

# Emergent Coding and Topic Modeling: A Comparison of Two Qualitative Analysis Methods on Teacher Focus Group Data

International Journal of Qualitative Methods  
Volume 22: 1–9  
© The Author(s) 2023  
DOI: 10.1177/16094069231165950  
[journals.sagepub.com/home/ijq](https://journals.sagepub.com/home/ijq)  
SAGE

Atsushi Miyaoka<sup>1</sup>, Lauren Decker-Woodrow<sup>1</sup> , Nancy Hartman<sup>1</sup>, Barbara Booker<sup>2</sup>, and Erin Ottmar<sup>3</sup>

## Abstract

More than ever in the past, researchers have access to broad, educationally relevant text data from sources such as literature databases (e.g., ERIC), an open-ended response from online courses/surveys, online discussion forums, digital essays, and social media. These advances in data availability can dramatically increase the possibilities for discovering new patterns in the data and testing new theories through processing texts with emerging analytic techniques. In our study, we extended the application of Topic Modeling (TM) to data collected from focus groups within the context of a larger study. Specifically, we compared the results of emergent qualitative coding and TM. We found a high level of agreement between TM and emergent qualitative coding, suggesting TM is a viable method for coding focus group data when augmenting and validating manual qualitative coding. We also found that TM was ineffective in capturing more nuanced information than the qualitative coding was able to identify. This can be explained by two factors: (1) the word level tokenization we used in the study, and (2) variations in the terminology teachers used to identify the different technologies. Recommendations include additional data cleaning steps researchers should take and specifications within the topic modeling code when using topic modeling to analyze focus group data.

## Keywords

focus groups, methods in qualitative inquiry, narrative analysis, qualitative evaluation, secondary data analysis

## Introduction

More than ever in the past, researchers have access to a variety of broad, rich, educationally relevant text data from a number of sources, such as literature databases (e.g., ERIC), open-ended responses from online courses/surveys, online discussion forums, transcribed audio of face-to-face classes or focus groups, digital essays, and social media. These advances in data availability (coupled with emerging analytic techniques) can dramatically increase the possibilities for discovering new patterns and testing new theories in educational contexts. This article reports on the similarities and differences in results from two different methods of analysis for qualitative focus group data about teachers' experiences when using four educational technologies: Topic Modeling (TM) and emergent qualitative coding using NVivo.

## Background about Topic Modeling

TM is an unsupervised machine learning method for classifying documents by detecting word and phrase patterns within the text and finding word groups and similar expressions that best characterize a set of documents. TM is often used to identify which underlying concepts are discussed within a collection of documents and can help determine which topics each document is addressing. For example, [Mi et al. \(2020\)](#)

<sup>1</sup>Westat, Rockville, MD, USA

<sup>2</sup>Independent Consultant, Cumming, GA, USA

<sup>3</sup>Worcester Polytechnic Institute, Worcester, MA, USA

### Corresponding Author:

Atsushi Miyaoka, Westat, 1600 Research Boulevard, RB2114, Rockville, MD 20850, USA.

Email: [atsushimiyaoka@westat.com](mailto:atsushimiyaoka@westat.com)



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE

and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

used TM to uncover the trends and foundations in research on students' conceptual understanding in science education. They categorized articles into 10 topics that considered information about the semantic cohesion and exclusivity of words to topics. Taking advantage of modern computational advancement, TM was also used to identify the underlying topics and the topic evolution in the 50-year history of educational leadership research literature (Wang et al., 2017). Researchers were able to map the temporal terrain of topics in the educational leadership field over the past 50 years and shed new light on the development and current status of the central topics in educational leadership research literature. In addition, TM was used to identify the future research directions in the field of personalized language learning (Chen et al., 2021). Moreover, TM was used to identify facts, relationships, and assertions that would otherwise remain buried in the mass of textual data. For example, Vijayan (2021) thematically modeled the literature related to teaching and learning during, and about, COVID-19. Specifically, abstracts and metadata of literature were extracted from Scopus (3461 documents) and through identifying the key research themes, the study uncovered inequities in education as a result of the digital divide.

In addition, researchers can use TM to examine online discussion forums or open-ended responses. For example, Peng et al. (2020) attempted to gain overviews of automatic tracking and understanding of temporal topic changes in small private online courses (SPOCs) discussion forums. They improved the TM by incorporating time, emotion, and behavior characteristics into the process, were also able to uncover students' temporal focuses (i.e., the changes of topic intensity and topic content), and reflected the evolution of a topics' emotional and behavioral tendencies. In another study, Ozeran and Martin (2019) conducted a pilot project that tested the application of four different algorithmic TM—i.e., Latent Dirichlet Allocation (LDA), phase-LDA, Dirichlet Mixture Model (DMM), and Non-negative Matrix Factorization (NMF)—to chat conversations. The purpose of their study was to determine if and how the TM approach could best identify the most common chat topics in a semester and whether these topics could inform library services beyond chat reference training. In a study by Li et al. (2021) TM was found to be useful for comparing an instructor's and their student's perceptions of online teaching and learning. While researchers identified instructional practices that were perceived by both instructors and students as effective, they identified a handful of different topics or themes that were divergent between instructors and students in their perceptions of online teaching.

