

# Using Text Mining Techniques to Analyze Students' Written Responses to a Teacher Leadership Dilemma

Yuejin Xu and Noah Reynolds

**Abstract**—This article describes the use of IBM SPSS Text Analytics for Surveys to analyze students' written responses to a teacher leadership dilemma. The purpose of this study was to examine the accuracy of the categories generated by IBM SPSS Text Analytics for Surveys. Our findings from the correlation analyses indicate that a significant interrater reliability existed between the text mining method from IBM SPSS Text Analytics for Surveys and human ratings.

**Index Terms**—Technology in education, text mining, teacher leadership dilemma, IBM SPSS text analytics for surveys.

## I. INTRODUCTION

The teacher leader concept has gained momentum worldwide and in the commonwealth of Kentucky. In 2008, Kentucky's Education Professional Standards Board (EPSB) made teacher leadership as one of the Kentucky Teacher Standards (standard 10), and started to require the public universities in Kentucky to redesign teacher leader master's and planned fifth-year programs to develop teacher leaders. The teacher leader concept is not about training teachers to become school administrators. Instead, it is about empowering teachers to take more active role in school improvement. A teacher leader is expected to identify leadership opportunities within the school, to use student data to improve instruction, to take initiatives in the community, and to collaborate with colleagues. Current researches indicate that teacher leadership is critical for sustained school improvement [1]-[3], and student learning [4]-[7].

### A. Teacher Leadership Pathways and Assessment

Teacher leadership can be fostered by both formal training and job-embedded support [7]. In Kentucky, the Teacher Leader Master (TLM) degree program has been created in public universities to help in-service teachers develop teacher leadership. A typical TLM degree program offers systematic study of leadership skills and strategies to foster teacher leadership. For example, the TLM degree program at Murray State University consists of core courses in teacher leadership, classroom management, curriculum development, instruction for diverse students, and research. In addition, it requires a teacher leader candidate to complete two leadership projects (a classroom-level and a school/ district-level project) in order to successfully exit the TLM program [8]. A second

approach is teacher-led job-embedded support. Semadeni (2010) described one model called Fusion, in which "teachers collaborate to study, experiment, and coach one another in research-based strategies" (p. 66) [9]. The Fusion model was reported to be effective in recognizing the leadership potential in teachers and offering leadership opportunities to further develop that potential.

Current approaches to assessing teacher leadership include objective selected-response type of survey [10], structured observation and interview [11], [12], and teacher leadership dilemma [13]. Structured observation, interview, and teacher leadership dilemma are likely to yield written responses. The dominant method to analyze the written responses in teacher leadership field is to use rubric and independent human raters, which is costly and time-consuming. Consequently, it is meaningful to explore the use of text mining techniques in the assessment of teacher leadership.

### B. Text Mining Techniques

Text mining aims at automating the process of extracting and categorizing useful information from textual data. To date, a variety of text mining techniques have been developed, such as, (1) latent semantic analysis, (2) probabilistic latent semantic analysis, (3) latent Dirichlet allocation, and (4) correlated topic model. Lee, Song, and Kim (2009) reported that generative models (e.g., probabilistic latent semantic analysis, latent Dirichlet allocation, correlated topic model) show higher performance than a discriminative model (e.g., latent semantic analysis) [14]. Built upon these techniques, a number of text mining software have been developed and applied in business [15], [16], [17], and education [18], [19].

### C. Text Mining Software

Both free open source and commercial text mining software are available. Nature Language Toolkit [20] is open source text mining software, however, it requires the mastery of Python language, which could be challenging to some educators without formal programming skills. Some commercial text mining software, such as IBM SPSS Text Analytics for Surveys and Provalis Research's WordStat, use graphic user interface, and seem to be more user friendly. In this article, we report the use of IBM SPSS Text Analytics for Surveys to analyze students' written responses to a teacher leadership dilemma.

