

## Content and Thematic Analysis Techniques in Qualitative Research: Purpose, Process and Features

Muhammet Özden<sup>1</sup> 

### To cite this article

Özden, M. (2024). Content and thematic analysis techniques in qualitative research: Purpose, process and features. *Qualitative Inquiry in Education: Theory & Practice*, 2(1), 64-81. <https://doi.org/10.14689/qietp.2024.5>

### Article Info:

Received: 13.06.2024

Revised: 26.06.2024

Accepted: 16.07.2024

### Abstract

It has been observed that qualitative research methodologies are increasingly used in educational research. However, many errors can occur in qualitative research at various stages, from planning the study to collecting data, from selecting participants to writing the findings. One of the processes where novice qualitative researchers experience the most problems is data analysis. At this stage, researchers often do not know how to structure and code the data, and they lack confidence in their knowledge and skills regarding the processes of developing categories and themes. Another significant issue for researchers is their inability to understand the purposes and usage of the techniques applied in qualitative data analysis. It is even observed that in some studies, the processes of one analysis technique are replaced with those of another technique with similar characteristics. The continued increase of these problems negatively affects the robustness of qualitative research. The present study has been prepared to contribute to solving these issues observed in the literature. The study explains the characteristics, purposes, differences, and necessary steps of content analysis and thematic analysis approaches, which are frequently used in the analysis of qualitative data. Additionally, the study highlights the principles that should be considered in conducting content and thematic analyses, with the hope of assisting researchers in carrying out a rigorous and robust process.

### Keywords

qualitative research, qualitative data analysis, content analysis, thematic analysis

---

<sup>1</sup>**Corresponding Author:** Kütahya Dumlupınar University, Primary Education, Kütahya, Türkiye,  
[muhammet.ozden@dpu.edu.tr](mailto:muhammet.ozden@dpu.edu.tr)

## Introduction

Qualitative research can be defined as “*the process of developing a field-specific explanation or theory by examining the meanings created by individuals*” (Özden & Saban, 2017, p.5). However, it can be said that the majority of published qualitative studies are aimed at explanation and understanding. It has been observed that theory development studies are less frequent for various reasons. Researchers investigate the characteristic values, meanings, beliefs, thoughts, experiences, and emotions in phenomena and situations to develop the mentioned explanations or theories (Tashakkori & Teddlie, 2003). In this process, they utilize various data collection tools such as observation, interviews, and document analysis. The analysis of the obtained data is one of the areas where qualitative researchers struggle the most (Yıldırım & Şimşek, 2013).

Qualitative data analysis is the process of classifying and interpreting data to explain the explicit and implicit dimensions, meaning-making structures, and representations found within the data. It is also used to explore and describe issues in the field and processes in practice. The ultimate aim is generally to reach generalizable statements by comparing various materials, texts, or a few situations. Qualitative data analysis is conducted for the purposes of description, explanation, and theory development. Qualitative data can be analyzed to *describe* a phenomenon. This phenomenon may have emerged as a result of the subjective experiences of an individual or a group. In such cases, the analysis focuses on the phenomenon, its specific characteristics, and the connections between them. The analysis can also focus on comparing various individuals or groups experiencing the phenomenon, their common characteristics, or the differences between them. The second aim may be to determine the reasons underlying the existing differences, that is, *explanation*. The third aim is to *develop a theory* about the phenomenon being studied based on the analysis of the data (Flick, 2014).

Within the framework of the mentioned aims, techniques such as content analysis or thematic analysis can be used in the analysis of qualitative data. There are other qualitative analysis techniques, such as conversation analysis, narrative analysis, discourse analysis, and constant comparative analysis. Here, the most commonly used techniques, content and thematic analysis, are preferred. Both techniques are frequently employed in qualitative research. Content analysis is an approach where both numerical and thematic analysis processes are utilized, and culture and context are considered important interpretive features (Grbich, 2013). On the other hand, thematic analysis is often seen as a weaker analytical technique compared to content analysis. Thematic analysis can be introduced as a descriptive analytical technique that provides qualitative researchers with the basic skills necessary to employ many other qualitative data analysis techniques (Vaismoradi, Turunen, & Bondas, 2013). However, it can be said that what is more important than the strength or weakness of the analytical technique is its appropriateness to the subject being investigated and the adherence to rules in the analysis processes. In other words, when the reader reviews the analysis processes, they should not be prompted to ask what distinguishes this analysis from thematic analysis or content analysis.