### *Previous Comparisons of TM and Traditional Qualitative Approaches*

Several researchers have analyzed the usefulness and results of TM compared to a more-traditional qualitative approach. Baumer et al. (2017), for instance, compared qualitative and computational methods from open-ended survey items (1095 responses). They found that the two analytical methods

produced some similar and some complementary insights about the phenomenon of interest. Further, they noted that those complementary topics did not strictly map either set of results onto the other. They also noted each approach involved an iterative process with non-negligible amounts of researchers' subjective judgment in the application process and in interpreting results. Similarly, Leeson et al. (2019) explored the potential of Natural Language Processing (NLP) to analyze qualitative data. NLP is a machine learning technique which helps a machine process and understand the human language with the use of computer algorithms. Leeson et al. (2019) pointed out that NLP is becoming more frequently employed and shows promise as a tool to analyze qualitative data in public health. In their study, they compared a qualitative method of open coding, thematic analysis and other traditional qualitative methods with NVivo, with two forms of NLP—TM and Word2Vec—to analyze transcripts from interviews conducted in rural Belize querying men about their health needs. Leeson et al. (2019) found that all three approaches identified conceptually similar results and suggested that the NLP can be a useful adjunct to qualitative analysis. These comparative studies also revealed limitations of TM, which we discuss below.

### *Limitations of TM*

Some of the well-known limitations of TM are worth noting. First, the LDA, one of the most commonly used algorithms for fitting a topic model, requires a fixed number of topics when fitting the model. Over the past 10+ years, although TM has been used to classify documents in a variety of purposes, there are still problems with choosing the optimal number of topics. Krasnov and Sen (2019) argued that the main problem with TM is the lack of a stable metric of the quality of topics obtained during the construction of the topic model. Rather, the model construction relies heavily on the researchers' subjective judgment. Further, the LDA approach is also limited by the bag-of-words approach to the corpus. More specifically, it bases the results on the frequency of each word and what words occur in the document, but not where they occurred. As such, much of the contextual information, (e.g., where in the document the word appeared) is lost. While TM has demonstrated abundant success on long documents, the approach faces a great challenge with short texts (Hong & Davidson, 2010; Zhao et al., 2011). This is mainly due to the fact that only very limited word co-occurrence information is available in short and sparse texts, making it difficult to derive which words are more associated in a document (Wang & McCallum, 2006).

### *Purpose of the Study*

In this study, we extended the use of TM on data collected from focus groups with teachers who implemented the following math technology interventions with their students:

*From Here to There (FH2T)*, a game-based perceptual learning intervention; *Dragon Box 12+ (DragonBox)*, a widely used game-based technology application; and two versions of ASSISTments (*Immediate Feedback* and *Active Control*). Particularly, we compared the results of qualitative coding and TM. Unlike the textual data typically used in text mining, and as described above, the focus groups in this study involved dynamic human communications—i.e., teachers sharing thoughts and experiences in a meaningful way while taking in, processing, and responding to both the facilitator and other participants. In this process, actors (in this case, teachers) constantly take turns (and listen simultaneously) in roles as speaker and listener (Watzlawick et al., 1967; DeVito, 2016). As such, the patterns of communication constantly evolve, and the directions and depth of information exchanged during the focus group can drastically differ depending on the group of teachers and the skills of the facilitator. In this study, we examined if TM could extract patterns that were consistent with more qualitative analysis approaches.

Specific questions we explored in this study were:

1. What themes emerged from the qualitative coding approach and the TM approach?
2. What limitations does TM have regarding the analysis of the focus group data?
3. What recommendations do we have for other researchers who may attempt to use the TM on data collected from focus groups?