The purpose of this study was to examine the accuracy of the categories generated by IBM SPSS Text Analytics for Surveys. Specifically, our research question was: How are findings generated by the text mining algorithm for IBM SPSS Text Analytics for Surveys comparable to the findings identified by an independent human rater using rubrics?

Manuscript received May 30, 2012; revised July 1, 2012. This work was supported in part by a 2010-2011 Regular Research Grant from Committee on Institutional Studies and Research (CISR), Murray State University and a grant from The Council on Postsecondary Education (Award No: PO2 415 1100002518 1) awarded to the first author.

The authors are with the College of Education, Murray State University, Murray, KY 42071 USA (e-mail: yxu@murraystate.edu).

## II. METHODOLOGY

### A. Participants and Setting

The participants in this study were 47 students at a public university in the United States. Twenty-two were undergraduate pre-service teacher education students recruited from one undergraduate course (Psychology of Human Development). Twenty-five were in-service teachers recruited from one graduate course (Methods of Research).

### B. The Teacher Leadership Dilemma Instrument

A teacher leadership dilemma was developed to measure teacher leadership. Participants were presented with the following hypothetical teacher leadership dilemma: "Suppose that you are a high school English teacher, you believe that a particular book (this book could be controversial) is extremely beneficial to your students' learning. You want to use it in your classroom. You tried to get your principal's approval but couldn't. What will you do next?" Participants were instructed to address all the aspects of the open-ended question and write down their response concisely and to the point. Word counts of each participant's written response revealed that the range of words written in the response were 17 words to 111 words with an average length of 55 words.

### C. Two Rubrics

Two rubrics were developed to analyze students' written response. One rubric focuses on the number of implicit stakeholders recognized. The hypothetical dilemma presents three explicit stakeholders: the teacher, the principal, and students. Implicit stakeholders in this dilemma could include parents, colleagues, school board, community, and profession. The more one can identify implicit stakeholders, the more likely the person possesses teacher leadership. The other rubric is used to determine the risk-taking level of the action one would like to engage in with 1 standing for minimum risk and 5 standing for extreme risk. One possible action of minimum risk could be to obey the principal's wishes. One possible action of extreme risk could be to use the book anyway without the principal's approval. It is argued that the willingness to take risks or leads also reflects one's leadership potential.

### D. Text Mining Procedures

IBM SPSS Text Analytics for Surveys was used in this study. We first prepared the data according to the software's data input requirements. Second, we used the New Project Wizard (Fig. 1), through which we defined variables for our analysis, and chose extraction options (Fig. 2).

Then, IBM SPSS Text Analytics for Surveys automatically extracted concepts and frequency of concepts (Fig. 3). The last step involved in building categories using rules and text matching techniques (Fig. 4).

We developed the text mining algorithm (rules and text matching techniques for IBM SPSS Text Analytics for Surveys) based on the two rubrics.

### E. Data Analysis

IBM SPSS Text Analytics for Surveys has the capacity to export categories into SPSS and EXCEL in dichotomies data

type and categories data type. To compare the findings from text mining and human rating, a Pearson correlation coefficient was calculated between the numbers of recognized implicit stakeholders reported by text mining and by the human rating. A second Pearson correlation coefficient was calculated between the risk-taking level reported by text mining and by human rating.

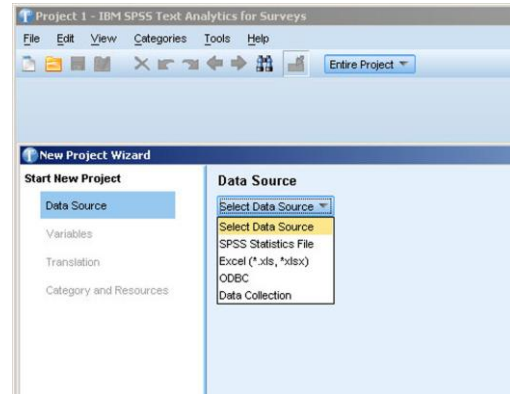


Fig. 1. The new project wizard.