In this study, the aim is to introduce the purposes, processes, and characteristics of content analysis and thematic analysis techniques and to explain how the data analysis process develops in both techniques. Information is provided under the following headings within the scope of the mentioned aim.

### **Definitions and Purposes of Content Analysis and Thematic Analysis**

Content analysis is also referred to as *qualitative content analysis* (Elo & Kyngäs, 2008) and *ethnographic content analysis* (Altheide, 1987). It is a systematic process of coding and categorization used to discover communication discourses and structures, patterns and trends of the words used, their frequencies, and their relationships in large amounts of information sources that are not immediately apparent. Content analysis is a systematic coding and classification approach that can be used to explore a large portion of existing textual information to identify the trends and patterns of the words used, their frequencies, relationships, communication structures, contexts, and discourses (Grbich, 2013). Thematic analysis, on the other hand, is a technique for describing and analyzing patterns within the data (Clarke & Braun, 2006). It involves conducting the analysis process based on the similarities, differences, and relationships in a data set. The term thematic relates to the aim of searching for themes clustered within the data (Gibson & Brown, 2009; Willig, 2013).

Both content and thematic analysis techniques aim to analytically examine and descriptively process narratives based on life stories by breaking them down into small analytical units (Sparker, 2005). Content analysis is a suitable technique for analyzing complex, significant, and sensitive phenomena (Elo & Kyngäs, 2008). It can be an appropriate analysis strategy to identify common points in the data, especially when what is known in a specific area is limited (Green & Thorogood, 2004). On the other hand, thematic analysis allows for rich and detailed descriptions of the data as a flexible and useful research tool (Braun & Clarke, 2006). This is because thematic analysis requires the investigation and identification of common characteristics within the data set (DeSantis & Noel Ugarriza, 2000).

Content and thematic analysis techniques are focused on the qualitative analysis of data. However, content analysis not only enables the qualitative analysis of data but also provides a quantitative representation. It is a descriptive approach used both in coding the data and interpreting the quantitative numbers of the codes (Grbich, 2013). In contrast, thematic analysis allows for a detailed and nuanced analysis of the data with a fully qualitative approach (Braun & Clarke, 2006).

### **Data Analysis Process in Content Analysis**

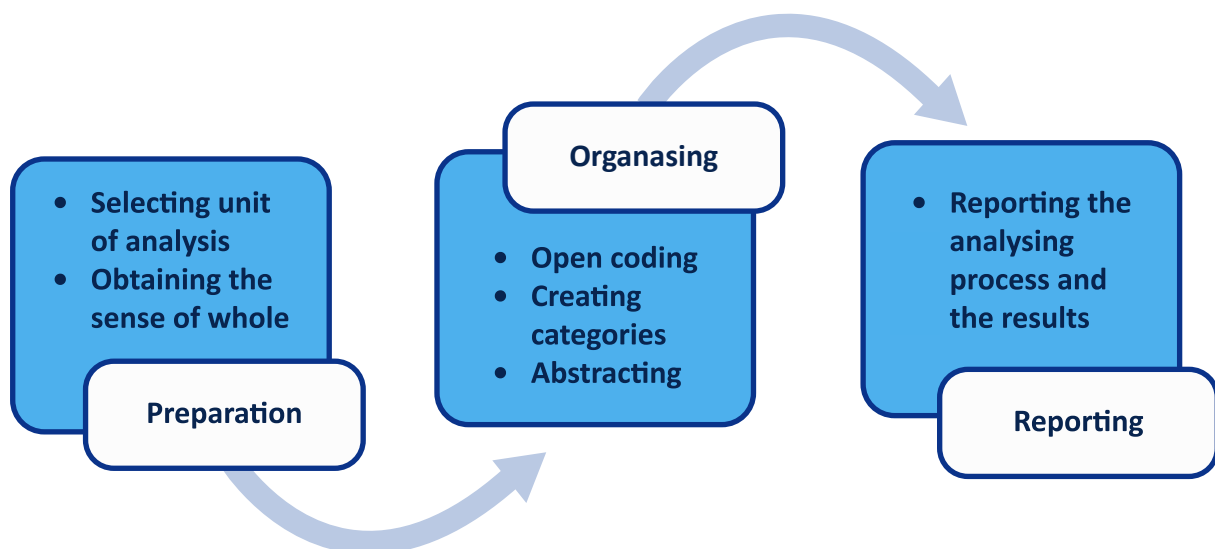
Inductive and deductive approaches can be used in the application of content and thematic analysis techniques. When what is known in a specific area is limited and there is insufficient information about the phenomenon being studied, an inductive approach can be used. In this approach, specific events are observed and relationships between these observations are established to reach generalizations. The inductive approach is used when there has been no prior research on the phenomenon, and thus, categories are derived directly from

