01 Aug 2022

Frame of Reference Training for Content Analysis with Structured Teams (Fort-Cast): A Framework for Content Analysis of Open-Ended Survey Questions using Multidisciplinary Coders

Sandra J. Galura

Kristin A. Horan

Joy Parchment

Daleen Penoyer

et. al. For a complete list of authors, see https://scholarsmine.mst.edu/psysci_facwork/320

Recommended Citation

Galura, S. J., Horan, K. A., Parchment, J., Penoyer, D., Schlotzhauer, A., Dye, K., & Hill, E. (2022). Frame of Reference Training for Content Analysis with Structured Teams (Fort-Cast): A Framework for Content Analysis of Open-Ended Survey Questions using Multidisciplinary Coders. *Research in Nursing and Health, 45*(4), pp. 477-487. Wiley.
The definitive version is available at https://doi.org/10.1002/nur.22227

Received: 1 November 2021 | Accepted: 26 March 2022

DOI: 10.1002/nur.22227

FOCUS ON RESEARCH METHODS

RESEARCH IN NURSING & HEALTH WILEY

Frame of reference training for content analysis with structured teams (FORT-CAST): A framework for content analysis of open-ended survey questions using multidisciplinary coders

Sandra J. Galura[1] | **Kristin A. Horan**[2] | **Joy Parchment**[1] | **Daleen Penoyer**[3] | **Ann Schlotzhauer**[2] | **Kenzie Dye**[2] | **Emily Hill**[2]

[1]College of Nursing, University of Central Florida, Orlando, Florida, USA

[2]Psychology Department, University of Central Florida, Orlando, Florida, USA

[3]Center for Nursing Research and Advanced Nursing Practice, Orlando Health, Orlando, Florida, USA

Correspondence

Sandra J. Galura, College of Nursing, University of Central Florida, 12201 Research Pkwy, Suite 300, Orlando, FL 32826, USA.
Email: sandra.galura@ucf.edu

Abstract

In the context of a global pandemic, the need for reliable analysis of qualitative data in healthcare has never been more pressing. Open-ended questions are a feasible way for both researchers and organizational stakeholders to gain deeper insight into complex situations when timely research is needed. However, the interpretation of brief, textual responses can prove problematic. Both manual and automated/semiautomated methods of coding qualitative data have been associated with errors and costly temporal delays. Data obtained from the qualitative analysis of open-ended questions have been questioned for lacking robust insights. The present article introduces an innovative, manual, team-based method of analyzing responses to open-ended survey questions. This method was developed and implemented at the outset of the COVID-19 pandemic to understand the needs of nurses and their perceptions of organizational strategies that were implemented to address pandemic-related challenges. This framework utilizes a dedicated project management structure, general purpose software for data collection and analysis, frame-of-reference training designed for an interdisciplinary team of coders, and data analysis procedures that align with qualitative content analysis procedures. In concert, these techniques empower researchd team members with varying backgrounds and disparate levels of experience to provide unique human insights to data analysis procedures, refine the coding process, and support the abstraction of meaningful themes that were used to prioritize organizational strategies and further support nurses as the pandemic progressed.

KEYWORDS

frame-of-reference training, open-ended questions, qualitative analysis

1 | INTRODUCTION

Open-ended survey questions can provide crucial information and elicit diverse responses not confined to the limits of standardized, closed-ended survey questions (Riiskjær et al., 2012). Specific to employee surveys, open-ended questions provide the individual employee with an opportunity to express immediate needs and truthful opinions while providing organizations with valuable feedback (Zull, 2016; Gilles et al., 2017; Singer & Couper, 2017). Despite this value, analyzing the free-text comments generated by these questions can be burdensome and costly, especially when working with large data sets and within condensed project time frame (Etz et al., 2018). As a result of these challenges, critical findings may be lost. While a variety of computer-aided text analysis methods have been introduced in an effort to efficiently analyze formidable sets of open-ended comments, these methods may not be feasible for all projects. This article introduces a manual, team-based framework, *frame-of-reference training for content analysis with structured teams* (FORT-CAST), a feasible method of completing a moderately large-scale evaluation of open-ended comments. This framework was applied to an organizational survey of nurses' needs and perceptions of organizational strategies during the COVID-19 pandemic.

2 | BACKGROUND

2.1 | Open-ended questions

Open-ended questions are frequently used in qualitative and mixed methods research studies to obtain a holistic and comprehensive understanding of complex situations, an understanding that would be otherwise limited with the use of closed-ended survey questions (Zull, 2016). Open-ended questions have been used to facilitate information gathering when the range of possible answers to a question is unknown, to avoid a long list of response options in closed-ended questions, to avoid the use of directive questions that might steer participants towards a specific answer, and to capture information on substantive issues not addressed by traditional closed-ended items (Borg & Zuell, 2012; Zull, 2016). Open-ended questions have also been used to motivate respondents to provide truthful opinions, express criticisms freely, and share personal experiences related to topics that may be of a sensitive nature (Zull, 2016). For these reasons, open-ended questions have experienced a resurgence in use in recent years by health services researchers and organizations due to their ability to elicit key information about complex healthcare and employee-related issues.

Despite this resurgence, limitations associated with open-ended questions have led to a debate among experts about their value and the quality of the insights they can provide. The content provided in free-text responses is often restricted by either space on paper surveys or word-count limitations in electronic surveys. As a result of these limitations, experts argue the resulting data likely lacks the richness necessary to provide substantial insights (LaDonna et al., 2018). Specific to employee surveys, responses have been largely found to be negative in tone, with the likelihood of writing comments inversely related to the employee's job satisfaction (Borg & Zuell, 2012; Gilles et al., 2017). Additionally, experts acknowledge responses to these questions are rarely analyzed using rigorous qualitative methods, with some further suggesting analysis may be more quantitative than qualitative if the analysis is based on the frequency of key words (Stoneman et al., 2013; Gilles et al., 2017; LaDonna et al., 2018).