## Methods

This methodological study uses focus group data from a portion of teachers who participated in a larger efficacy study examining the impact of four instructional technologies across a school year (Decker-Woodrow et al., 2023). For the efficacy study, a total of 52 seventh-grade mathematics teachers and their 4200 students from 11 middle schools (10 in-person schools and one virtual academy) were recruited from a large, suburban district in the Southeastern United States in the summer of 2020. Since the larger study was conducted between September 2020 and April 2021, during the peak of the COVID-19 pandemic in the United States, the school district offered the students and their families a choice of learning modality (100% in-person classroom or 100% asynchronous virtual academy) for the 2020-2021 school year. Although the modality selection was intended to be for the full school year, families were allowed to change their selection. In the larger study, 60.6% of students initially selected in-person learning at school, and the remaining 39.4% selected virtual learning at home. In terms of the change of learning modality, only 13.5% of the students changed their initial learning modality choices at some point over the school year (Lee et al., 2023). Regardless of learning modality, all study activities were administered online during students' regular math classes (for in-person students) or as part of learning activities (for virtual

students). All participating students worked individually at their own pace using their devices. Students were randomly assigned to one of four mathematics technology interventions; two game-based interventions (FH2T and DragonBox) and two online problem set interventions (the ASSISTments platform was used for both conditions with variation in presence of hints and timing of feedback). Ultimately, 3271 students and 34 teachers implemented the interventions during the 2020-21 school year, in the peak of the COVID-19 pandemic. The larger study was approved by the Institutional Review Board at a university in the Northeastern United States.

## Technology Interventions

*From Here to There! (FH2T)*. FH2T (<https://graspablemath.com/projects/fh2t>) is a game-based application that implements theories of perceptual learning and embodied cognition to address cognitive and affective factors that lead to low proficiency in mathematics. FH2T was designed to promote fluency in algebraic notation by using notation from the beginning and directly connecting actions and operations to mathematical principles. The goal of each problem within FH2T is to encourage flexibility in notational transformations rather than solving for “x”.

*DragonBox 12+ (DragonBox)*. DragonBox (<https://dragonbox.com/products/algebra-12>) is a game-based application that introduces advanced algebraic concepts to students of ages 12 – 17 years. Problems begin with images rather than algebraic notation and then move gradually to use of algebraic notation, with the goal of solving for “x.” Therefore, a design principle of DragonBox is that students do not perceive the game as a mathematics game and do not explicitly make a connection to mathematical properties.

*Problem Sets with Hints and Immediate Feedback in ASSISTments (Immediate Feedback)*. The Immediate Feedback condition used an online homework system that provides feedback to students as they solve traditional textbook-style problems called ASSISTments (<https://new.assistments.org/>). In this condition, students were provided immediate feedback and on-demand hints as scaffolds during problem-solving.

*Problem Sets with Delayed Feedback (Active Control)*. The Active Control condition also used ASSISTments; however, feedback on performance was not given until after students completed the assignment. As such, this condition served as a control condition because it mimicked traditional problem set assignments that students typically complete.

## Participants

The data used in this study came from the teachers who participated in focus groups in the spring of the 2020 school

year, at the end of the implementation of the larger efficacy study mentioned above (Decker-Woodrow et al., 2023). Within the larger efficacy study (Decker-Woodrow et al., 2023), focus groups were conducted to better understand (a) a teacher's perspectives of students' reactions to the mathematics technologies, (b) challenges teachers and students encountered while using the mathematics technologies, (c) impacts of the various applications on student learning, (d) the unique impact of the pandemic in instructing students, and (e) suggestions for changes or improvements to the mathematics technologies. Out of the 34 teachers who implemented the study, 16 (47%) participated in one of the four focus group sessions. Participation was voluntary and an additional incentive was provided to the participants. Teaching experience for these participants ranged from 1–20 years, with the average being 9.73 years.

### Data Sources

With input from study directors and the study liaison, the external evaluator developed the focus group protocol. The study school liaison facilitated the virtual focus groups with teachers via Zoom, and an experienced qualitative researcher from the research team served as an independent notetaker. Using the study school liaison as the facilitator served two purposes:

(1) knowledge of the technology programs allowed for organic probing and questioning, and (2) existing relationships with the teachers quickly established a comfortable and safe space for participants to engage. All focus groups were recorded, and audio recordings were transcribed to facilitate coding and analysis.

It is important to note that TM and qualitative coding, and the comparison of the two methods, were applied on the data collected from the following selected questions:

1. What were some student reactions to working with FH2T, DragonBox, and ASSISTments?
2. What did your students enjoy or dislike the most about the technologies?
3. What were the most challenging or easy tasks your students encountered?
4. Were there any unique differences between the two learning modalities – i.e., students who used the technologies in-person versus virtually?

We selected these questions for inclusion in this study because all participating teachers shared their experiences and provided roughly the same amount of data across all participating teachers. (As the focus group data was also used for the larger efficacy study (Decker-Woodrow et al., 2023), the qualitative coding was conducted on all data collected.) We discuss the analytical approach next.