the data (Hsieh & Shannon, 2005). On the other hand, if the aim of the research is to test a previous theory in a different situation or to compare categories obtained from different studies, a deductive approach may be appropriate. In this approach, the analysis is based on previous information, theories, or models, and the process progresses from general to specific (Elo & Kyngäs, 2008). In this study, the content analysis and thematic analysis processes are explained according to the inductive approach.

The data analysis process in content analysis consists of three stages: (i) *preparation*, (ii) *organizing*, and (iii) *reporting*. The tasks and procedures that researchers need to perform at each of these stages are explained below in Visual 1.

**Figure 1.**

*Data Analysis Process in Content Analysis*



### Stage One: Preparation

In the preparation stage, two fundamental processes are carried out: selecting the unit of analysis and obtaining the sense of whole. The unit of analysis is the focal point on which the researcher codes to answer the research questions. Depending on the research question, the unit of analysis can be a letter, word, sentence, phrase, interview, or entire observation protocols (Graneheim & Lundman, 2004). Furthermore, the unit of analysis can encompass an entire newspaper, a diary, a book or book chapter, a storybook, or a set of storybooks, among other possibilities. For instance, let's consider a phenomenological study on intercultural experiences of higher education students conducted by a researcher (see Ersoy, 2013). In this study, the unit of analysis would be the interviews conducted with each participant who has experienced intercultural encounters.

Selecting the appropriate unit of analysis not only guides the coding process but also influences the development of codes. Choosing the wrong unit of analysis can lead to two negative outcomes. First, if the chosen unit is too granular (i.e., at a much more micro level than what is actually needed), the researcher may overlook important contextual information and may need to conduct an analysis that requires more time and cost. An example of a unit of analysis that is too granular could be individual words. Second, if the chosen unit is too vague (i.e., at a high macro level), important relationships and contextual meanings within smaller (individual) units of content may be overlooked, leading to incorrect classification and interpretation of data. An example of a unit of analysis that is too vague could be all the journals written by 25 participants in a study or all the comments made by participants in an online support forum (Roller & Lavrakas, 2015).

Qualitative researchers should opt for a broader, more contextually based unit of analysis rather than a narrowly focused level of analysis (such as words or sentences). Researchers should avoid units that obscure meanings and nuances at an individual level. However, units of analysis that can be defined as entire interviews, observations, or focus group discussions offer researchers the possibility of obtaining reasonable and valid meanings in contexts that can be analyzed.

When starting the analysis, researchers need to decide whether they will analyze only the manifest content or also the latent content (Elo & Kyngäs, 2008). Manifest content refers to what is easily observable in the material, while latent content refers to the meanings inferred from the material (Saldaña & Omasta, 2021). Latent content is embedded within the material and is interpreted, interpreted, and explained by the researcher considering contextual conditions.

Decisions regarding the selection of manifest or latent content can influence the choice of the unit of analysis. A researcher analyzing manifest content may use quantitative content analysis. In content analysis, numerical data is preferred because the repetition of words in the analyzed document indicates the importance of those words. This numerical process relies on searching for keywords and word frequencies within the context. The units of content analysis that researchers can use during the process of analyzing manifest content may include the following (Grbich, 2013):

- *Word frequency*: This helps determine how often your key words occur in your documents. When counting word frequency, it is logical to exclude conjunctions such as “because, and, or, or, like, thing, which, etc.” as well as other common conjunctions used in your document set. After determining the frequency of key words, it would be useful to choose several other tools to describe the contextual use of the selected key words.
- *Keyword in context or concordance*: This approach displays every word in the document in alphabetical order and in context.
- *Category analysis or cluster analysis*: This is a stage where other related (synonymous) words are also included. For example, after the word “economy” is defined in the

document, other relevant words such as employment, unemployment, and inflation are also searched for. Then the computer shows how frequently each category occurs in the document.

