Using Digital Technology to Address Confirmability and Scalability in Thematic Analysis of Participant-Provided Data

Chung Joo Chung
Kyungpook National University, South Korea, cjchung@knu.ac.kr

J. Patrick Biddix
The University of Tennessee, Knoxville, pbiddix@utk.edu

Han Woo Park
YeungNam University, South Korea, hanpark@ynu.ac.kr

Using Digital Technology to Address Confirmability and Scalability in Thematic Analysis of Participant-Provided Data

Abstract

This article presents a technique for analyzing large-scale qualitative data to address considerations for scalability and confirmability in thematic analysis of participant-provided data. A network approach provides a consistent means of coding that scales with the size of the dataset and is verifiable using standardized methods. This form of data analysis can be used with smaller data sources including interview transcripts as well as large data sources such as open-ended survey responses. A constructivist (inductive) approach is maintained and needed, however, to aid in interpretation of latent constructs. In this article, we provide both a conceptual overview of the co-word analysis method and a practical example.

Keywords

Qualitative Research, Network Analysis, Co-Word Analysis, Thematic Analysis, College Students, Technology

Creative Commons License

Acknowledgements

Using Digital Technology to Address Confirmability and Scalability in Thematic Analysis of Participant-Provided Data

Chung Joo Chung
Kyungpook National University, Daegu, South Korea

J. Patrick Biddix
The University of Tennessee, Knoxville, USA

Han Woo Park
Yeungnam University, Gyeongsan
North Gyeongsang, South Korea

This article presents a technique for analyzing large-scale qualitative data to address considerations for scalability and confirmability in thematic analysis of participant-provided data. A network approach provides a consistent means of coding that scales with the size of the dataset and is verifiable using standardized methods. This form of data analysis can be used with smaller data sources including interview transcripts as well as large data sources such as open-ended survey responses. A constructivist (inductive) approach is maintained and needed, however, to aid in interpretation of latent constructs. In this article, we provide both a conceptual overview of the co-word analysis method and a practical example. Keywords: Qualitative Research, Network Analysis, Co-Word Analysis, Thematic Analysis, College Students, Technology

Introduction

Researchers working with large scale qualitative data sources are challenged with representing reality and perspective within and across data sources, which becomes exponentially more difficult as data volume increases (Twining, Heller, Nussbaaum, & Tsai, 2017). Although thematic coding and analysis software has made investigation of larger datasets more manageable, we find confirmability (i.e., the degree to which the analysis process is influenced by the researcher) and scalability (i.e., maintaining the core tenets of constructivism as volume of data increases) to be persistent challenges. We have found this especially problematic when working to separate researcher perspective from results that are grounded in large-scale participant-provided data (Shenton, 2004, p. 72).

Although we prefer member checking as a technique when working with smaller scale participant-provided data or co-constructed data such as interviews, we have found contradictions in verifying data when asking a few members to authenticate or corroborate findings for very large populations. Taking a pragmatic approach, our search for more precise and repeatable results to address our concerns for confirmability and scalability led us to social network analysis techniques. In this article, we present an example case for how to conduct a social network-based analysis of data (referred to as co-word analysis) taken from participant-provided qualitative (short answer) responses to a survey to demonstrate how it can be used to enhance confirmability and increase scalability in qualitative data analyses.

Thematic Data Analysis Approaches

Thematic analysis is a qualitative data analysis process that involves identifying patterns and themes in qualitative data to represent findings (Braun & Clark, 2006). Miles, Huberman, and Saldaña (2015) elaborated that qualitative data analysis is a "continuous, interactive enterprise." The main distinction for thematic analysis stems from whether a researcher begins with a set of themes based on the existing literature or a theory (*a priori)* or if the themes emerge from the data (*in vivo*) (Boyatzis, 1998). Thematic analysis can be completed using analog or digital methods which further can be operationalized using manual or automated methods. Table 1 summarizes various ways researchers might code data using combinations of these approaches.

Table 1. *Operational techniques for coding qualitative data*

Analog | Manual ⇒ analog (print) materials ⇒ highlighters, sticky notes	**Analog | Automated** ⇒ not applicable
Digital | Manual ⇒ digital files ⇒ word processor tools ⇒ coding software, manual	**Digital | Automated** ⇒ digital files ⇒ coding software, automated ⇒ social network software

Analog | Manual and Digital | Manual approaches are human-reliant techniques for data analysis. In both cases, the researcher reviews data sources using a line-by-line or approach to segment excerpts of data (e.g., Chenail, 2012) to identify and highlight important words or phrases (codes) and then compiles codes to create themes (Merriam & Tisdell, 2015). In recent years, software has gotten more complex, adding the ability to create and displays thematic models to show links between data (Paulus, Lester, & Dempster, 2014). Manual approaches are labor intensive and allow for nuance and consideration of context. A major advantage is that researchers remain "close" to the data, and become more intentional participants in the co-creation of results (Neuendorf, 2016; Richards, 1998).