2.2 | Methods of analyzing responses to open-ended questions

2.2.1 | Automated coding

Recent advances in technology have mitigated the burdens associated with manual coding allowing researchers to automatically code textual responses to open-ended questions. Compared to manual processes, automated coding processes are objective, less costly, and more efficient, and thus more valuable for the timelier analysis of large data sets (greater than 1500 comments; Etz et al., 2018; He & Schonlau, 2020; Schonlau & Couper, 2016). Defined as the use of a statistical learning model to predict the code of free-text responses, automated coding is accomplished through a series of steps which begins with the manual coding of a randomly selected subset of the data. This manually coded "training data set" is used to convert the answer texts of all open-ended responses into numerical variables (n-gram variables) which contain counts of indicators addressing how often a given word ($n = 1$) or sequence of words ($n = 2$) occurs in the narrative text. The resultant model is tested using the training data set, with the final algorithm predicting the most likely codes (Schonlau & Coup, 2016).

2.2.2 | Semi-automated coding

Compared to fully automated coding, semi-automated coding consists of a few additional steps. Similar to automated coding, the process begins with manually coding a subset of the larger data set (training data set). Text answers for both the training and test data are converted to the same numeric variables (n-gram). The resulting statistical learning algorithm is applied to the training data, which then estimates the probability for each text falling into a coding category. Next, a threshold of probability for automated coding is determined and computed based on expected accuracy. If the estimated categorization probability threshold exceeds the computed threshold, automated coding is accepted; if the estimated probability is lower than the computed threshold, manual coding is required (Schonlau & Couper, 2016). While both fully automated and semi-automated coding methods have been introduced in an effort to efficiently analyze formidable sets of open-ended comments, these methods have been found to have higher error rates than human coders (He & Schonlau, 2021), automatically classifying only 58% of textual data accurately (Schonlau & Couper, 2016).

2.2.3 | Computer-assisted content analysis

This method of coding involves categorizing content based on a content-analysis dictionary where coding rules are formulated as lists of words (Zull, 2016). Instead of defining codes, phrases are defined that are indicative of a particular category. When the words or phrases appear in a text-response, a corresponding code is assigned. To accomplish this method of analysis, suitable software must be available to support the development of the dictionary. In addition, participant responses must be available in machine-readable form, which is readily accomplished with today's widespread use of web-based surveys. Following completion of coding using this method, it is recommended that the quality of coding be tested manually using a subset of the larger data set. While sufficient time and effort on the front-end is required to develop and validate the content analysis dictionary (list of words), the use of such programs (e.g., MAXDictio, TextQuest, WORDSTAT) allows for the quick and reliable coding of large volumes of data (Popping, 2015; Zull, 2016).

2.2.4 | Manual data analysis methods

Manual coding, defined as the use of human coders to classify free-text responses, is facilitated using a coding manual or codebooks that are drafted based on preliminary coding categories and updated throughout the coding process as new codes are identified (Cascio et al., 2019). Codebooks provide definitions and can include example quotes to promote consistency in coding across team members. As compared with technology-aided methods, manual coding processes are viewed as more subjective, time-consuming, and expensive. To facilitate more timely manual coding, the use of multiple coders, or a multidisciplinary coding team, has been identified as beneficial. In addition, involving multidisciplinary coders in data analysis has been found to be a valuable strategy to enhance credibility, defined as the confidence in the truthfulness of the research conclusions (Lincoln & Guba, 1985), as multiple coder perspectives inform the data analysis process (Cascio et al., 2019; Church et al., 2019). However, challenges associated with using a team approach to coding may include varying skill levels of individual coders (Berends & Johnston, 2005) and varying knowledge of the phenomenon under study, both of which can be addressed with training and supportive strategies. Strategies reported in the literature to address these challenges include orienting team members to the nature, purpose, and objectives of the study, the structure of the data, the rationale for the approach to coding, and the final study deliverables (Berends & Johnston, 2005; Giesen & Roeser, 2020; White et al., 2012). In addition, incorporating practice coding sessions to build coding team members' confidence with the coding process has been identified as a key training strategy (Berends & Johnston, 2005; Giesen & Roeser, 2020; White et al., 2012). The use of codebooks and training manuals has been consistently identified as a critical strategy to support consistency among coding team members (Cascio et al., 2019; Giesen & Roeser, 2020; White et al., 2012). Detailed codebooks that identify codes, code definitions, and any conditions on when to double-code or to not apply codes, help foster understanding of the coding scheme and facilitate intercoder reliability (Giesen & Roeser, 2020).

Neither automated, semi-automated, computer-assisted, nor manual methods of coding are without error; however, the sources of error vary by method. Semi-automated systems have been found to have higher error rates, only classifying up to 58% of textual data correctly (Schonlau & Couper, 2016) with the primary error reported as generalization error, or out-of-the-sample error, a measure of how accurately a machine algorithm can detect meaningful patterns in data (He & Schonlau, 2021). Conversely, errors in manual coding are largely due to human error, unclear coding manuals, or misunderstanding the meaning of textual responses. Nevertheless, the rigor of all methods is challenged when coding more lengthy, ambiguous text responses (He & Schonlau, 2021).