- *Co-occurrence*: For this process, the researcher should search for specific words that are related. For example, words like security and terrorism. For this purpose, the researcher scans the document to check how closely some words are related to each other.

On the other hand, analyzing latent content requires a more nuanced understanding. A researcher focusing on implicit meanings in the data, consistent with research objectives, would approach the data with an ethnographic perspective. Such a researcher is expected to engage in processes such as description, exploring context, explaining meaning, and identifying patterns. A classic example of analyzing latent content is Little Orphan Annie. The research is significant for making visible the ideological representations of the emerging bipolar world rivalry after World War II. In the relevant study, questions such as (i) Who are Annie's friends and rivals? (ii) What are the goals endorsed by Annie and her friends? (iii) What recommendations do the characters make to achieve these goals? (iv) Which symbols do Annie, and her friends evaluate positively? and (v) Which symbols do they condemn? were addressed. Research data were obtained by examining 104 issues of the comic strip over two years (from April 1948 to July 1950). The analysis revealed that both the very poor (e.g., orphans) and the very wealthy (those Annie seeks help for orphans from) were Annie's most important friends, while Russian agents were depicted as a gang of young hoodlums extorting protection money from children. Annie's life goals were found to include making a lot of money, being charitable, being a law-abiding citizen, making a good marriage, and raising a large family. To achieve these goals, earning large amounts of money (i.e., money being both the end and the means), the use of power, and hard work were recommended. Symbols evaluated positively by Annie and her friends included orphans, workers, honest merchants, and smart businessmen, while symbols such as laziness, unwillingness to work, radicals, work camps, the Soviet Union, and Hitler were negatively evaluated (Shannon, 1954; Schreier, 2013). This study is important for explaining the visibility of latent content about conservatism, middle-class Americanism, capitalist values, and anti-Soviet ideology through content analysis.

Once the unit of analysis has been decided, the researcher should obtain a sense of the whole by reading the transcripts several times, striving to understand what is happening in the data (Morse & Field, 1995), and develop a holistic understanding. To achieve this, questions such as "who is telling the story, where is this happening? when did it happen? what is happening, why?" can be used during the process of reading the data with a comprehensive perspective that fosters a deeper understanding. Furthermore, during the preparation phase, written materials should be read several times to thoroughly understand the data. In this sense, during the preparation phase, no theory or previous research is more effective than the researcher's familiarity with the data. Once an understanding of the data relationship is developed, the analysis process can begin (Elo & Kyngäs, 2008). It can be said that determining the unit of analysis and becoming familiar with the data are critically important in the preparation phase. This way, the analysis can be conducted more meticulously, and the designated unit of analysis is also validated.

## **Stage Two: Organization**

This stage involves the processes of open coding, category formation, and abstraction. Open coding is a stage where notes are taken on the data and labels are assigned to the data using concrete or abstract descriptions. In vivo coding involves labeling the data using words or descriptions found in the documents we are working on during the analysis process. Abstract coding, on the other hand, involves assigning concepts related to the researcher's own theoretical sensitivity to the data in the documents we are working on. In the open coding process, any of these strategies or both can be used together. However, researchers using the abstract coding strategy can create a coding guide that includes conceptual explanations of the codes generated. This way, they can prevent possible inconsistencies in the data analysis process that spans over a long period. Whether using in vivo or abstract coding strategy, it is important to create a sufficient number of labels that explain all dimensions of the content. After the labeling stage is completed, a coding key is created to view the labels as a whole. Once this process is completed, relevant labels are grouped into categories based on their common characteristics.

The stage of open coding is followed by the process of category formation. In content analysis, the aim is to reach a condensed and broad definition of the phenomenon under investigation, making it crucial to create concepts and categories that describe the phenomenon (Elo & Kyngäs, 2008). A category is a group of content that shares a common characteristic. For example, codes such as “reflection on self”, “overcoming depression”, “overcoming insomnia”, “living with hope for a new day”, and “utilizing faith” come together to form the category of “increased personal strength” (Holley, 2020). As seen, category formation serves functions such as facilitating descriptive characterization of the content and expressing the visible content of the text.