### Analytical Approach

The qualitative coding and the TM were conducted independently of the other by two researchers. From the raw

transcription data, each researcher produced the coding taxonomy. The researchers then independently summarized the findings from their analysis. Once each analysis process was completed, the two researchers then came together, comparing and discussing results to understand the similarities and differences between the two methods. This independent analysis allowed the researchers to determine (a) what parallels existed between the two methods regarding both analysis process and analysis outcomes, (b) how these two methods could potentially provide an opportunity for triangulation for rigor, and (c) how these two methods could be used together in future projects (e.g., independent parallel analysis to provide validity scoring, initial TM to identify nodes followed by qualitative analysis, initial qualitative analysis followed by TM to provide context around nodes).

**Emergent Qualitative Coding.** During the protocol development, the qualitative researcher and the focus group facilitator identified an initial multi-level node taxonomy from which an initial tagging document was created. Using an emergent theme approach, after a cursory reading of the first two transcribed focus groups, the researcher then updated the tagging taxonomy and tagging document, leaving opportunity for “open” tagging when new nodes were identified. Also included for all tags was a valence code, which allowed the researcher to identify statement sentiment, using “neutral” when no sentiment could be determined (e.g., “The kids loved playing the games” was tagged “strongly positive”). These valence codes were as follows:

1. Strongly positive
2. Somewhat positive
3. Neutral
4. Somewhat negative
5. Strongly negative

Once all focus group transcriptions were entered into the tagging documents and then loaded into the NVivo qualitative analysis software, the researcher used a matrix analysis to combine themes and valence codes to conduct a strengths analysis on each of the identified nodes.

**Topic Modeling.** The TM analysis was applied to the same data mentioned above. The responses from each teacher were treated as a single document (16 documents in total from four focus groups). We conducted analyses in R, an open-source software. Several R packages including tidytext, dplyr, ggplot2, stringr, tidyr, textstem, cleanNLP, tm, ldatuning, and topicmodels offer various computer algorithms (i.e., computer codes) to process text data instantaneously.

NLP analysis involves a series of “conditioning” and “preprocessing” tasks – i.e., necessary data preparation tasks for computers and computer programs to understand human language. The conditioning tasks involved removing the contractions, removing special characters, and changing all



letters to lower cases. The preprocessing tasks involved part of speech tagging and splitting it into a sequence of a word such as a word-level token. A token is a meaningful unit of text that we are interested in using for analysis. The word-level token is the most popular, but the token can be either words, characters, or subwords (n-gram characters) (Silge & Robinson, 2017). Further, stop words were removed (i.e., the words that do not necessarily provide useful meaning, such as “the, be, they, and, to, or that”). Lastly, nouns were extracted before conducting LDA TM.

The number of topics that exist in the data is unknown prior to the analysis. As such, the TM researcher obtained multiple results with a varying number of topics. The TM researcher examined each and determined the one that best represented the topic and words associated with a given topic.

## Results

### *Findings from the Qualitative Coding*

The emergent qualitative coding analyses generated five high-level themes that spoke to the student’s reactions to working with the various mathematics technologies:

1. Resentment among students regarding differences in technologies and hardware.
2. Experiencing frustration and persistence learning.
3. Making connections.
4. Remote versus in-person instruction.
5. Reactions specific to student population—i.e., special education (SPED) students and accelerated students.

*Resentment among Students Regarding Difference in Technologies and Hardware.* This theme speaks for resentment among students due to the different design of the technologies—e.g., game play versus worksheet. For instance, students using both of the problem set conditions in ASSISTments (with feedback) and control (no feedback) experienced a high level of frustration due to not having any game play in their learning like their peers. Teachers also reported that they heard students who were using the FH2T and DragonBox applications similarly expressing frustration and a desire to “switch applications.” Additionally, resentment among students also occurred between those using Chromebooks and those using touch screen tablets. Students using Chromebooks experienced some technical difficulties with the drag/drop functions that touch screen tablet-using students did not.

*Experiencing Frustration and Persistence Learning.* Teachers often talked about student frustration and observed persistence in learning due to the ways in which the technology interfaces

worked. For example, some students became frustrated when they solved a problem one way but the application would not allow them to move forward because they had not solved the problem in the way the application wanted. Some teachers made a point of emphasizing the benefits of “frustration” on student learning. That is, the learning parts of the game-based applications pushed students into more “flexible thinking” and “persistence” by forcing students to try multiple strategies or ways of solving equations or trying problems multiple times. This persistence not only helped progress them through the game, but helped students understand the concept of there being multiple ways to do math or solve equations.

*Making Connections.* Teachers talked about how learning with the game-based technologies provided their students with “aha-ha” moments. They reported hearing students within both their regular instructional time and study assignment time making connections between the study applications and their classroom instruction. Teachers shared how excited students were when they made the connections, reporting student comments such as, “This is showing us how to combine like terms,” “This is using distributive property,” and “I never thought of it like this.” The teachers believed the students were happy when they remembered things from the applications. At the end of each level, students were also asked a series of questions that connected the completed level activities to basic math knowledge and skills and teachers thought this was useful.

