

MATHEMATICS TEXTBOOK USE IN ENGLAND: MINING OFSTED REPORTS FOR VIEWS ON TEXTBOOKS

Christian Bokhove

Keith Jones

University of Southampton

University of Southampton

c.bokhove@soton.ac.uk

d.k.jones@soton.ac.uk

According to TIMSS data, the use of textbook in mathematics classrooms in England is relatively low in comparison to other countries. Although the reasons for this might be varied, the pronouncements of Ofsted, the official body for inspecting schools in England, might have an influence. This paper reports on a text analysis of almost 10,000 publicly-available Ofsted secondary school inspection reports and mathematics- specific commentaries from the year 2000 until now. The analysis focused on what Ofsted said over this period about textbook use in general and about the use of mathematics textbooks in particular. The analysis was conducted by first ‘scraping’ the reports from the Ofsted website and then utilising basic text mining and analysis methods to extract information on these documents. While the analyses showed that the occurrence of comments by Ofsted on textbooks appeared to be relatively minor, interpreting these findings from text mining alone was not straightforward. A further qualitative analysis of a sample of Ofsted publications found mention of ‘over-reliance’ on textbooks. Such allusion to ‘over-reliance’ on textbooks might have negative connotations and may have contributed to the relatively low use of textbook in mathematics classrooms in England.

Keywords: TIMSS, textbook use, Ofsted, England, text mining, text analysis

INTRODUCTION

Textbook use in mathematics classrooms in England is, according to TIMSS data, “lower than that in the highest-attaining countries [in TIMSS]” (see Askew et al, 2010, p.34). Askew et al, go on to say that given the paucity of relevant research, “the reasons for these differences in textbook use are not clear”, although the lower proportion of teachers in England reporting use of textbooks may reflect “a greater use of internet resources”, “initiatives by the National Strategies promoting the use of alternative resources alongside published textbooks”, and “a view amongst the educational establishment that schools over-rely on textbooks rather than undertaking their own detailed planning [of lessons]”.

Askew et al do not elaborate on who might comprise the “educational establishment” that might hold the view that “schools over-rely on textbooks”, so this paper reports on research that set out to establish what Ofsted, the official body for inspecting schools in England, might have said about textbook use of the period from the year 2000 until now. The analysis was conducted by first ‘scraping’ almost 10,000 publicly-available Ofsted secondary school inspection reports and mathematics- specific commentaries reports from the Ofsted website and then utilising basic text mining and analysis methods to extract information on these documents. As interpreting the results of the text mining was not straightforward, a further

small-scale qualitative analysis was conducted as a way of illuminating the text mining analysis.

OFSTED, THE OFFICE FOR STANDARDS IN EDUCATION

Ofsted, the Office for Standards in Education, Children's Services and Skills to give its current full title, says that it "regulates and inspects to achieve excellence in the care of children and young people, and in education and skills for learners of all ages in England" (Ofsted, 2014, preface). It does this by conducting inspections of schools, colleges, and providers of further education and skills (including providers of initial teacher training).

The process of inspection has been subject to some critical scrutiny, mostly focused on issues of validity, reliability, and consistency. For example, Campbell and Husbands (2000) found the methodology of inspection to be insufficiently reliable for the consequences which flow from it, a finding echoed by Jones and Sinkinson (2000) and Sinkinson and Jones (2001). Inconsistencies affecting inspections were reported by Penn (2002), while more recently Baxter & Clarke (2013) found considerable tensions in the inspection system.

While the role played by Ofsted in promoting school improvement is currently being examined (e.g. Jones & Tymms, 2014), what remains as yet relatively unexamined is the influence of Ofsted on the professional practice of teachers in England. Here it is necessary to note that Ofsted has recently had to state publicly that it "does not favour any particular teaching style and inspectors must not give the impression that it does" (Ofsted, 2015, para 178). Ofsted goes on to state that, for example, inspectors "should not criticise teacher talk for being overlong or bemoan a lack of opportunity for different activities in lessons unless there is unequivocal evidence that this is slowing learning over time" and that "when observing teaching, inspectors should be 'looking at' and reflecting on the effectiveness of what is being done to promote learning, not 'looking for' specific or particular things" (ibid).