Categories should be comprehensive yet distinct from each other. Any data relevant to the research purpose should not be excluded just because there isn't a suitable category for it. However, category formation is not simply about grouping similar or related categories together. Grouping categories involves classifying data based on belonging to a specific group. This is important for facilitating comparisons between data that belong to the same category and those that do not. The purpose of category formation is to create a tool that explains the phenomenon under investigation, develops understanding, and produces knowledge (Cavanagh, 1997). When creating categories, researchers make decisions by interpreting which data will be included in a particular category (Dey, 1993). At this stage, creating 10-15 categories can be considered ideal (Morse & Field, 1995). However, there is no rule or consensus on the number of categories you should create in qualitative research. For example, Creswell (2013) claims that there could be 25-30 categories. At this point, the criterion should be that the emerging categories encompass all relevant data related to the research purpose in sufficient quantity and adequacy.

The final component of the organization stage is abstraction. Abstraction means to express the general definition of the research topic clearly and precisely through categories (Elo & Kyngäs, 2008). This process shares similarities with theme development in thematic analysis. In short, at this stage, researchers attempt to create an integrative structure that adds

meaning to the phenomenon under investigation and its visible aspects, passing through various categories (Graneheim, Lindgren & Lundman, 2017). Abstraction can be seen as an expression of the hidden content of the text and a way to connect meanings within categories (Graneheim & Lundman, 2004). Additionally, in the organization stage, definitions should be developed for each theme, category, and code. Examples from the data should be identified for each code and category to prepare for reporting the findings (Hsieh & Shannon, 2005).

### **Stage Third: Reporting the Analysis Process and Results**

During the reporting process, researchers should provide information on how they conducted data entry, transcribed the data, ensured data security, developed coding processes, and anonymized participants. In a way, readers should be enabled to follow each stage of the analysis process, and a verifiable chain of evidence should be established (Twining, Heller, Nussbaum, & Tsai, 2017). At this stage, researchers should report in detail how they conducted the content analysis process for readers. In other words, they should explain with authentic examples how they performed the tasks and processes at each stage of the content analysis process in their own research. This way, they can ensure the credibility of their research. Additionally, researchers can utilize data analysis software to manage data during the analysis process. Models, concept maps, and categories can be used to report the results (Elo & Kyngäs, 2008).

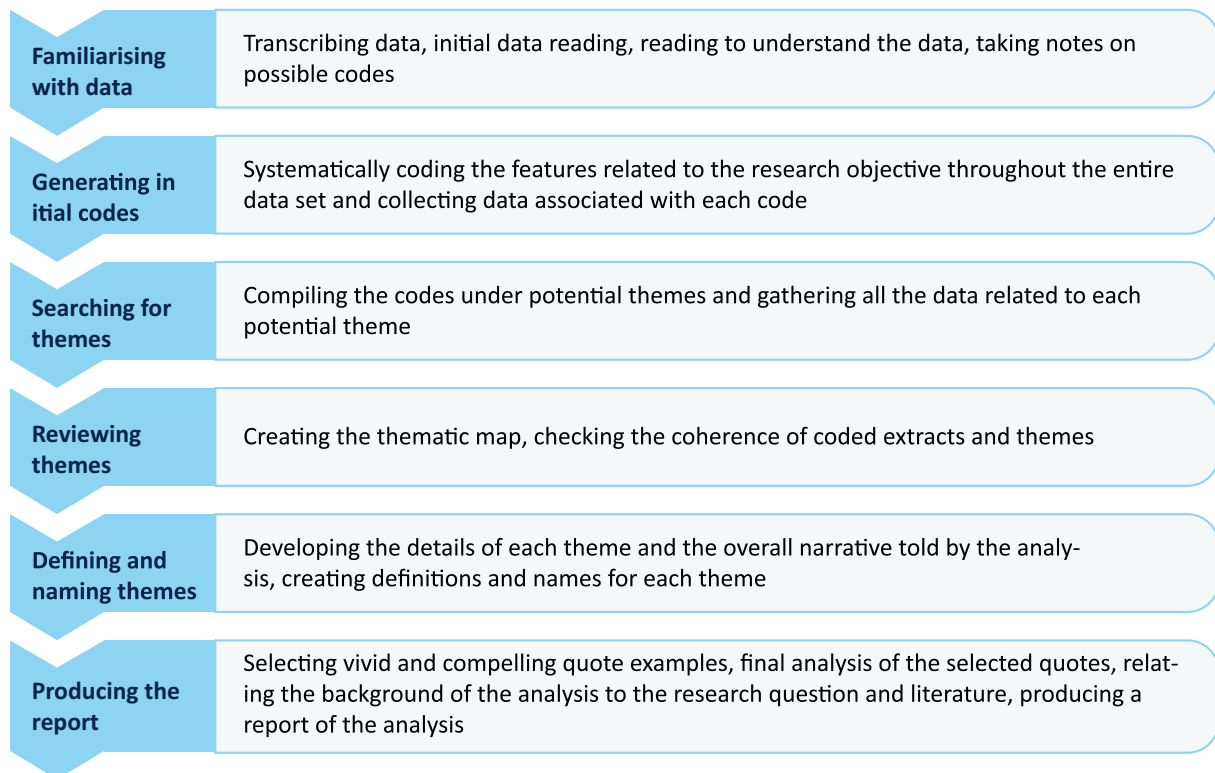
### **The Data Analysis Process in Thematic Analysis**