Digital | Automated and Analog | Automated approaches are computer-reliant techniques for data analysis. Both techniques approximate an *in vivo* or grounded approach to data analysis by seeking the most commonly occurring words in a dataset and providing descriptive statistics in terms of frequencies of use (Vlieger & Leydesdorff, 2011). Analog | Automated techniques are more conceptual than practical at this point, since automation or digital coding requires a digital data source. Digital approaches are highly systematic and efficient, but do not allow the researcher to connect context with meaning (Biddix, Park, & Wang, 2009; Richards, 1998). However, a major advantage is the ability to examine large amounts of data efficiently (Jung & Park, 2015).

Confirmability and Scalability Considerations

Operational techniques for analyzing textual data provide advantages and challenges related to both the process and conceptualization of data analysis. Primarily, there is concern for an inability to confirm results with most traditional coding techniques. Further, while

scalability can be substantially enhanced with the efficiency of digital and/or automated methods, this must be balanced against the potential loss of meaning and context (Twining et al., 2017). Additional discussion of these considerations follows.

The confirmability consideration. Qualitative researchers proposed confirmability as a concept to describe the extent to which results can be corroborated by others (Guba & Lincoln, 1981; Morse et al., 2002). Biddix (2018) described confirmability as a data analysis concern, verifiable when researchers include clear details about data analysis procedures such as how data sources became codes and codes became themes. This approach, sometimes called an audit trail (Merriam, & Tisdell, 2016) provides sufficient detail that another researcher could follow the same steps and arrive at similar results. Thomas (2006) recommended independent parallel coding as an alternative strategy that can be conducted during data analysis. The procedure involves two researchers independently coding a data source and comparing the two for congruence, consistency, and clarity. Put simply, confirmability is the extent to which the results can be achieved by others, ideally (although rarely possible) through replication (see also Elliott et al., 1999).

The scalability consideration. Scalability can be a challenge for qualitative data analysis. As the volume of qualitative data increases, maintaining the core tenets of constructivism becomes more difficult. Some computer assisted tools have been developed to aid in large-scale studies (Paulus, Lester, & Dempster, 2014), but often the result is data reduction at the cost of context. Raw and coded datasets can be analyzed to produce descriptive analysis including length and amount of data sources, word counts, frequency of codes, and prevalence metrics (Paulus, Evers, & de Jong, 2018). At a more complex level, some also can be used to automate coding and to produce network-like diagrams. One issue with this approach is the learning curve associated with digital tools – particularly complex functions like producing graphical representations of links in codes and themes (Belotto, 2018).

Co-Word Analysis of Thematic Data

Co-word analysis is a form of social network analysis (Danowski & Park, 2014; Hanneman & Riddle, 2005) in which the researcher identifies and models co-occurrences among words. Graphical representations aid in the interpretation of meaning (Leydesdorff & Vlieger, 2005). Researchers employ specialized software such as FullText.exe for English texts or KrKwic (Korean Key Words In Context) for Korean texts to search for words that appear, or co-occur, together (Park & Leydesdorff, 2004). Individual and co-occurring words are assigned descriptive statistics, which can be viewed in a variety of ways to identify patterns, or "recurring regularities" (Merriam & Tisdale, 2015, p. 206) in the data. A social network analysis feature is incorporated to visualize the connections in the data and more clearly identify emergent content, factors, and overall structure (Park, 2018; Park & Leydesdorff, 2013). Co-word analysis is concerned with finding shared meanings and interpretations among words with concepts in common (Doerfel, 1998) that can be mathematically valued. Researchers sometimes describe this as the "measurement of meaning" (Vlieger & Leydesdorff, 2011).

Leydesdorff and Welbers (2011) observed three non-exclusive capabilities of co-word analysis: inductive data analysis, large data analysis, and validation of content analysis using samples. Co-word analysis is regarded a blend between content analysis and factor analysis. As a form of content analysis, it is used to find meaning in documents from prominent words or phrases. As a form of factor analysis, it is used to detect correlations between words; the identification of latent concepts is also possible (Vlieger & Leydesdorff, 2011). Park and

colleagues (Biddix, Chung, & Park, 2015, 2016; Biddix, Park, & Wang, 2009; Park, 2012) proposed and demonstrated co-word analysis as an alternative operational techniques for coding thematic data. Because the techniques can handle small or very large volumes of data, this form of social network analysis has been useful in the study of "big data" (Lee & Park, 2019; Park & Leydesdorff, 2013).

Co-word analysis of thematic data begins by searching data using specialized data mining software. Units of analysis are texts, which can vary in size from sentences and paragraphs to sections and pages. While many thematic data analysis programs offer mining capabilities, such as producing descriptive frequencies for words using specified (*a priori*) and unspecified (*in vivo*) techniques, network analysis extends this step by identifying and then tracking relations between words. In particular, social media network tools also provide open data repository and sentiment analysis options for both qualitative and quantitative research (Smith, 2015). These relations, or links, are considered to "co-occur" which gives this form of inquiry its name, co-word or co-occurrence analysis (Chung & Park, 2010). Once the relevant unit of analysis is selected, the researcher decides how the word occurrences will be recorded (most/least frequent, weighting for moderate frequency, or using chi-square analysis). Co-word analysis programs typically utilize a chi-square analysis, which enables the researcher to calculate observed/expected values and assess the extent to which a word occurs above or below expectation (for more information, see Leydesdorff & Welbers, 2011).