METHODOLOGY

The procedure that was used for data mining was loosely based on the 'knowledge discovery in data' methodology using The Cross Industry standard Process for Data Mining (CRISP-DM, Bosnjak, Grljevic & Bosnjak, 2009). CRISP-DM distinguishes several phases. The first phase, Organizational Understanding, concerns an understanding of what data is actually on the web, what does it say, and how could it be useful for us. The second phase, Data Understanding, involves knowing the precise format of the data. In phase three, Data Preparation, the data is transformed into a format that is understandable for the tool that is to be used to perform the analyses. Phase four, Modelling, is the phase that is used for the actual analyses. Phase five, Evaluation, determines the truthfulness and usefulness of the analysis results. Finally, phase six, Deployment, involves the distribution and publication of the results of the analyses, as is done by this paper, and is therefore not given further consideration.

Organizational understanding

As noted above, the first phase of the analysis, Organizational Understanding, concerns an understanding of what data that is on the web, what it says, and how it could be useful for the

analysis. Here we can note that the Ofsted website provides publicly-available inspection reports for every school. Every report has a judgment attached to it which is mentioned on the website and within the report itself. The current judgments are: grade 1 (outstanding), grade 2 (good), grade 3 (requires improvement) and grade 4 (inadequate). Before January 2012 grade 3 (requiring improvement) was called ‘satisfactory’. The website also contains interim reports.

Data collection and data understanding

For the second phase, Data Understanding, a ‘scraper’ was set up using Scrapy (<http://scrapy.org/>) and this was used to ‘scrape’ the Ofsted website. The scraper collected the URLs of all historical inspection reports and interim reports since the year of first publication, 2000 (N=9559, 1.39 GB of data). This set of reports was then downloaded using a mass downloader. All documents were in PDF format. Table 1 provides an overview of all the Ofsted documents that were collected, and includes additional mined data, namely the average days between inspection date and report, as well as medians, minimum and maximum.

Table 1: overview of downloaded Ofsted documents

Year	Reports	Size *)	Avg days between inspection date and report	Med	Min	Max
2000	212	34.2 MB	171	158	91	1257
2001	278	38.7 MB	101	99	65	347
2002	178	30.5 MB	114	100	79	822
2003	190	33.7 MB	118	99	72	1409
2004	274	35.4 MB	137	94	51	1192
2005	302	51.1 MB	120	73	13	1178
2006	639	58.2 MB	106	26	10	2566
2007	884	122 MB	61	27	7	1975
2008	835	132 MB	42	29	6	1042
2009	896	128 MB	43	30	2	439
2010	1062	150 MB	36	26	10	286
2011	1139	193 MB	39	26	8	800
2012	1000	175 MB	29	22	4	212
2013	1481	239 MB	28	22	9	974
2014	189 **)	7.17 MB	35	31	-1 ***)	120
TOTAL	9559	1.39 GB				

Data preparation

For phase three, Data Preparation, the PDF files were first converted to plain text format. Two datasets were prepared from these ‘raw’ materials. One dataset consisted of three corpora of Ofsted files: 1) all files from 2000 to 2004, 2) all files from 2005 to 2009, and 3) all files from

2010 to March 2014. These files were all added as corpora to a software programme called TextSTAT (version 2.9c), “a simple programme for the analysis of texts. It reads plain text files (in different encodings) and HTML files (directly from the internet) and it produces word frequency lists and concordances from these files” (to quote from the TextSTAT website). Two mathematics-specific reports (Ofsted, 2008; 2012) were added separately making a total of five documents. A second dataset consisted of three year files for the years 2004, 2008 and 2013 respectively. These were imported into Rstudio web, a web-based frontend for the statistical package R. The tm (text mining) package (Feinerer & Hornik, 2014; Meyer, Hornik & Feinerer, 2008) was used to make a ‘corpus’ for every data set. The three corpora were then subjected to several transformations:

- Making all characters lower case
- Removing punctuation marks from a text document
- Removing any numbers from a text document
- Removing English stop-words
- Stripping extra whitespace from the documents

Finally, for this phase, what are called Document-Term matrices were constructed for every document, allowing us to perform several analyses of the documents. In the analysis reported in this paper we looked at the top 10 words and the relative frequency of words between the year-pairs 2004 and 2008, 2008 and 2013, 2004 and 2013.

Modelling

Both data sets were analysed using the respective software packages. The first set was explored quantitatively on frequencies of the following words: textbook, text book, over-reliance, resource (resources), math (maths, mathematics, mathematical), conceptual, procedure, procedural, algorithm, total number of unique words, and total number of words. Table 2 gives the frequencies for the five corpora. The last line shows the total number of documents in the data set/corpora.