2.3 | Addressing nurses' needs during the pandemic: A case study in the application of the FORT-CAST method

During the first wave of the COVID-19 pandemic, concerns expressed by organizational leaders over the well-being of their employees prompted the distribution of a web-based survey designed to assess nurses' needs and their perceptions of organizational strategies implemented in response to the pandemic. In an effort to provide employees with a voice to freely and honestly express their opinions and to provide organizational leaders with feedback about the effectiveness of organizational strategies, four open-ended questions were added to the survey: (1) what is one thing the organization could continue to do to support you, (2) what is one thing the organization could start doing to better support you, (3) what is one thing the organization could stop doing to better support you, and (4) is there anything else you would like to tell us about your workplace's response to the COVID-19 outbreak? Given the timely need for research, the resource intensive nature of the early stages of the pandemic brought about by the surge in patients and redeployment of staff, and the size and scope of the organization (large, urban, multi-campus system), administration of a survey incorporating open-ended questions was deemed the most feasible and appropriate method to obtain the desired data to achieve the study purpose in a timely manner. In addition, incorporating a team of multidisciplinary coders provided an opportunity to gain unique insights through the human interpretation of the data versus interpretation accomplished using computer algorithms. For these reasons, an innovative, team-based framework was developed to accomplish the manual coding and analysis of over 900 free text responses (less than 150 words) to the four open-ended survey questions.

2.4 | Innovation: FORT-CAST framework

FORT-CAST was developed to address the identified challenges associated with manual coding and the use of a team of multidisciplinary coders. Development of the framework consisted of four steps: defining the project management structure, creating a master data file for data entry and analysis, providing structured training for a team of multidisciplinary coders, and conducting data analysis that aligns with standard content analysis. Content analysis was defined as a systematic approach to the coding and categorizing of textual data to determine word trends, patterns, frequency of use, their relationships, and structures of communication (Gbrich, 2007; Mayring, 2000; Pope et al., 2006; Vaismoradi et al., 2013).

2.4.1 | Step 1: Define the project management structure—roles/responsibilities

First, a dedicated management structure (Giesen & Roeser, 2020) was developed by defining key project roles and responsibilities. Key roles included the project lead, lead trainer, primary coder, coding team, and the senior reviewers. The project lead was responsible for the overall management of all data analysis procedures, including managing the master spreadsheet used for data analysis, communicating with all members of the project team, and ensuring project timelines were met. The lead trainer's primary responsibility was to provide targeted training to all members of the coding team. Both the project lead and lead trainer collaboratively shared the role of primary coder throughout the data analysis process. Both represented different disciplines and were responsible for reviewing all coding completed by members of the coding team. The coding team consisted of members representing the disciplines of nursing and industrial and organizational psychology and were responsible for coding subsections of the overall data set under the guidance and support of the primary coders. Finally, senior reviewers, consisting of nurse leaders from the clinical study site and academic researchers representing the disciplines of nursing and psychology, were tasked with several key steps in the data analysis process, including developing the initial categorization scheme, validating all coding completed by the coding team, abstracting themes, and collaborating to achieve consensus of the final study results.

2.4.2 | Step 2: Create a master file for data entry and analysis

To facilitate data analysis, a master EXCEL spreadsheet was created. The EXCEL file consisted of several tabs/sheets. The first tab included three elements: (1) general instructions for the protocol to be followed by members of the coding team, (2) the initial categorization scheme established by the primary coders (project lead/lead trainer) and senior reviewers, which included initial parent (main category) and child (related subcategory) codes and their definitions as column headers, and (3) an example of a coded response (see Table 1). The second tab included individual coder assignments, or the subsections of data to be evaluated by each coder, represented as EXCEL line numbers. The third tab detailed all steps of the data analysis process. The final four tabs, one for each of the four open-ended questions, included the initial categorization scheme at the top of each page to serve as a guiding codebook to facilitate consistency among coders during the data analysis process, and all related participant responses for the specific question downloaded from the Qualtrics survey management platform.

Specific to our study, the initial categorization scheme represented on the master EXCEL spreadsheet was developed collaboratively. Each of the primary coders and senior reviewers was provided with the same sub-sample of the larger data set (20 randomly selected responses downloaded from the Qualtrics survey platform) and allotted time for independent review to develop their own individual interpretations of the data. Following independent review time, the primary coders and senior reviewers convened as a team to collaboratively discuss their individual interpretations of the data sub-sample. The group consensus was achieved on the initial categorization scheme with parent codes (main categories) and child codes (related subcategories) derived directly from the textual data and described on a manifest level or low degree of interpretation and abstraction (Lindgren et al., 2020). For example, “Communication” occurred frequently within the first 20 responses and, as such, it was identified as a preliminary parent (main) coding category and a definition of the category (“formal and informal messages sent within an organization; can vary in content, frequency, and level of transparency”) was added to the EXCEL spreadsheet to guide the coding team. Child codes (subcategories related to parent codes) were identified simultaneously with the identification of parent codes and were incorporated into the initial categorization scheme. Examples of child codes for the parent code of “Communication” were “communication content,” “communication frequency,” and communication transparency.” Child codes were similarly defined. Additional cell column headings incorporated into the spreadsheet provided space for coders to document illustrative quotes that reflected a specific code, add any general coding comments or questions for the primary coders to review, and make recommendations for new codes identified during the coding process.