Procedurally, the initial network analysis step fits a Digital | Automated categorization, but the secondary meaning making step is Manual. When content analysis is completed using a social network approach, the semantic or linguistic association between prominent words becomes the fundamental feature (Leydesdorff, 2001). However, since words are rarely spoken or written without context, meaningful analysis of text must also consider other associated words, phrases, or concepts (Neuendorf, 2016). As a result, co-word analysis is a multi-step process that first uncovers significant data points, identifies links between units, values those links and units, evaluates their position in the dataset, and then relies on the researcher to contextualize and interpret findings. This pairing of network analysis with thematic coding blends the efficiency of large-scale data automated analysis while allowing for a manual constructivist interpretation.

Co-Word Analysis Example

In the following example, we demonstrate how open-ended responses can be thematically analyzed using co-word analysis to address confirmability and scalability considerations. Data used for this example were derived from qualitative responses from a survey of college students enrolled in engineering programs at two large four-year research universities in the southeast Korea. The full questionnaire included demographic information, questions about mobile technology usage, ratings for use of mobile technology for specific activities, and open-ended questions about mobile technology use for academic projects. To demonstrate the use of the co-word analysis with thematic data for the purposes of this article, we selected one of the short-answer questions from the survey: *How does the use of mobile technology affect the academic performance of students?* We analyzed all 205 responses to this question. We chose this smaller set of data for illustrative purposes to show how the method can be used to improve validation. Following is the step-by-step analysis procedures. At the end of the section, we included results section to demonstrate how data are presented in text.

1. Prepare Data and Analyze Frequencies

First, we imported open-ended responses from the online survey platform as text files into KrKwic. A member of the research team initially screened the data, and removed blank or single word responses. During data cleaning and dataset preparation, researchers typically use a stop word or natural language processing dictionary. A stop word is a list of several commonly used words that co-word analysis software ignores such as articles (e.g., a, an, the) and conjunctions (e.g., and, but). After dataset preparation, we identified a listing of the top 40 word frequencies (words that appeared at least 3 times) as a group. We also specified automated data mining for stemming words. For example, learning also includes other versions of the word such as "learned." Table 2 displays the results.

Table 2. *Network metrics*

Words	Frequency	Ranking	nDegree	Ranking	nEigenvector	Ranking
LEARNING	65	1	14.936	2	37.587	2
USE	43	2	13.881	4	35.189	3
SEARCH	40	3	16.424	1	41.125	1
IMPROVEMENT	39	4	11.751	6	28.827	7
CLASS	36	5	13.964	3	33.365	4
EFFECTIVE	33	6	12.712	5	31.699	6
INEFFECTIVE	28	7	8.985	16	24.415	13
CONVENIENT	27	8	8.816	17	22.659	16
DEVICE	27	8	10.471	9	26.583	10
MOBILE	26	10	10.297	10	26.132	11
REFERENCE	26	10	11.373	7	31.97	5
UNKNOWINGNESS	25	12	6.717	25	16.618	26
INFORMATION	23	13	8.408	19	22.056	17
UNDERSTANDING	23	13	8.527	18	19.91	21
DISTRACTION	21	15	9.32	13	24.798	12
ACHIEVEMENT	17	16	9.738	11	27.755	9
NECESSARY	12	17	11.323	8	28.68	8
APPLICATION	10	18	7.853	22	18.648	22
ASSIGNMENT	10	18	6.518	26	16.332	27
CONTENT	10	18	6.003	30	14.454	30
FAST	10	18	9.668	12	24.167	14
STUDENT	10	18	4.523	33	10.657	34
POSSIBLE	9	23	9.262	14	21.129	20
SMARTPHONE	8	24	6.347	27	15.284	29
SOLUTION	8	24	7.451	23	18.019	24
VARIOUS	8	24	5.187	32	11.837	32
INTERNET	7	27	3.624	37	9.303	37
RELATED	7	27	9.046	15	23.144	15
USEFUL	7	27	6.2	29	17.833	25
FUNCTION	6	30	5.322	31	13.798	31
INTEREST	6	30	4.256	34	9.759	35
QUESTION	6	30	7.408	24	18.344	23
VIDEO	6	30	6.244	28	16.156	28
PERSONAL	5	34	3.537	39	10.723	33
MEANS	4	35	3.561	38	8.635	38
FILE	3	36	3.97	35	7.384	40
LAPTOP	3	36	3.15	40	7.416	39
MATERIAL	3	36	8.344	21	21.694	18
REVIEW	3	36	3.84	36	9.548	36
TEACHING	3	36	8.344	20	21.694	19

2. Generate Network Metrics

Next, a member of the team exported data from KrKwic into UCINET. For detailed procedures, refer to https://www.leydesdorff.net/software/fulltext/. The software was used to generate network metrics such as nDegree and nEigenvector (Borgatti et al., 2002), which are essential for understanding the importance of individual words as well as the overall structure of a network. Table 2 also displays these metrics. The degree centrality of each word is calculated based on the number of words adjacent to a given word in a text. nDegree stands for the normalized degree centrality that is the degree divided by the maximum possible degree. In a contrast to degree centrality, eigenvector value considers the centrality of words to which a given word is connected.