Table 2: frequencies for the corpora

	2000-2004	2005-2009	maths2008	2010-2014	maths2012
textbook	2597	304	7	30	5
text book	346	64	0	20	1
over-reliance	185	61	0	37	0
resource	29472	9371	7	3611	35
math	41463	17100	387	20224	485
conceptual	201	67	8	81	17
procedure	16200	10343	2	6207	0
procedural	23	2	2	4	1
algorithm	27	3	1	3	1
unique words	88212	121112	3911	46360	4503
total words	28464140	12676440	28685	11149713	36404
total no reports	1132	3556	1	4871	1

FINDINGS AND DISCUSSION

Table 2 shows there are distinct differences in frequencies of certain key words, even when taking into account the total number of (unique) words in the texts. There also seems to be an increase in the number of reports, with these on average containing fewer words. This might be explained by the introduction of more interim reports and letters. A qualitative analysis also shows there certainly are examples of quotes within the reports that are very telling with regard to the vision on learning and teaching mathematics. One report stated:

...They develop a methodical approach to finding solutions rather than relying on memorising a collection of algorithms...

Another:

...Weaker factor: teaching procedural rules rather than developing understanding...

Observations like these, however, must be scrutinized carefully and take into account the context of the word use as well. As it stands, the word counts might be an indication of certain developments, but may not yet be reliable enough.

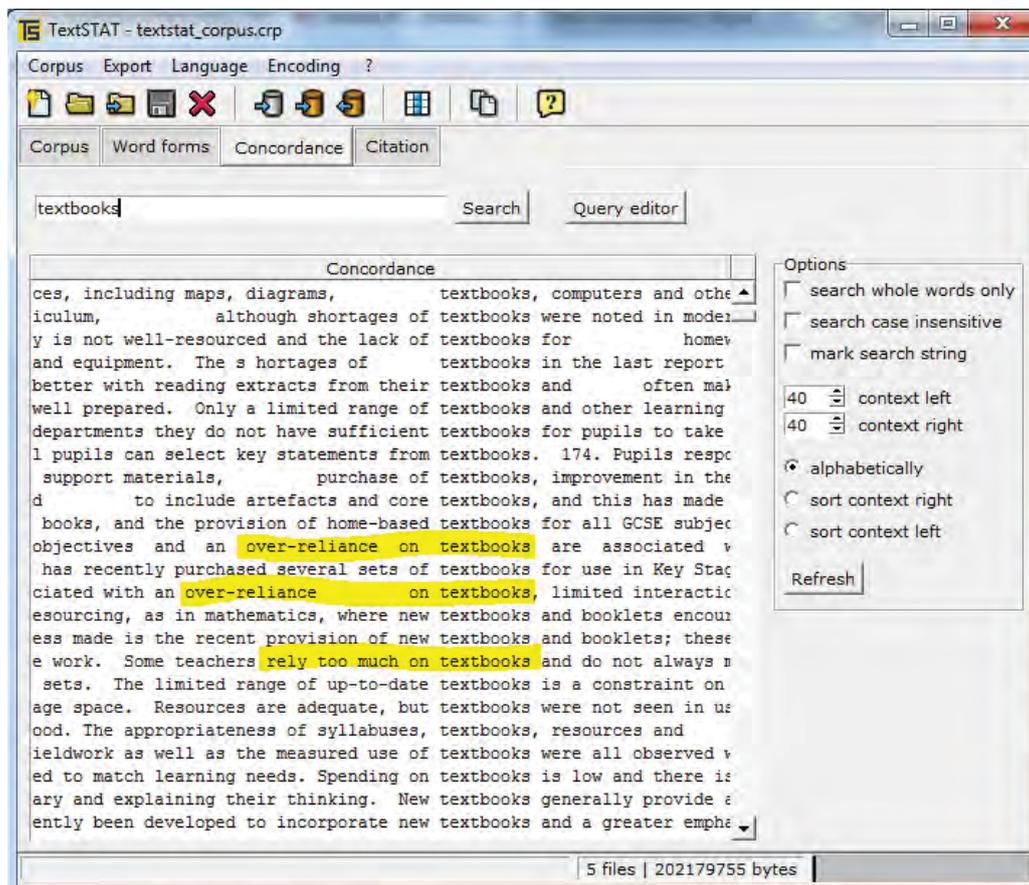


Figure 1: example of concordance use with key word ‘textbooks’, ‘over-reliance’ highlighted. The sharp decrease in the use of the words ‘textbook’ and ‘text books’ might indeed indicate that textbooks have been valued less and less over the years. As such, it is to be expected that comments on the ‘over-reliance’ of textbooks would decrease as well. For example, the 2005-2009 documents flagged up several uses of the word ‘over-reliance’, as also demonstrated in Figure 1. Examples included:

“...dominated by an over-reliance on textbooks...”