*Remote versus In-Person Instruction Reactions/Engagement.* In general, teachers talked about the challenges of remote learning (compared to in-person learning) in providing instructions and keeping students engaged. For example, teachers often mentioned how trouble-shooting logins, connectivity, or other technical issues took a great deal of extra work on the part of both the teachers and students for remote students. For example, downloading the DragonBox application on a personal device resulted in some additional work and challenges for students, teachers, and study stakeholders. Teachers also reported student engagement changing dramatically when students returned to in-person learning. Teachers reported that both teachers and students felt very overwhelmed with the remote learning and simply “shut down.” Teachers, though, did indicate that the students who did the study assignments while remote learning were much more prepared when they returned to school and that they understood more what was happening in class because of working in the games.

*Reactions Specific to Student Population.* The last theme identified addresses two specific student populations: special education students (SPED) and accelerated students. The

different applications posed some unique challenges for the SPED student population. (In the larger study, SPED students in self-contained classes were randomly assigned to either FH2T or DragonBox separately from the larger random assignment. These students were not assigned the ASSISTments application at teacher request.) Specifically, teachers reported their SPED students required much more explicit support and instruction with basic functions, such as logging in, navigating the game levels, and understanding why they had to “remember all the different ways to work through problems and equations.” These issues were predominantly experienced by students using the FH2T and DragonBox applications. In addition, while they liked the game-based programs more than the ASSISTments application, due to more pictures/visuals and less required reading, they did require more help solving the equations and struggled to complete levels independently.

Teachers also observed that the accelerated students were somewhat frustrated by having to start at the lowest level in each of the programs and wished the programs were adaptive. They indicated that a pre-test would help “level” a student so they could be challenged and would not have to start with the lowest-level material that they already knew.

### *Findings from the TM*

The TM generated three high-level themes that speak to the student’s reactions to working with FH2T, DragonBox, and ASSISTments: 1) Technology factors that influenced student engagement; 2) Non-technology factors that influenced student engagement; and 3) Student reactions toward mathematical instructions.

*Technology Factors That Influenced Student Engagement.* This theme illustrates teacher observations about the technology-related factors that influenced student reaction/engagement. This theme contained words such as tablet, technology, or virtual, error, issue, assistance, login, or skills. Teachers often cited that students who got the tablets showed higher engagement and they liked the puzzle aspects of the DragonBox game. Teachers observed some disconnect between the curriculum content (algebra expressions) and the technologies, especially with the accelerated students at the beginning. Teachers noted the challenges with working with SPED students. They often mentioned how difficult it was to monitor and provide instructions to students in the virtual environment and get the same level of feedback or engagement from them. Lastly, teachers mentioned that students wished to switch the technology they were using when they were stuck on a problem (the grass always looks greener on the other side of the fence). The top ~20 words associated with this theme were

tablet, time, person, week, program, password, error, matter, participation, activity, beginning, issue, couple, assistance, learning, list, login, process, random, reason, skill, technology, and virtual.

*Non-Technology Factors That Influenced Student Engagement.* Teachers observed some non-technology-related factors that also influenced student engagement. This theme was captured with words such as game, class, classroom, teacher, and feedback. In this theme, teachers often noted that the game-like or non-game-like instructions made a big difference in student engagement. They also noticed that students, particularly students in an in-person environment, enjoyed interacting with others (including both with other students and with teachers) when they worked on the instructional activities. The top ~20 words associated with this theme were dragon, game, class, device, grade, classroom, question, teacher, timer, test, email, feedback, language, engagement, amount, half, piece, and setting.

*Student Reactions Toward the Mathematical Instructions.* Teachers talked about students’ reactions toward the mathematical instructions provided via the learning tools. This theme was captured with words such as math, assessment, answer, step, question, method, and path. Teachers mentioned that some students were confused because of the lack of explicit instructions in the game-like tools (more evident among students who used DragonBox) or ways to navigate the tools to complete a lesson/chapter to advance to the next level. Once students advanced to the higher levels, students shared that the game became a little more challenging and engaging. While these observations were shared among students in the two game-based groups, teachers did not observe this reaction among students in the business-as-usual condition (i.e., problem sets in ASSISTments with delayed feedback). Lastly, teachers observed that those students who used FH2T experienced some challenges deconstructing numbers, getting the correct integer placement, or understanding the mathematical rules. The top ~20 words associated with this theme were student, time, assessment, people, Chromebook, math, school, activity, beginning, issue, message, answer, step, question, method, nature, path, phone, progress, reaction, struggle, and type.