3. Create a Co-Word Matrix

Next a member of the team produced, a matrix, or listing of word intersections, to identify co-occurring patterns among words (Table 3).

Table 3. *Co-word matrix*

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1	LEARNING	65																																							
2	USE	24	43																																						
3	SEARCH	17	8	40																																					
4	IMPROVEMENT	14	3	17	39																																				
5	CLASS	12	20	10	3	36																																			
6	EFFECTIVE	27	25	7	7	13	33																																		
7	INEFFECTIVE	11	12	4	3	6	31	28																																	
8	CONVENIENT	13	2	12	4	3	5	1	27																																
9	DEVICE	11	21	2	1	9	8	6	1	27																															
10	MOBILE	10	19	2	1	8	7	5	1	33	26																														
11	REFERENCE	16	9	20	5	11	7	3	9	2	2	26																													
12	UNKNOWINGNE	7	4	7	7	2	3	2	2	4	4	1	25																												
13	INFORMATION	9	6	7	6	2	6	2	4	3	3	1	3	23																											
14	UNDERSTANDING	1	2	7	8	5	4	2	3	1	1	6	6	2	23																										
15	DISTRACTION	15	4	6	3	9	5	3	3	3	3	6	0	2	1	21																									
16	ACHIEVEMENT	17	6	7	5	7	8	5	2	0	0	8	1	3	0	4	17																								
17	NECESSARY	6	6	8	4	2	5	0	4	1	1	5	1	5	2	1	1	12																							
18	APPLICATION	3	5	2	0	6	4	2	1	3	3	3	1	0	1	1	2	1	10																						
19	ASSIGNMENT	3	1	2	2	4	3	2	1	4	4	1	0	2	1	1	2	0	2	10																					
20	CONTENT	0	1	3	4	4	2	2	0	0	0	2	5	1	10	0	1	1	0	0	10																				
21	FAST	3	3	6	3	4	2	1	2	2	2	1	2	5	2	1	2	3	0	1	1	10																			
22	STUDENT	4	4	0	0	2	2	0	2	4	4	0	0	1	1	0	0	1	0	0	0	0	10																		
23	POSSIBLE	3	2	4	0	3	1	0	1	2	2	5	0	3	4	1	0	3	2	2	2	1	2	9																	
24	SMARTPHONE	2	5	0	1	5	3	1	1	0	0	0	1	2	0	2	0	3	0	0	0	1	1	0	8																
25	SOLUTION	3	2	4	4	1	3	0	1	0	0	2	0	0	3	0	1	4	0	0	1	0	0	2	2	8															
26	VARIOUS	4	0	1	2	0	3	0	1	2	2	0	0	4	0	0	0	0	1	1	0	0	2	3	0	0	8														
27	INTERNET	2	2	1	0	5	1	1	0	0	0	0	3	0	1	0	1	0	0	1	0	0	1	0	0	1	0	7													
28	RELATED	3	3	5	4	3	1	0	2	1	1	4	0	0	3	3	0	2	0	1	1	0	0	2	0	2	0	0	7												
29	USEFUL	2	7	4	0	4	3	2	0	0	0	2	2	1	0	1	1	1	2	1	0	0	1	0	0	0	0	1	1	7											
30	FUNCTION	5	3	1	1	1	3	1	1	3	2	0	1	0	0	0	0	1	0	2	1	0	1	1	0	0	0	0	0	0	6										
31	INTEREST	3	1	0	1	4	0	0	1	0	0	0	0	3	1	0	0	0	0	2	2	1	0	1	0	0	0	1	0	0	0	6									
32	QUESTION	1	0	4	3	1	2	0	0	0	0	1	0	0	0	1	1	2	0	0	0	1	0	0	1	2	0	0	0	0	0	0	6								
33	VIDEO	4	0	4	4	0	0	0	0	0	0	2	0	0	3	2	0	0	1	0	0	0	0	0	0	1	2	0	0	0	0	0	1	6							
34	PERSONAL	3	2	1	1	2	2	1	0	1	1	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5						
35	MEANS	3	1	0	0	1	4	1	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	4					
36	FILE	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	1	2	0	0	0	0	0	1	0	0	0	0	0	0	3				
37	LAPTOP	0	2	1	1	0	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3			
38	MATERIAL	1	0	3	3	1	0	0	0	0	0	1	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	3		
39	REVIEW	0	0	0	1	2	0	0	1	1	1	0	0	0	0	2	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	
40	TEACHING	1	0	3	3	1	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0	3	0	3

Researchers must decide how to organize and interpret these results. For example, should only Understanding and Content be interpreted as indicative of the responses, since they are more highly correlated or should Unknowingness be added to help make sense of the results? This somewhat subjective interpretation of the statistical measures is aided by the use of network visualization software, which helps to further identify optimal patterns in the data.