The data analysis process in thematic analysis consists of six stages. However, these stages should not be seen as a linear model where each stage is completed before moving on to the next. Instead, thematic analysis should be considered as an iterative process. The detailed explanation of how the thematic analysis process unfolds is provided in the following Figure 2 (Braun & Clarke, 2006).



**Figure 2.**

*Data Analysis Process in Thematic Analysis*



adapted from Vaismoradi, Turunen, & Bondas, 2013.

**Stage One: Familiarization with the Data**

When embarking on the analysis process, the researcher may have collected the data themselves, or the data might have been provided to them. It is preferable for the researcher or research team to collect and analyze their own data. If the data were collected by the researcher, they would begin the analysis process with some background knowledge and analytical thoughts about the data. In fact, the data collection process is, in a way, part of the analysis and should not be considered separate from it. Therefore, in qualitative analysis, who collects the data is an important consideration. Nonetheless, it is crucial for the researcher to immerse themselves in the data to understand its breadth and depth. Immersion often involves repeatedly but effectively reading the data, exploring meanings, and patterns. Ideally, the entire dataset should be read at least once before the researcher begins coding.

Taking notes for coding or marking the data to be coded will assist the researcher in the later stages. Once these processes are completed, it can be said that the researcher is ready for the more formal coding process. However, coding continues to evolve throughout the analysis process. Coding is a living, dynamic process influenced by ongoing developments.

## **Stage Two: Generating Initial Codes**

The second stage begins with the researcher immersing themselves in the data and becoming familiar with it. Consequently, the researcher creates an initial list of thoughts about what is in the data. This stage thus involves generating initial codes from the data. Codes can be described as identifying interesting features of the data and as “*basic units of raw data or information that can be associated with the phenomenon under investigation*” (Boyatzis, 1998, p.63). However, coding is distinct from themes. Themes emerge in subsequent stages of interpretive analysis where coding occurs. If a theme is a heading, then codes are subheadings under that heading. If a theme is a framework, then codes fall within that framework.

Coding may exhibit different characteristics depending on whether themes are data-driven or theory-driven. Initially, themes will be data-driven; however, the researcher will later approach the data with specific questions in mind. Coding also depends on whether the researcher will code the entire content of the dataset or only specific features of the dataset. This determination is primarily influenced by the research purpose/questions and the density of the dataset. Systematically working through the entire dataset, giving equal importance to data from each participant, and identifying interesting features in each participant can form the basis of recurring patterns, briefly themes.

During the stage of generating initial codes, it is particularly important for researchers to: a) code to generate as many potential themes/patterns as possible, b) code extensively where quotations will be taken from the data, c) code participants’ individual quotations under different relevant themes. As a result of these processes, a quotation may remain uncoded, may be coded once, or may be coded multiple times if necessary.