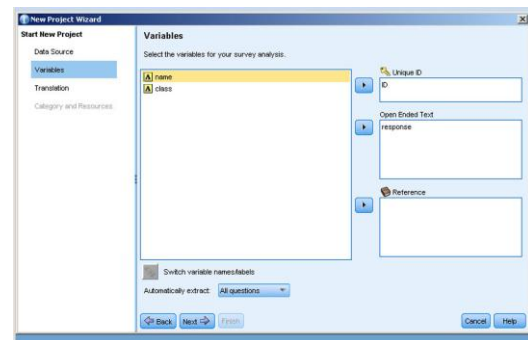


Fig. 2. Variables definition and options selection.

## III. RESULTS

### A. Recognized Implicit Stakeholders

IBM SPSS Text Analytics for Surveys yielded 241 unique concepts and their frequencies. Table I showed the unique concepts that appeared three times or more in the data according to the concepts extraction procedure.

IBM SPSS Text Analytics for Surveys offers some tools to further examine each unique concept in its occurrence in the data, and manually categorize each unique concept into types, and patterns. It also allows the users to create their own text matching rules and categorization algorithms in building categories. We applied the implicit stakeholder algorithm. Out of the 47 responses, 20 responses (42%) were successfully categorized as containing implicit stakeholders. The remaining 27 responses were placed in the no-recognized-implicit-stakeholder category. The results were exported into SPSS. For each participant, a total number of recognized implicit stakeholders were computed. The total number of recognized implicit holders has a range of 0 to 3. Table II reported the means and standard deviations of the number of recognized implicit stakeholders by the software and a human rater.

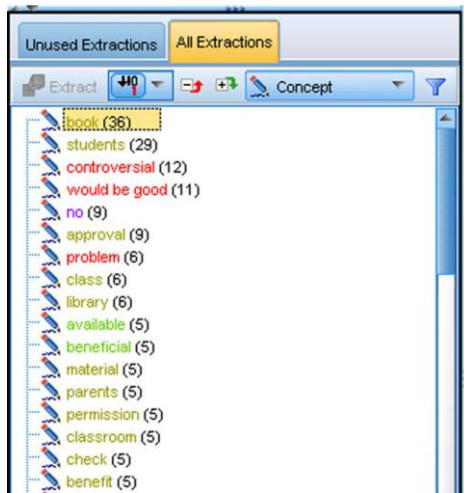


Fig. 3. Extracted concepts.

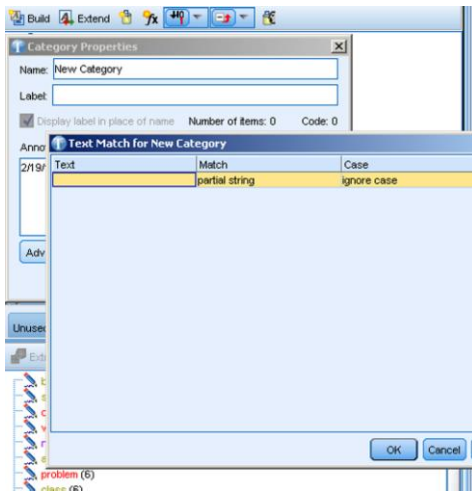


Fig. 4. Building categories.

The number of recognized implicit stakeholders generated from IBM SPSS Text Analytics for Surveys was significantly correlated to that from an independent rater using the rubric ( $r = .877, p < .001$ ).

**B. Risk-Taking Levels**

IBM SPSS Text Analytics for Surveys categorized 39 responses after the application of the risk-taking level algorithm. The researcher reviewed the remaining 8 responses (17%) and forced the response into appropriate categories based on the rubric (Fig. 5).

TABLE I: UNIQUE CONCEPTS AUTOMATICALLY GENERATED BY IBM SPSS TEXT ANALYTICS FOR SURVEYS SOFTWARE.