“...also sometimes an over-reliance on textbook activities...”

Even in these sections one could argue that textbooks and textbook activities are quite different. Further, the quantifier ‘sometimes’ seems less pronounced than ‘dominated’. ‘over-reliance’ is also used for different topics:

...is therefore an over-reliance on work sheets...

...over-reliance on web-based mathematics packages...

Work sheets can certainly be seen as different from textbooks, but arguably could be grouped under ‘resources’. Another example is:

...over-reliance on whole school initiatives....

This might not be the type of ‘over-reliance’ one is searching for, and it certainly is not connected to the topic of this paper, that of textbooks and resources. The same thing happened to ‘resources’ where URLs with ‘resources’ in them were included.

Nevertheless, a further qualitative analysis of a sample of Ofsted publications found mention of what Ofsted called ‘over-reliance’ on textbooks: “in over a third of classes there was an over-reliance upon a particular published scheme” and this “usually led to pupils spending prolonged periods of time in which they worked at a slow pace, often on repetitive, undemanding exercises, which did little to advance their skills or understanding of number, much less their interest and enthusiasm for mathematics” (Ofsted, 1993, p. 16).

Finally, there is the issue of ambiguity in terms. In mathematics education ‘procedural’ and ‘conceptual’ are often used to denote procedural skills and conceptual understanding. The words’ moderate occurrence and ratio might indicate a preference by Ofsted for ‘conceptual’ rather than ‘procedural’, an observation befitting criticisms along these lines. If one looks at the word ‘procedure’ then we are presented with a very high word count. This is mostly caused by the word ‘procedure’ being used for more purposes than learning and teaching, for example a ‘safety procedure’. Still, if only a small proportion of the word implies a ‘procedural’ aspect of learning, it could ‘normalize’ the ratio procedural/conceptual. It is exactly this *contextual* aspect that would need to be taken into account by future text analyses.

In addition to these challenges for interpretation, there also are numerous technical challenges. One issue concerns the processing phase. Even in this experiment the transformation of PDF files was not straightforward. Converting PDF documents to text format depends on whether the file is not password protected. Another challenge concerned special Unicode symbols, missing spaces and other formatting issues. This might have influenced some of the resulting data; it is however hypothesized that given the large amount of data and documents the influence was relatively small. One interesting aspect to explore would be whether the expressed views differ over judgments: are there differences and similarities between inadequate and outstanding schools, for example. In a next phase we aim to explore the usefulness of more advanced text analysis and mining methods like sentiment analysis, association mining, semantic analysis, and Latent Dirichlet Allocation (LDA) which is a form of so-called topic modelling (e.g. Blei, Ng, & Jordan, 2003). Worded in laymen

terms these techniques these techniques might allow us to discover themes and patterns in the reports' choice of words, for example whether certain words and topics occur more frequently for certain (types of) schools. As Figure 2 further exemplifies, these might be visualized as well. In any case, like any model, the biggest challenge in all of this seems to be the *interpretation* of the findings.

