MATHEMATICS EDUCATION IN SOUTH AFRICA: MANY PERSPECTIVES, MANY VOICES

David A. Thomas
University of Providence, U.S.A.

Heather A. Handy
Georgia Institute of Technology, U.S.A.

Gerrit H. Stols
University of Pretoria, South Africa

ABSTRACT: Mathematics education and its perceived significance as a foundation of national economic competitiveness have become a topic of intense scrutiny and debate in many countries. Researchers, government officials, and the media in South Africa bring to this debate their respective concerns, perspectives, recommendations, and voices. This study used automated textual analysis and visualization technologies to search a collection of 148 scholarly articles, dissertations and books, government-related publications, and media reports about mathematics education in South Africa for evidence of recurring, distinctive voices characteristic of these communities of thought and action. The analysis found that these communities characteristically do favor the use of different concepts in representing their concerns, perspectives, and recommendations. Characteristic preferences are also observed with respect to clusters of concepts, called themes. These and other findings are presented in both tabular and graphical formats.

Keywords: mathematics, education, perspectives, automated textual analysis, leximancer

INTRODUCTION

Much of what humanity knows and believes is expressed and preserved as unstructured data. We use text, graphics, and music to tell the stories of our lives, explain the history of the universe, and to reveal our thoughts and feelings. Our storehouse of knowledge and experience is vast, complex, messy, and growing exponentially. To cope with the information explosion, scholars in many knowledge domains rely on sophisticated information technologies to search for and retrieve records and publications pertinent to their research interests. But what is a scholar to do when a search identifies hundreds of documents, any of which might be vital or irrelevant to his/her work? More and more, scholars are turning to automated content analysis technologies to achieve what they do not have time to do themselves; characterize the global features of a large corpus of work and identify relationships between significant concepts and themes.

There are several reasons why one would want an automated system for content analysis of documents (Smith & Humphreys, 2006). Researchers are subject to influences that they are unable to report (Nisbett & Wilson, 1977) which may lead to subjectivity in data analysis and the interpretation of findings. Limiting researcher subjectivity often involves extensive investments of time and money to address interrater reliability and other sources of bias. One goal of automated content analysis is to reduce this cost and to allow more rapid and frequent analysis and reanalysis of text. A related goal is to facilitate the analysis of massive document sets and to do so unfettered by *a priori* assumptions or theoretical frameworks used by the researcher, consciously or unconsciously, as a scaffold for the identification of concepts and themes in the data (Zimitat, 2006). Since textual analysis technologies operate directly on words (as well as other symbols), a rationale for inducing relationships between words is needed. Beeferman observed that words tend to correlate with other words over a certain range within the text stream (Beeferman, Berger, & Lafferty, 1997). Indeed, a word may be defined by its context in usage (Smith and Humphreys, 2006, p. 262; Courtial, 1989; Leydesdorff and Hellsten, 2006; Lee and Jeong 2008).

Mathematics education and its perceived significance as a foundation of national economic competitiveness has become a topic of intense scrutiny and debate in many countries. Educators, researchers, government officials, the private sector, and the media bring to this debate their respective concerns, perspectives, recommendations, and voices. Motivated by a hunch that different communities of thought and action tend to speak with distinctive voices, this study used automated textual analysis and visualization technologies to search a collection of 148 recent (2011 – 2015) journal articles, dissertations and books, South African research organizations, Opinion Pieces on the Ministry of Education website, and Media24 reports identified by their use of the keywords *mathematics*, *education*, and *South Africa* for evidence of recurring, distinctive voices.

Research Questions

This study asks the following research questions relative to a sample of 148 documents (i.e., files) identified and downloaded from a variety of online sources:

1. Which concepts
   1.1. Occur most frequently?
   1.2. Co-occur most frequently?
   1.3. Are associated with particular communities of thought and action?
2. What themes (i.e., co-occurring sets of concepts) emerge?