In other words, network diagrams can be useful in making decisions about which co-words are most indicative of responses.

4. Create a Network Diagram

Next, we produced a visualization of the network to aid in the identification of co-occurring[1] words (Figure 1). This is done as a team to enable discussion of emergent clusters. Displaying the words in clusters aids in interpretation by showing relations and general patterns in the dataset. The visualization software is applied after data mining and is controlled by the user to graphically reconstruct the figure. Shapes and colors are used for visualization purposes. The size of the shapes indicate the prominence[2] of words. Words are clustered together to indicate frequently occurring groups. Lines between words signify frequently occurring words.
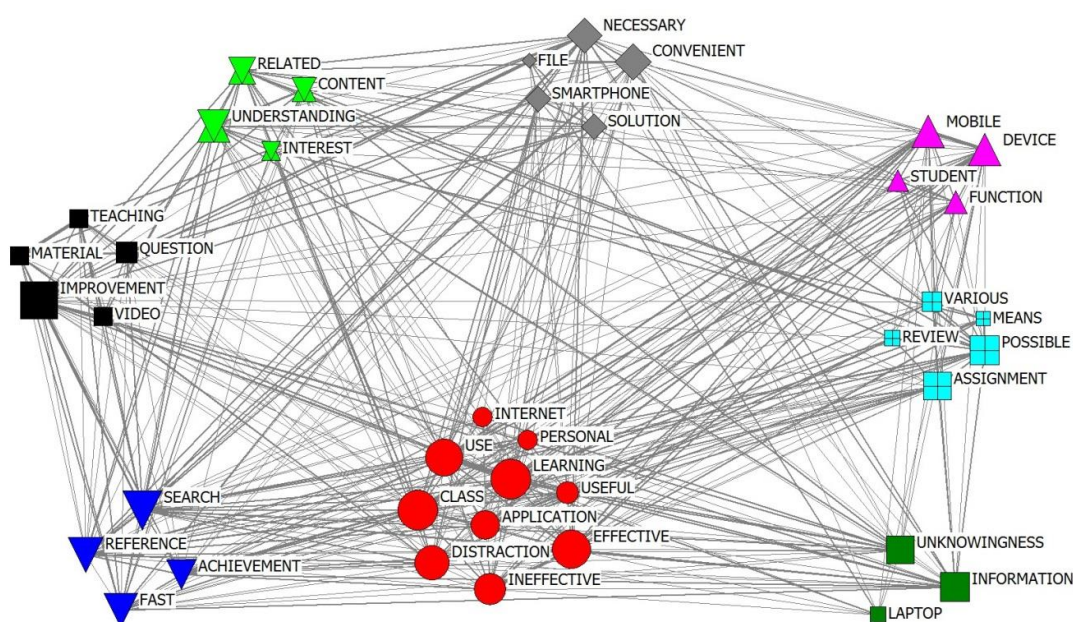
Figure 1. Network diagram

5. Interpret and Validate Data

To facilitate contextualization and presentation of data, researchers create summaries of the co-occurring words based on the most frequent responses (i.e., the most embedded clusters) (Rosen, Woelfel, Krikorian, & Barnett, 2003). Biddix, Chung, and Park (2015, 2016) demonstrated the use of this approach in a study of teaching and learning practices among

---

[1] This form of analysis is referred to CONCOR, which clusters network data by splitting blocks based upon the CONvergence of iterated CORrelations (CONCOR) with user control of the splits. Given an adjacency matrix, or a set of adjacency matrices for different relations, a correlation matrix can be formed by the following procedure. Form a profile vector for a vertex i by concatenating the ith row in every adjacency matrix; the i,jth element of the correlation matrix is the Pearson correlation coefficient of the profile vectors of i and j. This (square, symmetric) matrix is called the first correlation matrix. The procedure can be performed iteratively on the correlation matrix until convergence. Each entry is now 1 or -1. This matrix is used to split the data into two blocks such that members of the same block are positively correlated; members of different blocks are negatively correlated. CONCOR uses the above technique to split the initial data into two blocks. Successive splits are then applied to the separate blocks and are controlled by the user.

[2] The size of the concentric circles indicates the degree centrality among words. Prominence in this case refers to centrality, or how important certain words are to the overall structure of the network. The number of vertices adjacent to a given vertex in a symmetric graph is the degree of that vertex.

college students. Using this same technique, we created interpretable, generalized, and perhaps most importantly, contextualized responses as a team. We identified and included direct quotations to further evidence the alignment of generalized data analysis with actual data, as recommended by Braun and Clarke (2006). One member of the team completed the write-up by composing narrative themes. We provide additional details for this step in the following section, as this is critical in demonstrating how the technique can address scalability (dataset size is less relevant using this quasi-automated method) and confirmability (the process allows for checking and verifying that data themes and clusters match the text.