2.4.3 | Step 3: Provide structured frame-of-reference training

To address the varying skill levels and contextual knowledge of the nursing discipline among the multidisciplinary members of the coding team, frame FORT was integrated into the framework as an innovative training method. Traditional coding approaches have emphasized the use of multiple coders categorizing the same response as a way to identify problems with reliability and validity (Bachiochi & Weiner, 2002). When inter-rater agreement is high, there can be more confidence that the interpretations by raters are

TABLE 1 Data management EXCEL spreadsheet: Team protocol with embedded code book example

Evaluation of nurse needs during the corona virus pandemic—team protocol

General instructions for coder: you will be coding responses to 4 open-ended questions. Each worksheet tab (e.g., Q1, Q2, Q3, & Q4) represents participant responses to one of the four open-ended questions. Each line on the worksheet represents a single participant's response. After reading the response, select the code that represents the major theme identified in the participant's response and indicate in the corresponding cell the frequency in which the code appears. The frequency most likely will always be "1" (see examples below). Some codes have clear subcategories that were identified. These codes are represented by a parent code (main category) and a child code (subcategory). ***For responses that contain a parent/child code theme, you will score a "'1" for both the parent and corresponding child code***. Should there be a portion of the participant's response that reflects the code in a very comprehensive, well-stated way, please provide a direct quote in the column titled " Illustrative Quote" and include the corresponding code that the quote reflects. Should you identify a code that is not represented in the current worksheet, please provide a note in the column "New Codes"

Coder Name:

Respondent ID	Qualitative Question #1: What is one thing the organization could continue to do to support you during the COVID-19 pandemic?	Parent Code 1: Communication	Child Code 1a: Communication content	Child Code 1b: Communication frequency	Child Code 1c: Communication transparency	Illustrative quote(s)	General comment/ questions	New Codes
		Definition: formal and informal messages sent within an organization can vary in the content, frequency, and level of transparency.	Definition: Communication content	Definition: Communication frequency	Definition: Communication transparency	Definition: Reponses that either reflect a code/theme in a very comprehensive, well-stated way and/or a well-summarized call to action or a new call to action (telling them to do something new). Example: don't email everyday; continue daily updates.	Definition: General comments or questions related to the choice of coding categories.	Definition: New categories identified in participant responses that do not align with any a priori constructs.
1	Participant Response	1		1				
2	Participant Response	1		1				

more reliable rather than influenced by the idiosyncrasies of each coder's categorization. While there are several advantages to this approach, coding large data sets in a small-time frame may prove cumbersome when multiple coders must be assigned to each participant response. Alternative approaches can minimize the influence of bias while maximizing efficiency, by providing structure for single-coder rating systems in the analysis of large-scale qualitative data. FORT, a tool used in industrial and organizational psychology, minimizes bias in the interpretation and rating of work-related behavior (Schleicher et al., 2002; see Table 2). In keeping with its name, the training serves to create a consistent "frame of reference," or categorization schema for coding, among all coders. In other words, by following FORT, coders should be "on the same page." In addition, FORT focuses on coding categories that should exist regardless of the idiosyncrasies of each coder and helps the coder learn to recognize the category and separate it out from other noise. For example, when coders participate in a FORT session before coding for purposes such as performance appraisal, rating accuracy and agreement tends to be higher (Roch et al., 2012). FORT can be conducted easily using the following steps: (1) instruct coders about biases that can influence coding, (2) define categories of interest for coders and provide examples of category-related behavior, and (3) provide coders with practice coding and feedback on their practice coding (Schleicher et al., 2002).

Specific to our study, the lead trainer developed and conducted a 2-h FORT session consisting of three key steps. First, in a virtual meeting, all coders observed a presentation on rater biases. Second, the coding team was oriented to the initial categorization scheme or preliminary coding categories. Finally, coders were provided with two participant responses and given practice time to code. Codes were discussed among the group and the lead trainer gave feedback. This training provided the benefits of promoting reliability and validity of ratings in a single-coder approach and provided graduate student coders with theoretical and practical knowledge regarding rater bias and FORT procedures.

To further support content analysis skill acquisition, each coder was provided with a subsample of 25 participant responses to all four open-ended questions from the larger data set for practice. Each coder was allotted 5 days to complete practice coding of the subsample. The project lead then reviewed each coder's responses, providing additional feedback and guidance after which time coders were cleared to proceed with coding their assigned data subset.

2.4.4 | Step 4: Conduct data analysis

Quantitative content analysis has been identified as the classical method of analyzing responses to open-ended questions with coders coding the responses to open-ended questions on the basis of a predefined coding scheme (code book; Zull, 2016). Specific to our study, given the range of participant responses (up to 150 words), data were analyzed using an inductive content analysis, moving from the specific to general by first coding the textual responses (main/parent; subcategories/child), and then abstracting themes that cut across coding categories (Elo & Kyngas, 2008; Vaismoradi et al., 2013).

Each member of the coding team analyzed approximately 750 free-text responses to the four open-ended questions and was allotted 2 weeks to complete coding of their assigned data subset. During the coding process, coders reviewed each response and either assigned an existing parent and/or child code or made recommendations for the development of a new code. For example, initial codes largely focused on well-known issues encountered at the onset of the pandemic related to personal protective equipment and redeployment. As coding progressed, members of the team identified unanticipated codes such as "Positive Affirmations" defined as positive phrases or statements that described a particular outcome within the organization, and "Employee Voice" defined as soliciting input from employees and acting based on input. Coders also provided illustrative quotes or specific text responses that helped to further illustrate and support conclusions drawn from the coded data.

TABLE 2 Rater biases summarized in FORT training

Bias	Definition	Rater instructions
Similar to me error	Raters tend to rate responses or participants who are similar to them more positively, even when it is not related to the construct of interest. *Example:* If a participant mentions their job title, a rater who holds the same job title might rate their responses more positive, even if the construct applies to all job titles.	Try not to pay more attention to the words of someone who sounds like you, writes like you would write, or has a similar work history.
First impression and recency errors	Raters tend to weight performance of the first and last ratings differently than ratings in the middle. This could be due to memory effects, or due to the tendency to interpret ratings relative to each other (i.e., "this one is better than the last one").	Try to treat each new data point as individual, as if it is the very first text you've seen.
Rater fatigue	If raters code a large number of responses in one sitting, they will be more likely to tune into non-construct-relevant details as they become mentally fatigued.	Try not to do a "marathon" coding session. Instead, try to break it up into smaller codes.