METHOD

This is an informal survey of professional and public discourse relative to mathematics education in South Africa. Metaphorically, the study casts a wide net into an ocean of unstructured data to identify recurring concepts and themes associated with the ongoing debate of mathematics education in South Africa. All documents were analyzed using *Leximancer* (2015), a textual analytics tool that automatically extracts a dictionary of terms from source documents, discovers concepts, and constructs a thesaurus of terms associated with each concept. *Gephi* was then used to visualize and further explore these networks of concepts as identified in Leximancer.

In *Leximancer*, concepts are collections of words that "travel together" (i.e., co-occur) throughout the text. For example, in a document about climate change, the concept *carbon* might, in the thesaurus, be associated with the keywords *dioxide*, *carbonate*, *footprint*, and *sequester*. *Leximancer* weights these terms according to how frequently they occur in sentences containing the concept, compared to how frequently they occur elsewhere. A sentence (or group of sentences) is only tagged as containing a concept if the accumulated evidence (the sum of the weights of the keywords found) is above a set threshold. These data are used to make two determinations: (i) the most frequently used concepts within a body of text; and more importantly, (ii) the relationships between these concepts (e.g., the co-occurrence between concepts). Discovered concepts may be displayed in ranked lists, by frequency of occurrence, or in graphical format. This approach to concept discovery, in addition to being unbiased, relieves scholars of the task of formulating their own coding schemes, while permitting the introduction of undiscovered terms at the scholar's discretion.

Leximancer discovers and extracts thesaurus-based concepts directly from the text data (Smith and Humphreys, 2006) using Boolean algorithms. Consequently, concepts are robust statistical artifacts. These concepts are then coded into the text (i.e., tags are inserted) using the thesaurus as a classifier. This process employs two stages of co-occurrence information extraction—*semantic* and *relational*—using a different algorithm for each stage. Clusters of co-occurring concepts are then aggregated into themes. The algorithms used are statistical, but they employ nonlinear dynamics and machine learning. The resulting asymmetric concept co-occurrence information is then used to generate a concept map and tabular outputs. For an over view of Leximancer features see Thomas (2014).

Leximancer processes text in a series of four stages: Load Data, Generate Concept Seeds, Edit Emergent Concept Seeds, Develop Concept Thesaurus, Create Compound Concepts, Generate Thesaurus, and Run Project. Figure 1 shows the Leximancer Project Control Panel.

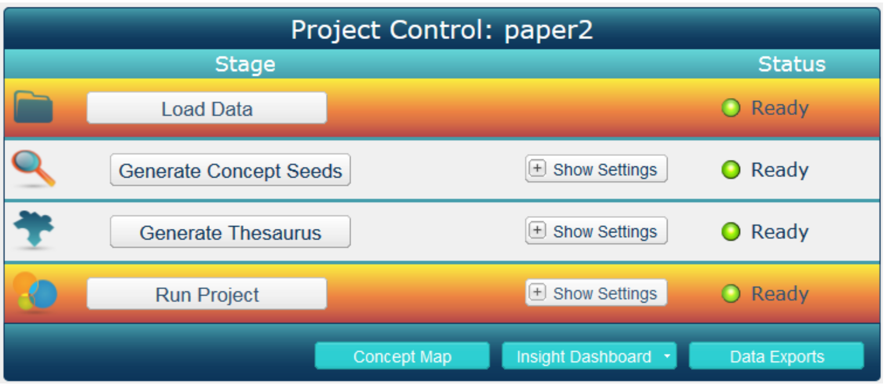

Figure 1. Leximancer Project Control Panel

At each stage, Project Control prompts the researcher to edit the manner in which *Leximancer* processes the data. The editing options provided at each of these stages enable the researcher to, in effect, change the size of the "sliding window", add or delete documents, merge similar concepts (e.g., student, students, pupils), delete irrelevant or distracting concepts (e.g., recurring formatting terms like *Figure* and *Table*), propose additional concepts, and so on. All such choices are reversible, so experimentation is both possible and desirable in refining the analysis and presentation graphics.