6a. Create and verify theme clusters. Although based on statistical measures and verifiable by review of frequency and correlation metrics, selecting the most "representative" clusters as themes is a constructivist-based "sensemaking" activity. We derived summary statements by viewing and interpreting the network metrics (Table 1), along with the social network diagram (Figure 1). This procedure is best completed separately by members of the research team and then compared as a validation check. As with identifying themes in traditional qualitative analysis, disagreements are discussed until consensus agreement on the summaries is reached. Two members of our research team followed this procedure, and generally agreed on the results after the initial round. To promote accuracy, we also used asked members of the population to review results (member-checking).

6b. Identify complementary quotations. After we selected representative phrases and verified them with a check against a randomly selected sample of the original data (intact responses), we selected complementary quotations to contextualize the results. Keeping good notes when creating the cluster summary is beneficial, since the validation procedure involves using software to perform a keyword search of the original data to locate and verify how the identified word clusters appear in context.

6c. Validate and contextualize findings. Graneheim and Lundman (2004) noted that "a text always involves multiple meanings and there is always some degree of interpretation when approaching a text" (p. 106). Although the initial list of co-words was statistically identified, the correlations among words may not reflect sentiments in the actual data. Further a concern is that some important clarifying words, such as "not" might be overlooked depending on the algorithm and specification of co-occurrence. However, the default specifications in most software is set to identify words co-occurring more than three times. So, a case where "not" might appear with "distraction" would be visible in the output. To address this issue and further and investigate the potential for misspecification, we returned to the initial data and reviewed responses using the listed phrases. This process is best considered iterative, meaning that there may be some trial-and-error in the checking procedures.

7. Organize and Write Results

Co-word analysis procedures yield several different types of output files that are used in data transfer, analysis, and in interpretation. For the purposes of presenting analysis and results, we typically provide the same three visuals we presented in this article: Network metrics (Table 2), Co-word matrix (Table 3), and the Network diagram (Figure 1). We find that a good organizational strategy for results is to use subsections for each open-ended response or research question (depending on the unit of analysis). Then, the most frequently co-occurring responses, as interpreted from both hierarchical and co-word analyses, should be displayed in phrases and then reworded to create summaries. The final presentation may also use conceptual themes derived by the researcher. Following is a brief example of a summary result section.

Sample Results Section

College students enrolled in engineering programs at two large four-year research universities in the southeast Korea were asked to respond to the following question: *How does the use of mobile technology affect the academic performance of students?* A total of 396 students from 10 classrooms responded to the survey and 205 provided open-ended responses. The total word count for all responses was 1,928.

Table 2 shows the most frequently occurring word was learning (65 times), followed by use (43), search (40), improvement (39), class (36), effective (33), ineffective (28), convenient (27), device (27), mobile (26), and reference (26). All other words were used 25 times or less. Figure 1 displays word groups. We used this visual to create the following summary statements to describe primary themes in the data.

- Learning and achievement is both effective and ineffective with the Internet (can be enhancing or distracting
- Videos of teaching material for questions and improvement
- Content related to interests improves understanding
- Searching for references, solutions, and files is convenient with a smartphone (in class)
- Using mobile devices makes various assignments and review possible
- Using smartphones in class is a distraction

As a final step, we grouped similar thematic statements, added explanatory narrative, and provided complementary quotations. Following is a listing of the narrative themes derived from the open-ended responses in this data. Keywords from the analysis are **bolded**.

Enhanced, but Distracted Learning

While students described numerous advantages of using mobile devices for learning related to convenience and the ability to enhance comprehension (even during class), they also mentioned the problems of distraction in nearly every example. Two students used the image of a double-edged sword to convey this dilemma. One noted, "I think it is a double-edged sword. It's easy to study with mobile devices, but there is a possibility that it will fall into a side path." Following are examples of additional quotations related to enhanced, but distracted learning.

> "It is **effective** if it is **used** correctly, but it is **ineffective** at the same time because there are many **uses unrelated** to the **class**."

> "It can help me **understand** in depth what I want to know but there is a concern that my attention may be **distracted**. I would recommend using them on more lessons and books."

> "You can **conveniently** find the **materials** you want, but they are too easily exposed to out-of-school **materials** and often interfere with your studies."

Convenient Access in Class and During Study

Students appreciated the convenient ability to locate material and access information, as needed. They described doing this both during study and in class. The primary motivator

was the ease of connection to information and the speed of finding an answer. A few students discussed determining credibility when describing this convenience.

> "It is easy to access **information** that you do not know before, which is a great help in studying."

> "First, it is easy to find the data you want anywhere, so you can easily access the **information** you need."

> "It makes you feel **convenient** in studying. Access to more information through associative **search**."

Enhances Learning

Beyond merely convenience, mobile devices were useful in helping students find alternative explanations for concepts, supported homework by allowing rapid access to information or videos online, and enhanced learning by providing the ability to explore concepts more deeply.

> "Enhance comprehension by acquiring information other than **class**."

> "**Useful** because you can find out words or terms you do not **understand** during **class** while **searching** the internet."