### *Comparison of Findings*

As shown in [Table 1](#), the TM results show a high degree of agreement with the qualitative coding results. The biggest difference between the TM and qualitative coding was the organization/classification of themes at the higher level (e.g., five themes vs. three themes). At the lower level, both approaches identified many of the same sub themes and therefore, the results tell similar stories about the student’s

**Table 1.** Themes and Sub Themes Identified by Qualitative Coding and TM.

Qualitative Coding	TM
<p>Theme 1. Resentment among students</p> <p>Sub themes:</p> <ul style="list-style-type: none"> <li>- Game versus traditional problem sets (ASSISTments with hints and immediate feedback and ASSISTments control with delayed feedback)</li> <li>- Desire to “switch applications” (students with FH2T and DragonBox)</li> <li>- Tablet versus Chromebooks</li> </ul>	<p>Theme 1. Technology factors that influenced student engagement</p> <p>Sub themes:</p> <ul style="list-style-type: none"> <li>- Tablet versus no tablet</li> <li>- Virtual versus in person</li> <li>- A puzzle aspect of DragonBox</li> <li>- Disconnect between the content and tool (accelerated students)</li> <li>- Students with special needs</li> <li>- Wishing for different tools when they got stuck</li> </ul>
<p>Theme 2. Experiencing frustration and persistence learning</p> <p>Sub themes:</p> <ul style="list-style-type: none"> <li>- Frustrated students when they got stuck (students with FH2T)</li> <li>- The learning tools pushed students into more “flexible thinking” and “persistence”</li> <li>- The puzzle aspect of FH2T and DragonBox</li> </ul>	<p>Theme 2. Non-technology factors that influenced student engagement</p> <p>Sub themes:</p> <ul style="list-style-type: none"> <li>- Game versus non-game</li> <li>- Ability to communicate with other students (in person)</li> <li>- Instructional/classroom management</li> <li>- Student’s organizational skills</li> <li>- Graded or not</li> </ul>
<p>Theme 3. Making connections</p> <p>Sub themes:</p> <ul style="list-style-type: none"> <li>- The study applications and their classroom instruction (students with DragonBox)</li> <li>- Interactions with other students in the classroom</li> </ul>	<p>Theme 3. Reactions toward mathematical instructions</p> <p>Sub themes:</p> <ul style="list-style-type: none"> <li>- Students did not quite recognize what they were doing</li> <li>- Getting up to the higher level made it more challenging</li> <li>- Challenges with the deconstructed numbers</li> </ul>
<p>Theme 4. Remote versus in-person instruction reactions/engagement</p> <p>Sub themes:</p> <ul style="list-style-type: none"> <li>- Completeness of instructional activities/assignments</li> <li>- Engagement level</li> <li>- Graded or not</li> <li>- Hard to get remote students to participate overall</li> </ul>	
<p>Theme 5. Specific to student population</p> <p>Sub themes:</p> <ul style="list-style-type: none"> <li>- Explicit support needed with SPED, even the basic functions</li> <li>- More help needed to solve the equations due to more pictures/visuals</li> <li>- More vocal about the tool not being the game (accelerated students)</li> <li>- Accelerated students wanted to skip the lower-level problems</li> </ul>	

reactions to working with the different mathematics technologies.

In examining the coding and narrative findings from the two different methods (TM and qualitative coding), the two researchers found that the TM method was less effective in capturing the nuanced information that the qualitative coding was able to identify. For example, the qualitative coding was better able to provide details about students’ reactions specific to each of the learning tools and thus was able to provide more nuanced findings than TM. Similarly, the qualitative coding was better able to identify differences and information specific to student populations (i.e., students in SPED and accelerated students) than the TM methodology.

Regardless of the analytical approach, data consistently suggested the participating teachers appreciated being part of this study and felt that their students benefited overall from participating. They agreed that the more interactive and “game-like” the application was, the more that the students actively engaged. While teachers generally felt positively toward the applications used in the study, and the practice provided by each program, they did not feel they could specifically pinpoint how or if student knowledge and learning capability had grown. They did note though that some students struggled, for example, with “decomposing [the equation] in the way the program wanted them to do it.”

## Conclusion

In our study, we wanted to examine the feasibility of using TM compared to qualitative coding on data collected from teacher focus groups. Unlike other forms of text data such as literature, books, or essays where a type of information (e.g., information about actors, information about data source, information about findings) is organized within a section, data generated from focus groups reflected dynamic human communications (i.e., participants share thoughts and experiences in a meaningful way while taking in, processing, and responding to the moderator and other participants). The facilitator can also steer the participants back to the focus group questions or go along with the direction of the focus group discussions, depending on the research questions posed. In this study, we questioned if TM could extract patterns that were consistent with established qualitative analysis approaches within such a complex context. Particularly, we compared the results of the qualitative coding and TM and found a high degree of agreement.