Unique Concepts	Frequencies
book	36
students	29
controversial	12
would be good	11
no	9
approval	9
problem	6
class	6
library	6

Unique Concepts	Frequencies
available	5
beneficial	5
material	5
parents	5
permission	5
classroom	5
check	5
benefit	5
would recommend	4
intention to use	4
reading	4
child	4
learning	4
school	4
work	4
extra credit	3
no intention to use	3
good	3
list	3
students to read	3
book outside of class	3
offer	3
ideas	3
time	3
content	3

The means and standard deviations of the risk-taking levels by the software and a human rater were reported in Table III. The risk-taking level generated from IBM SPSS Text Analytics for Surveys was significantly correlated to that from an independent rater using the rubric ( $r = .484, p = .001$ ).

**IV. DISCUSSION AND CONCLUSION**

IBM SPSS Text Analytics for Surveys is a linguistics-based solution specifically designed for categorizing survey text responses [21]. In the present study, from the 47 responses to a teacher leadership dilemma, IBM SPSS Text Analytics for Surveys can automatically generate 241 unique concepts, which offers some insight to understanding the linguistic structure of our raw data. The power of text mining techniques can be maximized only through the use of proper categorization algorithm. In this sense, the accuracy of text mining techniques is dependent on the algorithm that one develops, which in turn depends on ones' understanding of the data structure and possible expression of patterns.

TABLE II: MEANS AND STANDARD DEVIATIONS OF THE NUMBER OF RECOGNIZED IMPLICIT STAKEHOLDERS BY SOFTWARE AND HUMAN RATER.

	Mean	Standard Deviation	Sample Size
Software	.49	.66	47
Human rater	.53	.75	47

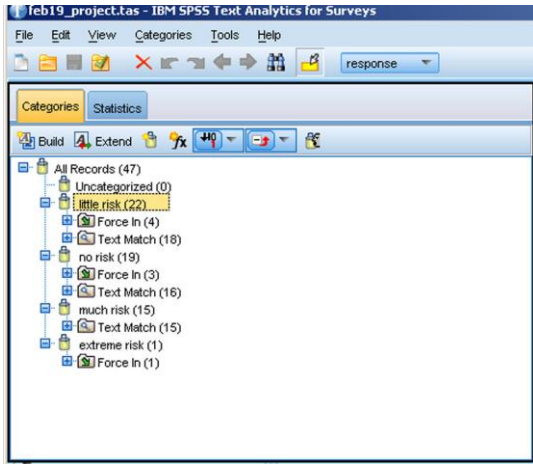


Fig. 5. Automated and forced categorization.

Our findings from the correlation analyses indicate that a significant interrater reliability existed between ratings generated from the algorithms developed for IBM SPSS Text Analytics for Surveys and human ratings. However, our findings also indicated that the strength of the interrater reliability is modest especially in the risk-taking level aspect. A close examination of the responses that were rated differently by the human rater and the software revealed that the algorithm works more accurately when it can cover most of the natural language expressions of a particular concept. For the implicit stakeholders recognized in this dilemma, the kinds of expression are relatively limited and the implicit stakeholders algorithm seems to be able to accurately categorize most of the expressions. Whereas to categorize raw data into appropriate risk-taking levels, the text mining algorithm need go beyond simple text matching into the realm of understanding the meaning of words in the overall context, which is not any easy task. This partially explained the modest strength of the interrater reliability in the risk-taking level aspect.

TABLE III: MEANS AND STANDARD DEVIATIONS OF THE RISK-TAKING LEVELS BY SOFTWARE AND HUMAN RATER.

	Mean	Standard Deviation	Sample Size
Software	2.02	1.03	47
Human rater	2.17	.73	47

A limitation to the present study was that only one independent human rater was employed. It would be meaningful to have two independent raters to further establish the reliability and validity of human rating from the

rubric. In addition, the risk-taking levels algorithm needs to be further improved based on a larger data set. Nevertheless, we believe that the text mining software has great potential in the field of education especially since it is less costly and time consuming than using independent human raters.