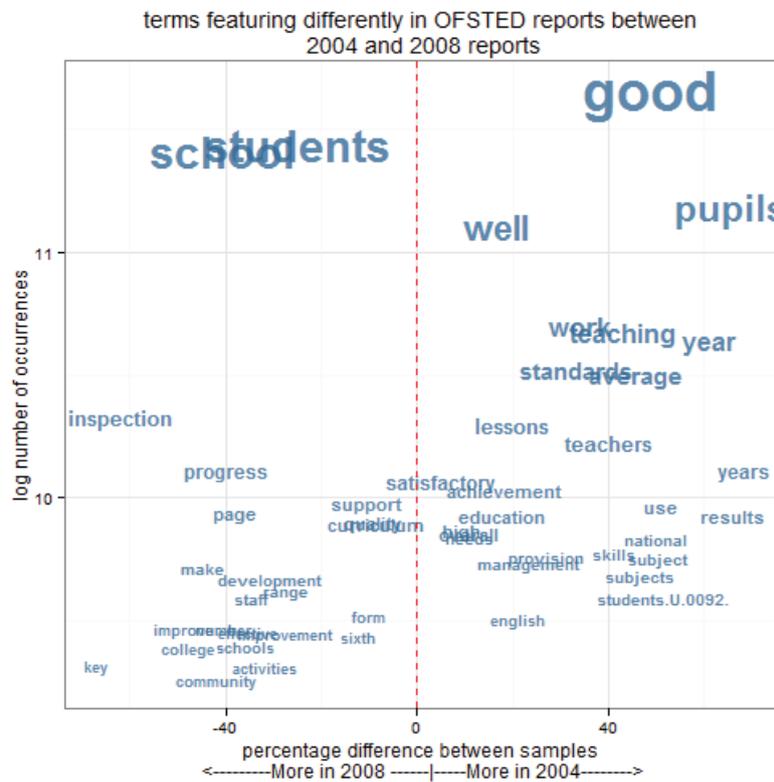


Figure 2: example visualization for Ofsted reports in 2004 and 2008

CONCLUSION

This paper reports on the first results of text mining analysis of the most recent publicly-available Ofsted inspection reports and interim reports from 2000 to March 2014. During this period it may be expected that policy changes have taken place, which might be concurrent with the relevant head inspector. This analysis focuses on what Ofsted has written about (mathematics) textbooks in documents over this period, and how this might relate to changes in Ofsted leadership. Although we conclude that interpretation of the current results proved to be difficult, this paper serves two purposes: first, to show that web-scraping and text analysis methods can be used to analyse policy changes (in this case on text-books and mathematics); second, to demonstrate an application of basic text analysis techniques. In relation to Ofsted's own policies, their re-issued Schools inspection policy FAQs (Ofsted, 2014) contains the following Q & A which is illustrative for this topic:

Q16. Does Ofsted discourage the use of textbooks to support teaching?

A16. No. Ofsted has no preference as to whether textbooks or other teaching materials are used. Inspectors are interested in the standards achieved by pupils and the progress made by them. Teachers are free to use whatever resources they see fit to prepare for, and teach a

lesson including textbooks. Inspection reports rarely criticise the use of textbooks unless they are out of date. Inspection reports are more likely to criticise an over-reliance on worksheets, which may be poorly reproduced, incomplete or insufficiently linked to earlier work.

In future research we aim to utilize more sophisticated techniques to see whether claims like these can be further substantiated.

References

- Askew, M., J. Hodgen, et al. (2010). *Values and variables: Mathematics education in high-performing countries*. London: Nuffield.
- Baxter, J., & Clarke, J. (2013). Farewell to the tick box inspector? Ofsted and the changing regime of school inspection in England. *Oxford Review of Education*, 39(5), 702-718.
- Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022. doi:10.1162/jmlr.2003.3.4-5.993.
- Bosnjak, Z., Grljevic, O., & Bosnjak, S. (2009). CRISP-DM as a framework for discovering knowledge in small and medium sized enterprises' data. In *Proceedings of the 5th International Symposium on Applied Computational Intelligence and Informatics* (pp. 509-514). IEEE.
- Campbell, J. & Husbands, C. (2000). On the reliability of Ofsted inspection of initial teacher training: a case study. *British Educational Research Journal*, 62(1), 39-48.
- Feinerer, I., & Hornik, K. (2014). *tm: Text Mining Package*. R package version 0.5-10. <http://CRAN.R-project.org/package=tm>
- Jones, K., & Sinkinson, A. (2000). A critical analysis of Ofsted judgements of the quality of secondary mathematics initial teacher education courses. *Evaluation & Research in Education*, 14(2), 79-93.
- Jones, K., & Tymms, P. (2014). Ofsted's role in promoting school improvement: the mechanisms of the school inspection system in England. *Oxford Review of Education*, 40(3), 315-330.
- Meyer, D., Hornik, K., & Feinerer, I. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, 25(5), 1-54.
- Ofsted (nd). <https://www.gov.uk/government/organisations/ofsted>
- Ofsted (1993). *The teaching and learning of number in primary schools: National curriculum mathematics attainment target 2*. London: HMSO.
- Ofsted (2014). *Schools inspection policy: some FAQs*. London: Ofsted.
- Ofsted (2014). *Inspecting schools: handbook for inspectors*. London: Ofsted.
- Penn, H. (2002). 'Maintains a Good Pace to Lessons': Inconsistencies and contextual factors affecting OFSTED inspections of nursery schools. *British Educational Research Journal*, 28(6), 879-888.
- Sinkinson, A., & Jones, K. (2001). The validity and reliability of Ofsted judgements of the quality of secondary mathematics initial teacher education courses. *Cambridge Journal of Education*, 31(2), 221-237.