Sampling and Selection of Documents

The sample of documents used in this study is neither random nor formally systematic. No generalizations are made about the population based on the characteristics of the sample. Documents associated with different communities of thought and action were identified. The research assumption of this paper was that *researchers' perspectives* about mathematics in South Africa are reflected by articles about mathematics in three South African journals published by different research organizations in 2014 and 2015:

- South African Journal of Education, the official journal of the Education Association of South Africa
- African Journal of Research in Mathematics, Science and Technology Education, the official journal of the Southern African Association for Research in Mathematics and Science Education (SAARMSTE)
- Pythagoras, the official journal of the Association for Mathematics Education of South Africa (AMESA)

The second assumption was that the *government's voice* about mathematics in South Africa are reflected by:

- *Media statements* about school mathematics in 2014 and 2015 by Angie Motshekga, the minister of Basic Education.
- Publications about mathematics education by the *Human Sciences Research Council*, South Africa's statutory research agency.
- Publication about mathematics education by *JET Education Services*, an independent, non-profit organization that works with government, the private sector, international development agencies and education institutions to improve the quality of education and the relationship between education, skills development and the world of work.

The third and last assumption was that the media's voice about mathematics in South Africa is reflected by:

- News paper reports about school mathematics in the archives the website of *Media24*, the largest media house in South Africa. Media 24 is a division of the South African media company Naspers.

Table 1 lists the number of documents associated with each community of thought included in the sample and their respective labels in the figures and tables in this paper.

Table 1. Number of documents included in sample

Community of Thought and Action	Label	#
Human Sciences Research Council	hsrc	8
JET Education Services	jet	9
South African Journal of Education	sage	15
African Journal of Research in Mathematics, Science and Technology Education	ajr	18
Pythagoras	pythagoras	15
Media24	media	74
Opinion Pieces by Minister Angie Motshekga	motshekga	9

RESULTS AND FINDINGS

Concepts

Leximancer depicts concept relationships in graphical representations called network spanning trees (See Figure 2). In such trees, concept nodes, individual files, and file folders appear as circular nodes. Rather than plot all 148 file nodes (which would hopelessly clutter figures), only the 7 file folders (e.g., hsrc, media, etc.) and 60 discovered concepts are plotted. The position of each folder reflects the aggregated concepts of its constituent files. In *Leximancer* network spanning trees, co-occurrences appear as segments. Concept proximity is strongly related to concept co-occurrence, that is, concept nodes positioned near to one another co-occur more frequently than more widely separated concepts. But the edges in *Leximancer* network spanning trees indicate only the most likely co-occurrences associated with each concept. All co-occurrences are seen in Figure 3. This *hair ball* representation was created using *Gephi* (2015), a network visualization tool.

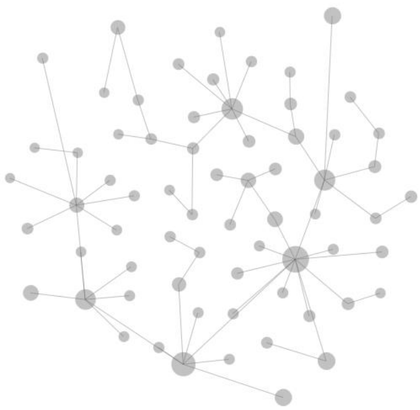

Figure 2. *Leximancer* Network Spanning Tree

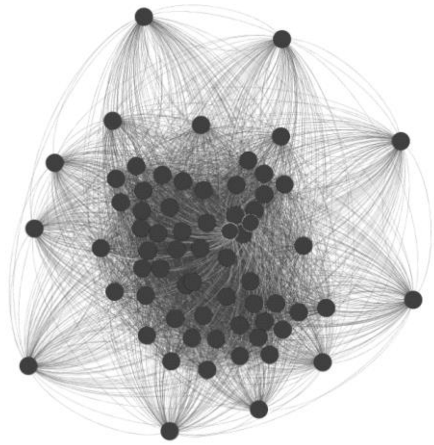

Figure 3. *Gephi* Network

In Figure 4, all discovered concepts and file folders appear as node labels. The 20 most frequent concepts occurring across the entire data set appear in Table 2. Count refers to the frequency of occurrence of a concept. Relevance is the percentage frequency of text segments which are coded with that concept, relative to the frequency of the most frequent concept in the list.