Abbreviation: FORT, frame-of-reference training.

This coding process not only led to the identification of additional parent code categories and child code subcategories, but also quantified their frequency of occurrence resulting in a rank-order of coding categories for each open-ended question. For example, in response to what the organization could continue to do to support nurses, parent and child codes specific to communication rank-ordered the highest in frequency. This prioritized, rank-ordering provided organizational leaders with clear direction on where to prioritize additional strategies.

Following completion of coding the entire data set, the primary coders (project lead and lead trainer) collaboratively reviewed all coded responses. Once all coding was reviewed, the primary coders began the initial abstraction process by identifying more abstract themes cutting across code categories. For example, specific to the parent code of "communication" and child codes of "content, frequency, and transparency," the more condensed categories were interpreted as "organization" and "pandemic" with the final abstracted theme interpreted as "knowledge of current state" (see Figure 1). This progression of abstraction/interpretation was then visually represented in a causal chain display that illustrated the progression of the analysis from the descriptive analysis of free-text responses (manifest content) to the more interpretive, abstraction of the underlying meaning (latent content; Lindgren et al., 2020).

Following completion of all descriptive (coding) and interpretive (abstraction) data analysis procedures, the senior reviewers were presented with the completed, master EXCEL which contained all individually coded participant responses, the final coding categories, and a preliminary draft of the abstracted themes as illustrated in the causal chain displays. After allowing time for individual review, a meeting was convened with the project lead, lead trainer, and senior reviewers, and following collaborative review and discussion, consensus was reached on the final abstracted themes (Elo et al., 2014). The final causal chain displays, which helped illustrate the connections between coding categories and the final abstracted themes, were used to develop infographics, or illustrations that used graphic elements to present findings in a visual manner (Infographic, n.d.), summary reports, and recommendations for presentation to organizational stakeholders (Elo et al., 2014; Giesen & Roeser, 2020). Application of the FORT-CAST framework in our study accomplished the manual content analysis of a moderately large data set of open-ended questions within 5 weeks from the time of receipt of collected data to presentation of the final results.

2.4.5 | Case study application summary

The FORT-CAST framework was the optimal approach for this study given the aim of the study, the sensitive nature of the data being collected, the tight project timeline, and the challenges faced in collecting data within the context of a worldwide pandemic. By understanding nurses' needs, we hoped to obtain actionable data from participants that would inform the continuation, development, and/or cessation of organizational strategies that would be used to support nurses' well-being as the pandemic progressed. These needs and perceptions would likely vary by campus. The most feasible way of collecting data broadly was by providing a readily available, web-based survey that employees could complete at their convenience. Other methods of qualitative data collection, such as observations (which primarily serve to gain insight into actual behavior rather than opinions) did not align with the aim of our study. One-on-one interviews, semi-structured interviews, and focus groups, while valuable for gaining insights into personal experiences as well as opinions, were not feasible given the context of the pandemic, the urgent need for research, and the limited resources.

Similarly, other qualitative methods of analysis did not align with the primary aim of our study. Ethnography, the analysis of patterns in behavior and thoughts of study participants, seeks to gain a deeper understanding of a specific culture. Phenomenology seeks to describe the meaning of experiences through the identification of themes. The grounded theory seeks to develop theories that are

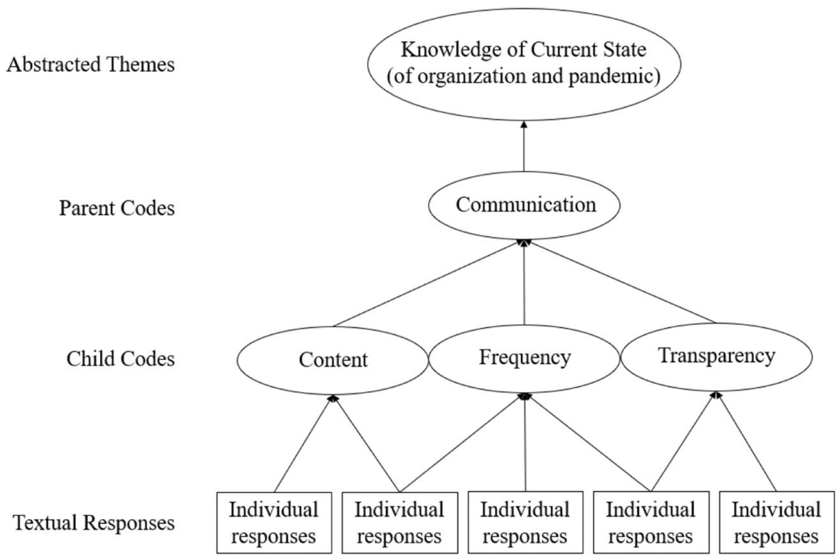

FIGURE 1 Visual representation of organization of coding and abstraction. This is a simplification of the coding and abstraction framework. Individual responses could be coded directly to one or more parent codes (bypassing child codes), a parent code and one or more of its child codes, multiple child codes subsumed under the same of different parent codes, and so forth. Similarly, an abstracted theme might be based on one or more parent codes.

grounded in data obtained from real-world observations (Polit & Beck, 2017). Again, these methods did not align with the study's aim to describe and quantify nurses' needs and perceptions of organizational strategies and make valid inferences of the analyzed data to the study context to provide actionable guidance for organizational leaders (Krippendorff, 2018).