> "**Search** and **understand** necessary materials. You can **understand** the subjects better with **searching** for answers. Possible to investigate materials related to **class**."

> "It is important. **Internet**, wikis, **videos**, etc. are **used** to **understand** the concept and the programmes and it reduces the time for calculation."

Too Distracting

Several students discussed only the distractions of mobile device use. They believed that the distraction outweighed the benefits for themselves and for most students.

> "Unlike the past, I cannot concentrate on my time because there are much more apps and icons that **distract** me."

> "I do not think the **use** of **mobile devices** has a positive **effect** on my studies. It would be fine if it had only the **functions** to be **used**, but usually it would do a lot of **personal** things to do rather than lessons."

> "It is **effective** if it is **used** correctly, but it is **ineffective** at the same time because there are many **uses** unrelated to the **class**."

Final Considerations

The purpose of this article was to demonstrate a technique for enhancing confirmability and scalability of qualitative data, while maintaining the core values of constructivism. As it

becomes both easier to collect large volumes of qualitative data and more commonplace for participants to provide it online, interpretive techniques for analyzing large-scale open-ended or document-based data are needed. In this article, we demonstrated a solution using co-word analysis paired with network visualization.

One possible concern for this analysis is the time cost for the analysis – both in terms of learning the software and in performing and interpreting the automated analysis followed by human analysis. The software tools used for this analysis are commonly employed in many academic fields including communications, sociology, and increasingly, education. The basic functions demonstrated in this article for data mining and network visualization can be performed with little prior knowledge of network analysis. Our added challenge was in translating the Korean text to English. This step was an advanced function enabled by the software that would not be necessary for data that did not require translation. We also performed additional validation checks in the full dataset to ensure the accuracy of the translation.

We close by emphasizing the important role of the researcher for final interpretation of data, consistent with the goals of constructivism (Merriam & Tisdell, 2015). Similar to early users of qualitative and later mixed methods analysis techniques (Creswell, 2008), as the use of co-word analysis for qualitative data continues to develop, researchers are advised to provide readers with additional insight about the procedure. We acknowledge that for smaller datasets, such as the one used for this demonstration, this technique can more labor intensive in that both the software and the human element are needed for analysis. In larger datasets, however, the technique can considerably reduce the time cost of initial analysis (scalability) as well as verification process of ensuring accurate interpretation (confirmability).

References

Belotto, M. J. (2018). Data analysis methods for qualitative research: Managing the challenges of coding, interrater reliability, and thematic analysis. *The Qualitative Report, 23*(11), 2622-2633. https://nsuworks.nova.edu/tqr/vol23/iss11/2

Biddix, J. P. (2018). *Research methods and applications for student affairs*. San Francisco, CA: Jossey Bass.

Biddix, J. P., Chung, C., & Park, H. W. (2015). The hybrid shift: evidencing a student-driven restructuring of the college classroom. *Computers & Education, 80*, 162-175.

Biddix, J. P., Chung, C., & Park, H. W. (2016). Faculty use and perception of mobile information and communication technology (m-ICT) for teaching practices. *Innovations in Education and Teaching International, 53*(4), 375-387

Biddix, J. P., Park, H. W., & Wang, T. (2009). Co-word analysis of open-end answers from Chinese Internet users: An alternative content analysis method for qualitative research. *The Society for Humanities Studies in East Asia, 16*, 415-447.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *UCINET 6 for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development.* Thousand Oaks, CA: Sage.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.

Chenail, R. J. (2012). Conducting qualitative data analysis: Qualitative data analysis as a metaphoric process. *The Qualitative Report*, *17*(1), 248-253. https://nsuworks.nova.edu/tqr/vol17/iss1/13

Chung, C., & Park, H. W. (2010). Textual analysis of political messages: The inaugural addresses of two Korean presidents. *Social Science Information, 49*(2), 215-239.

Creswell, J. W. (2008). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Danowski, J. A., & Park, H. W. (2014). Arab Spring effects on meanings for Islamist Web terms and on Web hyperlink networks among Muslim-majority nations: A naturalistic field experiment. *Journal of Contemporary Eastern Asia, 13*(2), 15-39.

Doerfel, M. L. (1998). What constitutes semantic network analysis? A comparison of research and methodologies. *Connections*, *21*(2), 16-26.

Elliott, R., Fischer, C. T., & Rennie, D. L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *Research Studies in Psychology and Related Fields, 38*, 215-229.

Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today, 24*(2), 105-112.

Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation*. San Francisco, CA: Jossey-Bass.

Hanneman, R. A., & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California, Riverside; at http://faculty.ucr.edu/~hanneman/nettext/

Jung, K., & Park, H. W. (2015). A Semantic (TRIZ) network analysis of South Korea's "Open Public Data" policy. *Government Information Quarterly, 32*(3), 353-358.

Lee, Y.-J., & Park, J.-Y. (2019). Emerging gender issues in Korean online media: A temporal semantic network analysis approach. *Journal of Contemporary Eastern Asia, 18*(2), 118-141.