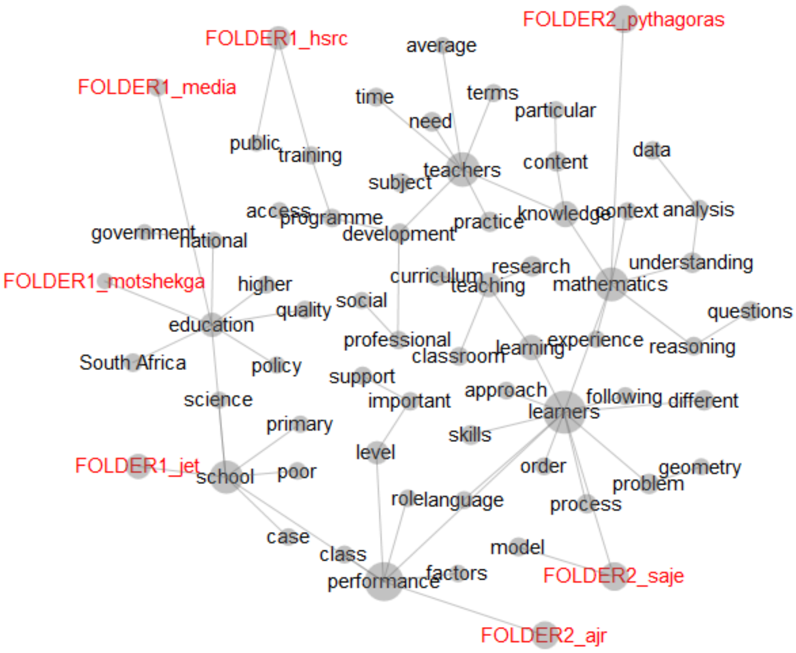

Figure 4. Concepts and Folders

Table 2. The 20 most frequent concepts across all the documents

	Concept	Count	Relevance		Concept	Count	Relevance
1	learners	4201	100%	**11**	different	875	21%
2	performance	3854	92%	**12**	problem	850	20%
3	mathematics	2770	66%	**13**	questions	829	20%
4	teachers	2668	64%	**14**	understanding	820	20%
5	school	2642	63%	**15**	research	791	19%
6	knowledge	1388	33%	**16**	curriculum	717	17%
7	learning	1320	31%	**17**	practice	698	17%
8	education	1288	31%	**18**	process	670	16%
9	teaching	1221	29%	**19**	need	665	16%
10	level	1137	27%	**20**	skills	664	16%

- Research questions 1.1 asks, which concepts occur most frequently across the entire document set? The answer to this question is found in Table 2: *learners, performance, mathematics, teachers, school, knowledge, learning, education, teaching, level, different, problem, questions, understanding, research, curriculum, practice, process, need,* and *skills*.
- Research question 1.2 asks, which concepts co-occur frequently? The answers to this question are implicit in Figure 4. For instance, the proximity of *government* and *national* suggests that they co-occur frequently. The same may be said of the concepts *mathematics* and *understanding*. The figure also suggests that the concepts *government* and *understanding* are unlikely to co-occur frequently.
- Research question 1.3 asks, which concepts are associated with particular communities of thought and action? The answers to this question are found in Tables 3 – 6.

Table 3. Most frequent concepts: *hsrc* and *jet*

	hsrc				***jet***		
	Word-like	Count	Likelihood		Word-like	Count	Likelihood
1	government	97	44%	**1**	professional	180	41%
2	training	147	37%	**2**	government	87	39%
3	public	93	37%	**3**	school	890	34%
4	higher	130	32%	**4**	policy	95	31%
5	quality	109	29%	**5**	training	121	31%
6	access	72	29%	**6**	poor	98	30%
7	national	74	28%	**7**	primary	84	27%
8	programme	132	27%	**8**	national	71	27%
9	social	64	23%	**9**	education	337	26%
10	policy	69	23%	**10**	case	105	26%