3 | DISCUSSION

The benefits of the FORT-CAST framework lie in its supportive project management structure, methodological appropriateness for achieving specific study purposes, training to address challenges associated with inter-rater reliability, unique insights provided by multidisciplinary human coders, and processes incorporated into content analysis procedures to ensure the credibility (Lincoln & Guba, 1985) of the abstracted themes. Content analysis has been regarded by some quantitative researchers as too simplistic for statistical analysis and by some qualitative researchers as not sufficiently qualitative in nature. Yet, content analysis has been widely used in nursing research, offering several major benefits including the ability to make valid inferences from data to their context, the content-sensitive nature of the method, and the method's concern with meanings, intentions, consequences, and context (Elo & Kyngas, 2008).

The manual method of content analysis incorporated into our framework may prove to be the most feasible option for many studies. The equipment necessary to accomplish automated/semi-automated machine learning methods of coding, as well as the computer software and user skill necessary to develop coding dictionaries for use in the computer-assisted content analysis may be prohibitive. Further, though technology-aided methods of content analysis have mitigated the time-consuming nature of manual analysis methods, these methods have been found to have higher error rates than human coders (He & Schonlau, 2021, Schonlau & Couper, 2016). To address these potential sources of human error associated with manual coding (He & Schonlau, 2021, Schonlau & Couper, 2016), FORT-CAST incorporates a clearly defined codebook (the initial categorization scheme of coding categories and their definitions) on each tab of the EXCEL file to guide coders in their data analysis, frame-of-reference training, and practice coding sessions with feedback from primary coders. These strategies serve to ensure clarity and understanding among multidisciplinary coders and enhance inter-rater reliability of the members of the coding team.

Having multidisciplinary representation on the human coding team proves to be advantageous throughout the coding process. Research has demonstrated the beneficial use of multidisciplinary coders as mitigating the subjectivity associated with manual qualitative processes and, by doing so, enhancing the quality and credibility (Lincoln & Guba, 1985) of analyses (Berends & Johnston, 2005; Church et al., 2019). Having data reviewed and interpreted from a human perspective rather than a computer algorithm, incorporating diverse perspectives, and providing opportunities throughout the review process to discuss coding disagreements and/or new code recommendations helps to provide unique insights and further refine the coding system throughout the data analysis process.

Finally, understanding that responses to open-ended questions are rarely analyzed using standard qualitative methods (Etz et al., 2018; Gilles et al., 2017; Stoneman et al., 2013), we ensured our framework aligned with the defined phases of content analysis (preparation, organization, reporting) and incorporated strategies throughout all phases to enhance credibility (Lincoln & Guba, 1985; see Table 3). For example, credibility (Lincoln & Guba, 1985) is enhanced in the organizing phase (where coding categories and subcategories are formulated) through the collaborative development of the initial coding categorization scheme by the project lead, lead trainer, and senior reviewers, the ability of the coding team to recommended the consideration of new codes as the coding process evolves, the levels of review provided by the primary coders throughout training and following the completion of all coding activities, and the final review of all coding categories and abstracted themes by the senior reviewers.

Application of the FORT-CAST framework is not without its challenges and limitations. While the use of general-purpose software (e.g., EXCEL) and manual data analysis processes may prove more feasible in many studies, the time associated with the completion of all steps of the framework likely exceeds the more efficient timeframes that could be achieved with automated/semi-automated data analysis methods. In the case study described above, analysis was conducted in a 5-week timeframe and results were presented promptly enough to provide valuable information to leaders as they dealt with ongoing pandemic surges. Nonetheless, the FORT-CAST framework may not be appropriate in more urgent research situations.

A recognized expert qualitative methodologist not familiar with the research study was unavailable to serve on the research team in our case study; this was a limitation of our application (Giesen & Roeser, 2020). When possible, such a methodologist should be included in the FORT-CAST framework. This role should be responsible for the development of the initial coding categorization scheme and providing support and feedback to the coding team. While these tasks can be accomplished by the project lead and lead trainer as they were in our study, they might have been better supported by an expert. Inclusion of a recognized expert in qualitative methodology further enhances the credibility (Lincoln & Guba, 1985) of the data and reported findings in a FORT-CAST framework.

The credibility (Lincoln & Guba, 1985) of study findings in a FORT-CAST framework is established through frequent collaborative discussions among the project lead, lead trainer, and senior reviewers. While consensus of the final study results can be achieved and validated as representing reality through these collaborative discussions, credibility (Lincoln & Guba, 1985) can be further enhanced with the use of a more robust audit trail. Internal audits, conducted by members of the research team at various intervals to assess inter-rater reliability, and an external audit conducted by an expert in qualitative data analysis not associated with the study can

TABLE 3 FORT-CAST framework

Content analysis phase (Elo & Kyngas, 2008)	Management structure (Giesen & Roeser, 2020)	Components
Preparation	Project lead Lead trainer Senior reviewers	**1.** Survey development; development of open-ended questions **2.** Selection of the unit of analysis **3.** Development of management structure
Organizing	Project lead Lead trainer Multidisciplinary coders	**4.** Raw data download (Project Lead): Survey data downloaded into data management platform (EXCEL) **5.** Data immersion (Primary Researchers): (sub-sample) – independent team member review **6.** Initial constructs (Primary Researchers): Collaboration to reach consensus on a priori constructs to guide open coding. **7.** Data Management Spreadsheet/Master Codebook Development (Project Lead/Lead Trainer **8.** Training (Lead Trainer): FORT training (coding team) **9.** Training (Project Lead): Practice Coding (coding team) **10.** Open Coding (Multidisciplinary coders): coding of full data set **11.** Open Coding Review(Project Lead/Lead Trainer): Final review of all coded responses **12.** Rank Order of Coding Categories (Project Lead/Lead Trainer): Final rank ordering of identified coding categories. **13.** Preliminary Abstraction (Project Lead/Lead Trainer) coded data abstracted into subcategories and higher-order categories.
Reporting	Project lead Lead trainer Senior reviewers	**14.** Independent Review (Primary Researchers): Review of final coded responses, rank-ordered coding categories, preliminary abstracted themes **15.** Collaborative Discussion (Primary Researchers): Collaborative review and discussion of preliminary abstracted themes. **16.** Final Consensus (Primary Researchers): Consensus on final abstracted themes. **17.** Final Deliverables: Graphic displays, infographics, summary reports, recommendations

Abbreviation: FORT-CAST, frame-of-reference training for content analysis with structured teams.

be incorporated into the framework to further enhance credibility (Lincoln & Guba, 1985; White et al., 2012).