Leydesdorff, L. (2001). *A sociological theory of communication: The self-organization of the knowledge-based society*. Parkland, FL: Universal Publishers.

Leydesdorff, L., & Vlieger, L. (2005). Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment. *Journal of the American Society for Information Science and Technology, 57*(12), 1616-1628.

Leydesdorff, L., & Welbers, K. (2011). The semantic mapping of words and co-words in contexts. *Journal of Informetrics, 5*(3), 469-475.

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation* (4th ed.). San Francisco, CA: Jossey-Bass.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2015). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage.

Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods, 1*(2), 13–22.

Neuendorf, K. A. (2016). *The content analysis guidebook* (2nd ed.). Thousand Oaks, CA: Sage.

Park, H. W. (2012). Examining academic internet use using a combined method. *Quality & Quantity, 46*(1), 251-266.

Park, H. W. (2018). YouTubers' networking activities during the 2016 South Korea earthquake. *Quality & Quantity*, *52*(3), 1057-1068.

Park, H. W., & Leydesdorff, L. (2004). Understanding KrKwic: A computer program for the analysis of Korean text. *Journal of the Korean Data Analysis Society, 6*(5), 1377-1387.

Park, H. W., & Leydesdorff, L. (2013). Decomposing social and semantic networks in emerging "big data" research. *Journal of Informetrics, 7*, 756-765.

Paulus, T., Evers, J. C., & de Jong, F. (2018). Reflecting on the future of QDA software: Special issue of The Qualitative Report. *The Qualitative Report, 23*(13). 1-5. https://nsuworks.nova.edu/tqr/vol23/iss13/1

Paulus, T. M., Lester, J. M., & Dempster, P. G. (2014). *Digital tools for qualitative research*. Thousand Oaks, CA: Sage.

Richards, L. (1998). Closeness to data: The changing goals of qualitative handling. *Qualitative*

Health Research, 8(3), 319-328.

Rosen, D., Woelfel, J., Krikorian, D., & Barnett, G. (2003). Procedures for analyses of online communities. *Journal of Computer-Mediated Communication, 8*(4).

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information 22*, 63-75.

Smith, M. (2015) Catalyzing social media scholarship with open tools and data. *Journal of Contemporary Eastern Asia, 14*(2), 87-96.

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation, 27*(2), 237-246.

Twining, P., Heller, R. S., Nussbaaum, M., & Tsai, C. (2017). Some guidance on conducting and reporting qualitative studies. *Computers & Education, 106*, A1-A9.

Vlieger, E., & Leydesdorff, L. (2011). Content analysis and the measurement of meaning: The visualization of frames in collections of messages. *The Public Journal of Semiotics, 3*(1), 28-50.

Author Note

Chung Joo Chung is an Associate Professor in the Department of Journalism and Mass Communication at Kyungpook National University. He conducts research on new media and technology, social networks, data science, AI from the perspective of social science. He has published articles in prestigious journals, such as *Journal of Computer-Mediated Communication*, *Scientometrics*, *Computers and Education*, *Social Science Computer Review*, *Technological Forecasting and Social Change*, and *Telecommunications Policy*. He also contributes to start-up activities and communities. Please direct correspondence to cjchung@knu.ac.kr.

J. Patrick Biddix is a Professor and Associate Director of the Postsecondary Education Research Center (PERC) at the University of Tennessee. His research and teaching focus on research design and assessment, student engagement and involvement, and postsecondary outcomes. Dr. Biddix is the author of *Research Methods and Applications for Student Affairs* (Jossey-Bass, 2018) and co-authored the 2nd editions of *Assessment in Student Affairs* (Jossey-Bass, 2016) and *Frameworks for Assessing Learning and Development Outcomes 2.0* (CAS, 2020). In 2015, he received a Fulbright Scholar Award to study college student communication and technology use in Montreal, Canada. Please direct correspondence to pbiddix@utk.edu.

Han Woo Park (Corresponding Author) is a Professor in the Dept. of Media & Communication, Interdisciplinary Graduate Programs of Digital Convergence Business and East Asian Cultural Studies, and Founders of Cyber Emotions Research Institute (at YeungNam University) and WATEF (World Association for Triple Helix & Future Strategy Studies), South Korea. He was a pioneer in network science of open and big data in the early 2000s (often called *Webometrics*) when he used to work for Royal Netherlands Academy and lead the World Class University project. He has published more than 100 articles in *SSCI* Journals. He is currently Chief Editors for *Journal of Contemporary Eastern Asia* and *Quality & Quantity*. Several publications were included in top 10 list of downloads and citations. He has been co-awarded the best paper in EPI-SCImago in 2016 and included in the list of core-candidates of the Derek de Solla Price Memorial Medal in 2017 and 2019. Please direct correspondence to hanpark@ynu.ac.kr.

Article Citation

Chung, C. J., Biddix, J. P., & Park, H. W. (2020). Using digital technology to address confirmability and scalability in thematic analysis of participant-provided data. *The Qualitative Report, 25*(9), 3298-3311. https://nsuworks.nova.edu/tqr/vol25/iss9/7