Table 4. Most frequent concepts: *sage* and *ajr*

	sage				***ajr***		
	Word-like	Count	Likelihood		Word-like	Count	Likelihood
1	model	228	41%	**1**	geometry	152	53%
2	role	127	37%	**2**	problem	307	36%
3	factors	112	34%	**3**	factors	117	36%
4	skills	226	34%	**4**	context	143	33%
5	learning	412	31%	**5**	experience	106	32%

6	policy	90	30%	**6**	curriculum	203	28%
7	reasoning	151	27%	**7**	mathematics	781	28%
8	content	167	27%	**8**	understanding	231	28%
9	learners	1102	26%	**9**	questions	232	28%
10	class	119	26%	**10**	process	187	28%

Table 5. Most frequent concepts: *Pythagoras* and *Media 24*

	pythagoras				***media***		
	Word-like	Count	Likelihood		Word-like	Count	Likelihood
1	understanding	322	39%	**1**	science	75	23%
2	analysis	218	39%	**2**	national	33	12%
3	average	136	38%	**3**	language	42	11%
4	reasoning	184	33%	**4**	subject	57	10%
5	particular	131	33%	**5**	need	63	09%
6	context	140	33%	**6**	education	118	09%
7	data	187	32%	**7**	school	236	09%
8	questions	260	31%	**8**	training	34	09%
9	different	274	31%	**9**	programme	40	08%
10	content	192	30%	**10**	time	50	08%

Table 6. Most frequent concepts: *motshekga*

	motshekga		
	Word-like	Count	Likelihood
1	quality	39	10%
2	national	23	09%
3	support	27	06%
4	access	15	06%
5	subject	32	06%
6	education	70	05%
7	programme	26	05%
8	public	12	05%
9	development	28	05%
10	social	12	04%

Themes

Unlike concepts, which are unbiased statistical artifacts, themes are clusters of concepts that reflect the intent and interests of the researcher. For example, consider Figure 5. Two circles sketched informally by the authors suggest clusters of concepts that frequently co-occur with *schools* and *mathematics*. *Leximancer* performs clustering automatically and analytically while providing the researcher a measure of control over the way the themes are displayed. Figure 6 illustrates this approach to include all sample documents. Theme names correspond to the most frequently occurring concept in the theme. Alternatively, Figure 7 lists the themes and their comparative relevance while Table 6 identifies the concepts subsumed under each theme. Figure 8 displays all concepts, themes, and folders.

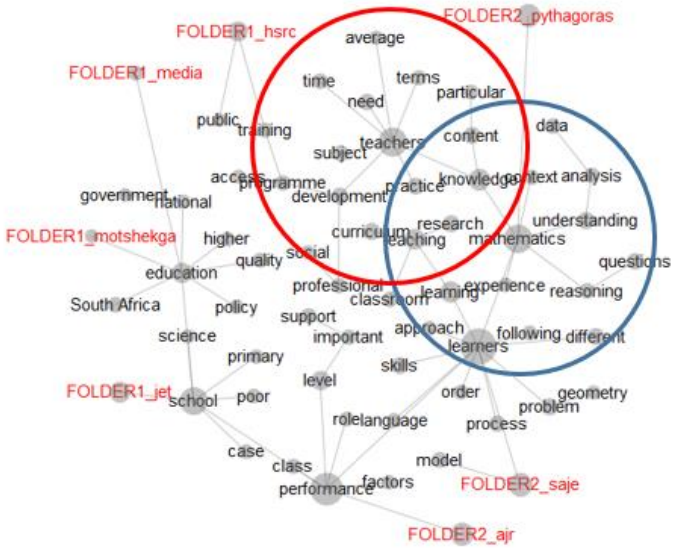

Figure 5. Themes Sketched by Authors: *teachers* and *learners*

Research question #2 asks, “What themes (i.e., co-occurring sets of concepts) span the concepts in a useful manner?” Figures 6, 7, and 8 and Table 7 identify a set of themes that aggregate concepts into categories resonate with important elements of mathematics education discourse. This set of themes serves the authors’ purpose in making that connection.