Application of FORT-CAST does not require a quantitative assessment of inter-coder reliability (ICR); this may be viewed by some as a limitation. A commonly cited reason for performing an ICR assessment in qualitative research is to assess the rigor of the coding categorization scheme and the robustness of the coding process. However, ICR is not universally accepted as beneficial and is seen by some researchers as contradicting the interpretive nature of qualitative research. Likewise, others view ICR as unnecessary, arguing that alternative concepts such as trustworthiness, defined as research findings that are "worthy" of attention (Lincoln & Guba, 1985), may be just as acceptable (O'Connor & Joffe, 2020). While the minimum number of coders necessary to complete an ICR assessment may be met in studies utilizing FORT-CAST, the skills and resources necessary to complete an ICR assessment may not be available. As an alternative, the training, practice coding, and ongoing review of coding embedded in the FORT-CAST framework provides a more feasible strategy to address credibility (Lincoln & Guba, 1985) and increase intercoder reliability before coding of the full data set commences and throughout the coding process.

Experts have questioned the ability of open-ended questions to generate robust insight (LaDonna et al., 2018). Researchers should carefully consider their study's aims and research design in weighing the advantages and disadvantages of open-ended questions. For example, in our case study, the decision to include these questions was intentional. We were purposeful in the design of the questions, ensuring they were clearly stated, used as an adjunct to the primary quantitative survey, and aligned appropriately with the primary study aim to gain an understanding of nurses' needs and their perceptions of organizational strategies to guide the ongoing organizational response. We found application of the FORT-CAST framework and its embedded qualitative content analysis processes yielded valuable insights by prioritizing nurses' needs through the quantification and rank-ordering of descriptive coding categories. The insights gained provided valuable feedback to leaders as to which organizational initiatives and strategies were beneficial and critically necessary to effectively support the well-being of their employees as the pandemic continued. Though not the primary aim of the analysis, the nature of the data allowed for further abstraction and interpretation as patterns threaded throughout coding categories emerged, enabling a deeper understanding of the textual responses provided by the study participants. Thus, open-ended questions were appropriate in our case study. This may not be the case for all research; study conceptualization should be an intentional and carefully evaluated process.

4 | CONCLUSION

The use of open-ended questions to explore complex phenomena in a variety of settings is on the rise, yet the textual responses to these questions are rarely analyzed using standard qualitative methods (Gilles et al., 2017; Stoneman et al., 2013). The FORT-CAST framework serves as a feasible approach to accomplish the manual analysis of moderately large data sets of responses to open-ended questions and could be easily applied in complex circumstances that warrant timely research. Involving multidisciplinary coders in the data analysis process proves valuable in gaining unique insights that help further refine codes and abstracted themes. Though the FORT-CAST framework aligns with the defined phases of qualitative content analysis (preparation, organization, reporting), experts express caution in the ability of open-ended questions to provide robust qualitative insights. Careful consideration should be given to the conceptualization of a study, ensuring the analysis of data occurs in tandem with a primary survey, the research question is focused and appropriate, and data analysis procedures offer robust insights to the phenomenon of interest (LaDonna et al., 2018).

AUTHOR CONTRIBUTIONS

Sandra Galura: Substantial contributions to the analysis and interpretation of the work; drafting and revising the work for intellectual content, final approval of the published version, and agreement to be accountable for all aspects of the work. **Kristin Horan**, **Joy Parchment, and Daleen Penoyer**: Substantial contribution to the conception and design of the work, analysis, and interpretation of data; drafting and revising the work for intellectual content, final approval of the published version, and agreement to be accountable for all aspects of the work. **Ann Schlotzhauer, Kenzie Dye, and Emily Hill**: Substantial contribution to the analysis and interpretation of data, final approval of the published version, and agreement to be accountable for all aspects of the work.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Not applicable.

ORCID

Sandra J. Galura http://orcid.org/0000-0003-0621-3974
Ann Schlotzhauer http://orcid.org/0000-0002-7400-3613

REFERENCES

Bachiochi, P. D., & Weiner, S. P. (2002). Qualitative data collection and analysis. In Rogelberg, S. G.(Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 161–183). Blackwell Publishing.