We also uncovered strengths of qualitative coding and several weaknesses of TM. More specifically, TM was ineffective in capturing more nuanced information that the qualitative coding was able to identify. This can be explained by two factors: (1) the word level tokenization we used in the study, and (2) variations in the terminology teachers used to identify the different technologies. Because we use the word-level tokenization, n-gram “From Here to There” were treated in four tokens such as “From,” “Here,” “To,” and “There,” instead of one token, and lost the context. We recommend fixing n-grams into single words to keep them together during the data cleaning and tokenization processes. Variations with the differences in technology reference can also cause problems. For example, teachers referred to the DragonBox as DragonBox, Dragon, or Box. The computer will treat these as being different words, rather than as the same thing. To solve this, we recommend choosing a single spelling and replacing any other variants in your text with that version. This can be done using global find/replace functions within any analysis software.

The similarities and differences in findings presented here are from selected focus group questions (see section above on analysis approach). The selection of questions was guided by the amount of data available—e.g., if all teachers shared their experiences with the mathematics technologies. Therefore, because TM was only investigated on a subset of the focus group data, we cannot extrapolate the performance of TM to data from other questions.

Lastly, as [Leeson et al. \(2019\)](#) argued, TM can be used either before the qualitative coding to identify the primary coding themes, in parallel, or after the qualitative analysis to validate the accuracy of the coding process or identify potential unrecognized nodes. While having the TM analysis prior to qualitative coding is useful in that it can guide decision-making when determining nodes (i.e., themes) for the tagging and coding process, we found the use of TM post-qualitative analysis to be more useful. When TM occurs parallel to (or separate from) the qualitative coding

process it adds an element of validity to the qualitative analysis process.

Qualitative researchers, as they should, constantly question themselves while coding. This questioning takes the form of asking, “Is this coding capturing what the participant really said?” or “Am I reading more into this response because of my personal experiences in this area?” In the past two decades, qualitative researchers have worked to address the issues of reliability, validity, and significance by using coding methods that offer opportunities for quantification of data ([Lewis, 2009](#); [Hayashi et al., 2019](#); [Eldh et al., 2020](#)). [Vaismoradi et al. \(2013\)](#) delineate between “qualitative content analysis,” and “thematic analysis,” and indicated that content analysis offers more options to measure frequency, as a proxy for significance, as long as this is done with caution. What our team found through the process of using TM at the end of the qualitative analysis was both the quantification as well as the validation of the coding process verified the qualitative findings. If the findings of the TM process had been significantly different from the qualitative analysis, this could have served as an indicator to the qualitative analysis team to review their methodology and either make an argument for keeping the existing analysis or adjust the coding process and re-run the analysis. In this situation, the use of the TM process served to reassure the evaluation team that their methodology and analysis was tested, and the outcomes were affirmed by the TM results.

This study successfully extended the application of TM on data collected from focus groups. Used together, study results demonstrated that TM is a viable method for coding focus group data in a variety of ways. It can easily be used prior to qualitative analysis to identify nodes (i.e., themes) or in parallel or after qualitative analysis to identify strengths and weaknesses that might be inherent in the qualitative analysis. Of benefit to the qualitative coder is the rapid nature of technology that allows for faster coding using TM. In this study, we utilized an unsupervised machine learning algorithm to conduct the topic modeling analyses in R, an open-source software (i.e., no fees or licenses). It includes several packages for preparing the analysis data set (e.g., a series of conditioning and pre-processing tasks described). The analysis started with preparing a data set. TM was conducted with the iterative process of exploring and analyzing text data aided by computer codes and an LDA algorithm that can identify concepts, patterns, topics, keywords, and other attributes in the data. In our study, use of software to prepare the data set (described earlier) took about 15 minutes and calculation of TM took (including the iterative process) about 2 hours. Depending on the size of the vocabulary and complexity of the topic, the TM may take longer (i.e., longer than 2 hours). Learning a programming language can be challenging, but R has a large user community (roughly two million researchers), and many user-friendly resources are available such as [Silge & Robinson \(2017\)](#). The use of such technology can allow researchers to either keep the validated findings or pivot in their approach without a great



deal of time, effort, and cost loss. For the clients or recipients of the findings, the benefit is in knowing that the research team has done due diligence in presenting credible and useable findings.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A180401 to Worcester Polytechnic Institute. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