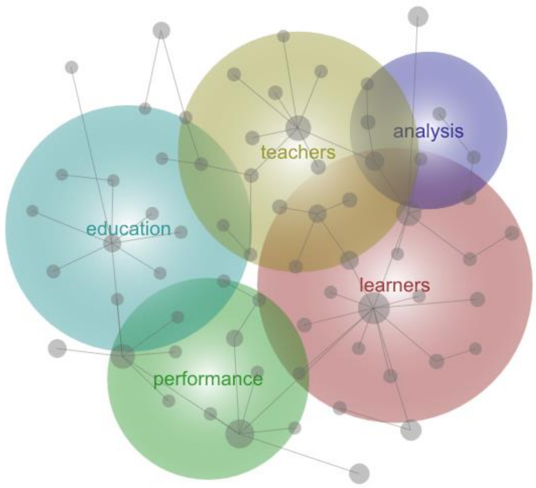

Figure 6. Themes Spanning the Entire Data Set

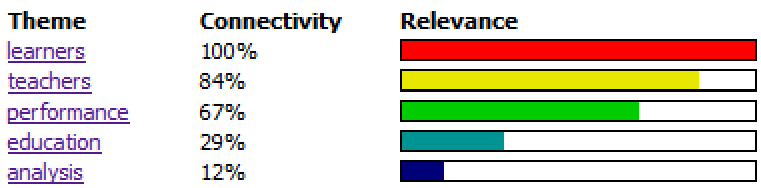

Figure 7. Theme Connectivity & Relevance

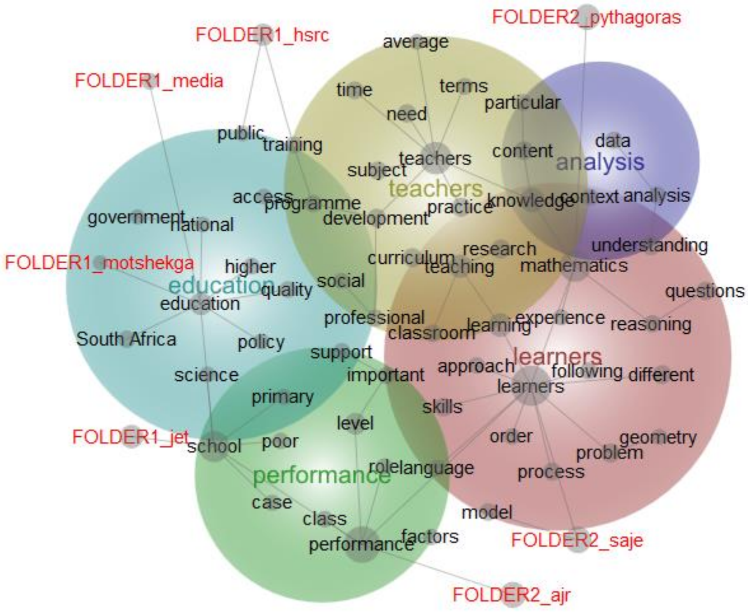

Figure 8. Themes, Concepts, and Folders

Table 7. Themes & Associated Concepts

Theme	Concepts
learners	learners, mathematics, learning, understanding, problem, different, skills, questions, process, model, classroom, reasoning, order, following, approach, experience, geometry
teachers	teachers, knowledge, teaching, research, curriculum, practice, content, development, need, subject, time, program, professional, training, terms, average
performance	performance, school, level, important, support, class, language, role, poor, factors, case, primary
education	education, South Africa, quality, higher, policy, social, science, national, access, public, government
analysis	analysis, data, context, particular

DISCUSSION

Some longtime users and consumers of conventional statistical methods find it difficult to see the purpose, value, or even the validity of textual (i.e., content) analyses of unstructured data (e.g., documents, interview transcripts, blog postings). For instance, when data take the form of words rather than numbers, how are research questions, hypotheses, sampling, data analysis, and conclusions related? Hypothesis testing and its attendant concerns with sampling, bias, error estimates, and other measures of uncertainty appear to have little or no traction in textual analysis. And while a deeper gaze into *Leximancer* reveals that it is based on Bayesian thinking, this insight offers little comfort to researchers unfamiliar with Bayesian methods and purposes.