REFERENCES

- [1] J. Kise and B. Russell, *Differentiated School Leadership*, Thousand Oaks, CA: Corwin, 2008.
- [2] L. Lambert, *Leadership Capacity for Lasting School Improvement*, Alexandria, VA: ASCD, 2003.
- [3] R. J. Marzane, T. Waters, and B. McNulty, *School Leadership That Works: From Research to Results*. Alexandria, VA: ASCD, 2005.
- [4] R. S. Barth, "Teacher leader," *Phi Delta Kappan*, vol. 82, pp. 443-449, 2001.
- [5] J. G. Gabriel, *How to thrive as a teacher leader*. Alexandria, VA: ASCD, 2005.
- [6] H. M. Marks and K. S. Louis, "Does teacher empowerment affect the classroom? The implications of teacher empowerment for instructional practice and student academic performance," *Educational Evaluation and Policy Analysis*, vol. 19, pp. 245-275, 1997.
- [7] J. York-Barr and K. Duke, "What do we know about teacher leadership? Findings from two decades of scholarship," *Review of Educational Research*, vol. 74, no. 3, pp. 255-316, 2004.
- [8] Y. J. Xu and G. Patmor, "Fostering leadership skills in pre-service teachers," *Presented at AdvancED/UK College of Education Innovation Summit*, Lexington, KY, 2010, unpublished.
- [9] J. Semandeni, "When teachers drive their learning," *Educational Leadership*, vol. 67, no. 8, pp. 66-69, 2010.
- [10] R. Smith, M. Ross, and R. Robichaux, "Creation and validation of a measure of leadership density," *The Journal of Research for Educational Leaders*, vol. 2, no. 2, pp. 79-111, 2004.
- [11] J. Huff, "Measuring a leader's practice: Past efforts and present opportunities to capture what educational leaders do," *Presented at 2006 AERA Meeting*, unpublished.
- [12] J. W. Triska, *Measuring Teacher Leadership*, 2007, unpublished dissertation.
- [13] Y. J. Xu and N. Reynolds, "Academic standing, prior teaching experience, and students' responses to a teacher leadership dilemma," *Presented at the 2011 Association for Psychological Science Annual Convention*, Washington, DC, May 26 - 29, 2011.
- [14] S. Lee, J. Song, and Y. Kim, "An empirical comparison of four text mining methods," *Journal of Computer Information Systems*, vol. 51, no. 1, 2010, pp. 1-10.
- [15] R. Arora and C. Stoner, "A mixed method approach to understanding brand personality," *Journal of Product and Brand Management*, vol. 18, no. 4, pp. 272-283, 2009.
- [16] M. Pullman, K. McGuire, and C. Cleveland, "Let me count the words: Quantifying open-ended interactions with guests," *Cornell Hotel and Restaurant Administration Quarterly*, vol. 46, no. 3, pp. 323-343, 2005.
- [17] C. Chen-Huei, S. Atish, and Z. Huimin, "A text mining approach to Internet abuse detection," *Information Systems and e-Business Management*, vol. 6, no. 4, pp. 419-439, 2008.
- [18] M. Abdous and W. He, "Using text mining to uncover students' technology-related problems in live video streaming," *British Journal of Educational Technology*, vol. 42, no. 1, pp. 40-49, 2011.
- [19] A. Abd-Elrahman, M. Andreu, and T. Abbott, "Using text data mining techniques for understanding free-style question answers in course evaluation forms," *Research in Higher Education Journal*, vol. 9, pp. 1-11, 2010.
- [20] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*, O'Reilly Media, 2009.
- [21] IBM Corporation, "Analyzing survey text: A brief overview," 2010, pp. 1-12.