### ORCID iD

Lauren Decker-Woodrow  <https://orcid.org/0000-0001-9404-567X>

### References

- Baumer, E. P. S., Mimno, D., Guha, S., Quan, E., & Gay, G. K. (2017). Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology*, 68(6), 1397–1410. <https://doi.org/10.1002/asi.23786>
- Chen, X., Zou, D., Xie, H., & Cheng, G. (2021). Twenty years of personalized language learning: Topic modeling and knowledge mapping. *Educational Technology & Society*, 24(1), 205–222.
- Decker-Woodrow, L., Mason, C., Lee, J., Chan, J., Sales, A., Liu, A., & Tu, S. (2023). The impacts of three educational technologies on algebraic understanding in the context of COVID-19. *AERA Open*.
- DeVito, J. A. (2016). *The interpersonal communication book* (14th ed.). Pearson Education Limited.
- Eldh, A. C., Arestedt, L., & Bertero, C. (2020). Quotations in qualitative studies: Reflections on constituents, custom, and purpose. *International Journal of Qualitative Methods*, 19(1), 160940692096926. <https://doi.org/10.1177/1609406920969268>
- Hayashi, P. Jr., Abib, G., & Hoppen, N. (2019). Validity in qualitative research: A processual approach. *The Qualitative Report*, 24(1), 98–112. <https://nsuworks.nova.edu/tqr/vol24/iss1/8>
- Hong, L., & Davidson, B. D. (2010). Empirical study of topic modeling in twitter. *Association for computing machinery*. [https://snap.stanford.edu/soma2010/papers/soma2010\\_12\\_.pdf](https://snap.stanford.edu/soma2010/papers/soma2010_12_.pdf)
- Krasnov, F., & Sen, A. (2019). The number of topics optimization: Clustering approach. *Machine Learning and Knowledge Extraction*, 1(1), 416–426. <https://doi.org/10.3390/make1010025>
- Lee, J. E., Ottmar, E., Chan, J. Y. C., Booker, B., & Decker-Woodrow, L. (2023). Relations between learning modality choices and movement on student mathematics achievement during the COVID-19 pandemic: A mediation analysis.
- Leeson, W., Resnick, A., Alexander, D., & Rovers, J. (2019). Natural language processing (NLP) in qualitative public health research: A proof of concept study. *International Journal of Qualitative Methods*, 18(1), 160940691988702–160940691988709. <https://doi.org/10.1177/1609406919887021>
- Lewis, J. (2009). Redefining qualitative methods: Believability in the fifth moment. *International Journal of Qualitative Methods*, 8(2), 1–14. <https://doi.org/10.1177/160940690900800201>
- Li, Q., Zhou, X., Bostian, B., & Xu, D. (2021). How can we improve online learning at community colleges? Voices from online instructors and students. *Online Learning*, 25(3), 157–190. <https://doi.org/10.24059/olj.v25i3.2362>
- Mi, S., Lu, S., & Bi, H. (2020). Trends and foundations in research on students' conceptual understanding in science education: A method based on the structural topic model. *Journal of Baltic Science Education*, 19(4), 551–568. <https://doi.org/10.33225/jbse/20.19.551>
- Ozeran, M., & Martin, P. (2019). “Good night, good day, good luck:” Applying topic modeling to chat reference transcripts. *Information Technology and Libraries*, 38(2), 49–57. <https://doi.org/10.6017/ital.v38i2.10921>
- Peng, X., Han, C., Ouyang, F., & Liu, Z. (2020). Topic tracking model for analyzing student-generated posts in SPOC discussion forums. *International Journal of Educational Technology in Higher Education*, 17(1), 35. <https://doi.org/10.1186/s41239-020-00211-4>
- Silge, J., & Robinson, D. (2017). *Text mining with r: A tidy approach*. O'Reilly Media Inc.
- Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & Health Sciences*, 15(3), 398–405. <https://doi.org/10.1111/nhs.12048>
- Vijayan, R. (2021). Teaching and learning during the COVID-19 pandemic: A topic modeling study. *Education Sciences*, 11(7), 347. <https://doi.org/10.3390/educsci11070347>
- Wang, X., & McCallum, A. (2006). Topics over time: A non-Markov continuous-time model of topical trends. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 424–433. Philadelphia.
- Wang, Y., Bowers, A. J., & Fikus, D. J. (2017). Automated text data mining analysis of five decades of educational leadership research literature: Probabilistic topic modeling of EAQ articles from 1965 to 2014. *Educational Administration Quarterly*, 53(2), 289–323. <https://doi.org/10.1177/0013161x16660585>
- Watzlawick, P., Bavelas, J. B., & Jackson, D. D. (1967). *Pragmatics of human communication: A study of interactional patterns, pathologies, and paradoxes*. W.W. Norton & Company.
- Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., & Li, X. (2011). Comparing twitter and traditional media using topic models. In *Proceedings of the 33rd European conference on advances in information retrieval*, pp. 338–349. Springer.