This paper asks and answers questions about the content of 148 documents using *Leximancer*, an automated qualitative approach founded on Bayesian methods. In this study, data acquisition, management, and analysis are focused on the identification of recurring concepts, their relationships to one another (i.e., co-occurrence and themes), and the communities of thought and action (e.g., publications) from which they were drawn. The goal of this inquiry is to determine whether different communities of thought and action favor the use of different and/or distinctive language (i.e., concepts) in representing their interests, findings, and recommendations. In a time when clear communication among stakeholders is needed to build consensus, understanding and accommodating these differences is essential if mathematics education in South Africa is to speak with a unified voice.

Leximancer identified a total of 60 recurring concepts in 148 documents drawn from 7 communities of thought. The identification and graphical representation of these concepts was performed automatically with limited editorial oversight (e.g., merging word variants) by the authors. This approach dramatically limits researcher bias in the identification of concepts and determining their locations in graphical representations (see Figure 4). Concepts determined in this manner are robust statistical artifacts, reliably and consistently identified in repeated analyses of the document sets. Repeated analyses do not necessarily yield identical graphical representations, however. Each variant on the network spanning tree displays the relationships between concepts from a different perspective, so to speak. For instance, the tree might be rotated, reflected, dilated, sheared, or otherwise rearranged in a manner that at first glance appears substantively different, but which on closer inspection portrays the same information in a different layout, including the relative positions the folders.

The answer to the research question 1a, “Which concepts occur most frequently?” was generated automatically by *Leximancer*: learners, performance, mathematics, teachers, school, knowledge, learning, education, teaching, level, different, problem, questions, understanding, research, curriculum, practice, process, need, and skills. These concepts appear in Table 2, where Count refers to the frequency of occurrence of a concept, and Relevance is the percentage frequency of text segments which are coded with that concept, relative to the frequency of the most frequent concept in the list, learners. This table also makes it clear that the first 5 concepts in the table are most characteristic of the data.

The answer to research question 1b, “Which concepts co-occur most frequently?” is more difficult to characterize. In Figure 4, concepts that are near to one another co-occur more frequently that more widely separated concepts. This representation suggests at a glance which concepts co-occur frequently without leaving out any information. In other words, Figure 4 is not so much a definitive answer to the research question as it is a model of co-occurrence from which specific inferences may be drawn. So, one answer to this question is, “The concepts located near one another in Figure 4”.

Alternatively, a table containing frequency of co-occurrence data could be created. Table 8, which shows the co-occurrence of the 10 most frequently occurring concepts, would be over 36 times as large if all concepts were included. Massive tables, while they have their uses, are not particularly helpful in this context.

Table 8. Partial co-occurrence matrix

	learners	performance	teachers	mathematics	school	knowledge	learning
learners		1626	1024	1228	1009	562	675
performance	1626		746	883	1056	371	401
teachers	1024	746		765	728	646	375
mathematics	1228	883	765		518	498	445
school	1009	1056	728	518		206	299
knowledge	562	371	646	498	206		208
learning	675	401	375	445	299	208	

The answer to research question 1c, “Which concepts are associated with particular communities of thought and action (i.e., folders)?” is based on the proximity of concepts and folders in Figure 4. One answer to this question is, “Concepts located near a particular folder are more characteristic of the documents in that folder than concepts located further away. Documents in folders widely separated from one another emphasize the use of different concepts. These differences are apparent in Tables 3 – 6, which identify the most frequent concepts in each folder. As with co-occurrence, this representation is not so much a definitive answer to the question as it is a model from which specific inferences may be drawn.

Discovered concepts and themes by themselves offer little more than labels for important ideas and their associations. The ideas themselves are embedded in the context blocks tagged by *Leximancer* in the source documents. As such, each concept is directly linked to the context blocks where it occurs. Boolean searches may be used to focus on specific co-occurrences of concepts. For example, a search for text blocks containing the concepts *learners* and *teachers* and *performance* and *education* and *analysis* pointed to 4 text blocks, all from the same HSRC document.

