2014.05.001

Waine, L. (2001). Writing speed: What constitutes 'slow'? An investigation to determine the average writing speed of year 10 pupils. In Rose, R. & Grosvenor, I. (Eds.), *Doing research in special education: Ideas into practice*, 75–87. London: David Fulton Publishers.