Berends, L., & Johnston, J. (2005). Using multiple coders to enhance qualitative analysis: The case of interviews with consumers of drug treatment. *Addiction Research & Theory*, *13*(4), 373–381. https://doi.org/10.1080/16066350500102237

Borg, I., & Zuell, C. (2012). Write-in comments in employee surveys. *International Journal of Manpower*, *33*(2), 206–220. https://doi.org/10.1108/01437721211225453

Cascio, M. A., Lee, E., Vaudrin, N., & Freedman, D. A. (2019). A team-based approach to open coding: Considerations for creating intercoder consensus. *Field Methods*, *31*(2), 116–130. https://doi.org/10.1177/2F1525822X19838237

Church, S. P., Dunn, M., & Prokopy, L. S. (2019). Benefits to qualitative data quality with multiple coders: Two case studies in multi-coder data analysis. *Journal of Rural Social Sciences*, *34*(1), 1–14. https://egrove.olemiss.edu/jrss/vol34/iss1/2

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative content analysis: A focus on trustworthiness. *SAGE Open*, *4*(1), 1–10. https://doi.org/10.1177/2158244014522633

Elo, S., & Kyngas, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

Etz, R. S., Gonzalez, M. M., Eden, A. R., & Winship, J. (2018). Rapid sense making: A feasible, efficient approach for analyzing large data sets of open-ended comments. *International Journal of Qualitative Methods*, *17*(1), 1–10. https://doi.org/10.1177/1609406918765509

Gbrich, C. (2007). *Qualitative data analysis: An introduction*. Sage Publications.

Giesen, L., & Roeser, A. (2020). Structuring a team-based approach to coding qualitative data. *International Journal of Qualitative Methods*, *19*, 1–7. https://doi.org/10.1177/2F1609406920968700

Gilles, I., Mayer, M., Courvoisier, N., & Peytremann-Bridevaux, I. (2017). Joint analyses of open comments and quantitative data: Added value in a job satisfaction survey of hospital professionals. *PLoS One*, *12*(3), e0173950. https://doi.org/10.1371/journal.pone.0173950

He, Z., & Schonlau, M. (2020). Automatic coding of text answers to open-ended questions: Should you double code the training data? *Social Science Computer Review*, *38*(6), 754–765. https://doi.org/10.1177/0894439319846622

He, Z., & Schonlau, M. (2021). Coding text answers to open-ended questions: Human coders and statistical learning algorithms make similar mistakes. *Methods, Data, Online First*, *15*(1), 103–120. https://doi.org/10.12758/MDA.2020.10

Infographic. (n.d.). *Merriam-Webster's online dictionary*. Retrieved January 3, 2022, from https://www.merriam-webster.com/dictionary/infographic

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage Publications.

LaDonna, K. A., Taylor, T., & Lingard, L. (2018). Why open-ended survey questions are unlikely to support rigorous qualitative insights. *Academic Medicine*, *93*(3), 347–349. https://doi.org/10.1097/ACM.0000000000002088

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.

Lindgren, B. -M., Lundman, B., & Graneheim, U. H. (2020). Abstraction and interpretation during the qualitative content analysis process. *International Journal of Nursing Studies*, *108*, 108. https://doi.org/10.1016/j.ijnurstu.2020.103632

Mayring, P. (2000). Qualitative content analysis. *Forum: Qualitative Social Research*, *1*(2). https://doi.org/10.17169/fqs-1.2.1089

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, *19*, 1–13. https://doi.org/10.1177/2F1609406919899220

Polit, D. F., & Beck, T. (2017). Nursing research: Generating and assessing evidence for nursing practice (10th ed.). Wolters Kluwer.

Pope, C., Ziebland, S., & Mays, N. (2006). Analysing qualitative data. In Pope, C. & Mays, N.(Eds), *Qualitative research in health care* (pp. 63–81). Blackwell Publishing.

Popping, R. (2015). Analyzing open-ended questions by means of text analysis procedures. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *128*(1), 23–39. https://doi.org/10.1177/0759106315597389

Riiskjær, E., Ammentorp, J., & Kofoed, P.-E. (2012). The value of open-ended questions in surveys on patient experience: Number of comments and perceived usefulness from a hospital perspective. *International Journal for Quality in Health Care*, *24*(5), 509–516. https://doi.org/10.1093/intqhc/mzs039

Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational & Organizational Psychology*, *85*(2), 370–395. https://doi.org/10.1111/j.2044-8325.2011.02045.x

Schleicher, D. J., Mayes, B. T., Day, D. V., & Riggio, R. E. (2002). A new frame for frame-of-reference training: Enhancing the construct validity of assessment centers. *Journal of Applied Psychology*, *87*(4), 735–746. https://doi.org/10.1037/0021-9010.87.4.735

Schonlau, M., & Couper, M. P. (2016). Semi-automated categorization of open-ended questions. *Survey Research Methods*, *10*(2), 143–152. https://doi.org/10.18148/srm/2016.v10i2.6213

Singer, E., & Couper, M. (2017). Some methodological uses of responses to open questions and other verbatim comments in quantitative surveys. *Methods, Data, Analysis*, *11*(2), 115–134. https://doi.org/10.12758/mda.2017.01

Stoneman, P., Sturgis, P., & Allum, N. (2013). Exploring public discourses about emerging technologies through statistical clustering of open-ended survey questions. *Public Understanding of Science*, *22*(7), 850–868. https://doi.org/10.1177/0963662512441569

Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing and Health Sciences*, *15*(3), 398–404. https://doi.org/10.1111/nhs.12048

White, D. E., Oelke, N. D., & Friesen, S. (2012). Management of a large qualitative data set: Establishing trustworthiness of the data. *International Journal of Qualitative Methods*, *11*(3), 244–258. https://doi.org/10.1177/160940691201100305

Zull, C. (2016). GESIS survey guidelines: Open-ended questions. GESIS–Leibniz Institute for the Social Sciences. https://www.gesis.org/en/gesis-survey-guidelines/instruments/questionnaire-design/open-ended.questions

How to cite this article: Galura, S. J., Horan, K. A., Parchment, J., Penoyer, D., Schlotzhauer, A., Dye, K., & Hill, E. (2022). Frame of reference training for content analysis with structured teams (FORT-CAST): A framework for content analysis of open-ended survey questions using multidisciplinary coders. *Research in Nursing & Health*, 45, 477–487. https://doi.org/10.1002/nur.22227