Despite widespread acceptance of the notion that improving student performance may have a high economic and social payoff, policy analysts in all countries have surprisingly limited hard data on which to base educational strategies for raising achievement. In South Africa this question is all the more pressing. South African students score at low levels in mathematics and language tests even when compared with students in other African countries. Further, the South African government’s own evaluations of ten years of democracy show little improvement in educational outcomes despite significant policy changes. While some reasons for this poor performance may be evident, and there is widespread agreement that the main challenge in South Africa is the quality of education, there is little empirical analysis that helps policy makers understand the low level of student performance in South African schools or how to improve it.

Depending on a researcher's interests, the same set of documents can be searched for text blocks containing particular sets of concepts. Those text blocks reveal context and purpose in ways that the concept lists themselves cannot. Viewed in this manner, textual analytics technologies like *Leximancer* offer more than summary findings. They provide a dynamic modeling environment for continued exploration and discovery.

In a subsequent article, similarities and differences between the document sets will be explored in greater depth and the broader implications of automated textual analysis in mathematics education research will be considered.

LIMITATIONS

In *Leximancer*, themes are variable concept groupings used to facilitate data exploration. Unlike concepts, which are robust statistical artifacts, themes are conveniences imposed by the researcher. The themes and subsumed concepts seen in Figure 8 and Table 7 reflect the authors' subjective preferences in this study.

In selecting sample documents for this study, several methodological issues arose for which the authors could find no clear guidance. For instance, selecting sample documents for content analysis is inherently different than selecting sample measurements for statistical analysis. We wondered on what basis a 500-word news story and a 5000-word research report might be considered comparable data. We settled on this interpretation: the most important determination with regard to a particular concept is whether it is present in a given document (and therefore characteristic of the community of thought from which the document was drawn), not how many times it occurs. Longer documents provide more opportunities to detect the co-occurrence of different concepts. Whether a concept occurs only once or many times in a particular document, that concept is formally associated with the document and its community of thought.

REFERENCES

Beeferman D, Berger A, Lafferty J 1997. A model of lexical attraction and repulsion. In *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, pp. 373-380

Courtial JP 1989. Qualitative models, quantitative tools and network analysis, *Scientometrics*, 15(5-6): 527-534.

Gephi: the open graph viz platform. http://gephi.github.io (Retrieved July 10, 2015)

Lee B, Jeong YI 2008. Mapping Korea's national R&D domain of robot technology by using co-word analysis, *Scientometrics*, 77(1): 3-19.

Leximancer: textual analytics. http://info.leximancer.com (Retrieved February 1, 2015)

Leydesdorff L, Hellsten L 2006. Measuring the meaning of words in contexts: an automated analysis of controversies about 'monarch butterflies', 'frankenfoods', and 'stem cells'. *Scientometrics*, 67(2): 231-258.

Media24: print media division of the South African media company Naspers. http://www.media24.com (Retrieved October 5, 2015)

Nisbett RE, Wilson TD 1977. Telling more than we can know: verbal reports on mental processes, *Psychological Review*, 84(3): 231-259.

Pythagoras. http://www.pythagoras.org.za/index.php/pythagoras (Retrieved October 6, 2015)

African Journal of Research in Mathematics, Science and Technology Education, Southern African Association for Research in Mathematics and Science Education. http://www.saarmste.org/journal (Retrieved October 6, 2015)

Smith A, Humphreys M 2006. Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping, *Behavior Research Methods*, 38(2): 262-279.

Smith A 2011. Leximancer Manual (Version 4.0). https://www.leximancer.com/site media/lm/science/Leximancer_Manual_Version_4_0.pdf (Retrieved November 10, 2014)

Thomas D 2014. Searching for significance in unstructured data: text mining with Leximancer. *European Education Research Journal*, 13(2): 235-256.

Zimitat C 2006. A lexical analysis of 1995, 2000 and 2005 Ascilite conference papers. *Proceedings of the 23rd Annual Ascilite Conference*, pp. 947-951.