## **Stage Three: Searching for Themes**

The third stage begins after the completion and consolidation of initial coding for all the data. At this stage, the researcher has a long list of different codes derived from the entire dataset. Therefore, the focus shifts to the more comprehensive analysis of themes. For this purpose, different codes are classified under potential themes, and the themes are integrated with all relevant coded quotations. Essentially, the researcher has started analyzing the codes and is contemplating how to bring different codes together to form inclusive themes. It may be helpful at this stage to visually present different codes grouped under themes. The researcher can use tables and mind maps or engage in organizing codes within clusters of themes by writing the name of each code and a brief description on a separate sheet of paper. These activities emerge as the researcher begins to think about the relationships between codes, themes, and different levels of themes. Through these processes, some initial codes may be used to form main themes, while others may form sub-themes, and perhaps some may be discarded. Additionally, at this stage, the researcher may also identify sets of codes that do not belong anywhere. This group of codes, likely labeled as “other,” is probably a temporary theme.

The researcher concludes this stage by associating candidate themes, sub-themes, and related coded data quotations. At this point, the researcher should strive to develop a rationale for the importance of individual themes. However, deciding whether themes will be kept as they are, merged, reviewed, separated, or discarded without thoroughly examining all quotations would be a mistake.

#### **Stage Four: Reviewing Themes**

The fourth stage begins with the researcher identifying a series of candidate themes and involves the development and review of these themes. At this stage, it may become apparent that some candidate themes lack sufficient supporting evidence or are not genuine themes due to the data having different characteristics. Moreover, at this stage, two themes may be structured to create a new theme, or one theme may give rise to two new themes. Throughout this process, there should be clear and identifiable distinctions between themes, while the data under each theme should cohesively fit together.

This stage encompasses two levels: reviewing and refining themes. The first level requires reviewing coded data quotations, which involves reading through amalgamated quotations for each theme and considering whether these quotations form a consistent pattern. If the researcher finds that candidate themes are creating a coherent and consistent pattern, they can proceed to the second level of this stage. However, if the candidate themes are not coherent, the researcher should check whether the theme itself is problematic or if some quotations within the theme do not align with it. In either case, the researcher needs to work on the theme, create a new theme, find a place for quotations that do not serve the existing theme, or remove them from the analysis. If the researcher believes that candidate themes adequately capture the lines of coded data, they are ready to proceed to the second level of this stage.

The second level involves similar processes, but the operations are conducted across the entire dataset. At this level, the researcher must pay attention to the relevance of individual themes to the dataset. However, the researcher should also consider whether the candidate thematic map accurately reflects the evidence from the entire dataset as a whole. At this stage, the researcher should re-read the entire dataset for two purposes. First, as discussed earlier, to determine if the themes regarding the dataset are working. Secondly, to read the dataset to code the data that may have been overlooked during the early stages of coding into the themes. The need for re-coding from the dataset can be explained by the fact that coding is an ongoing organic process.

After completing all these processes, if the thematic map created by the researcher is working, they can proceed to the next stage. However, if the thematic map does not fit the dataset, the researcher should return to the review and correction process until a satisfactory thematic map is established. This allows the researcher to identify potential new themes, and if the data is relevant and significant, they will start re-coding. It is almost impossible to determine when the coding process will end. However, if the refining process does not yield anything new, it may be time to stop. If the re-coding process evolves into fine-tuning and addressing minor details within a functioning coding framework, coding should be halted.

At the end of this stage, the researcher is expected to have an understanding of the different themes they have created, their compatibility with each other, and their characteristics reflecting the story of the data.

### **Stage Five: Defining and Naming Themes**

The fifth stage begins with the researcher establishing a satisfactory thematic map of their data. At this point, the researcher proceeds to identify and clarify the themes, which they will present for analysis, and analyze the data within these themes. Identifying and clarifying involves determining the essence of each theme and deciding which dimensions of the data each theme covers. The researcher should return to the amalgamated data excerpts for each theme and organize them into a coherent and cohesive narrative. It is expected that the researcher not only interprets the content of the presented data excerpts but also determines what each theme is about and why.

The researcher should write detailed analyses for each theme. They should consider where each theme fits within the overarching narrative of the study. This allows for a focus on both the function of the themes and their relationship with other themes. As part of the refining process, the researcher should determine whether a theme contains sub-themes. Sub-themes can be particularly useful for structuring broad and complex themes and can also be used to demonstrate the hierarchical meaning within the data.

At the end of this stage, the researcher should clearly define what the themes are or are not. Researchers can evaluate the adequacy and functionality of each theme by writing a few sentences explaining the scope and content of each theme. If a theme is not working, further refinement and review work should be carried out on the theme. Theme names should be concise, precise, effective, and should give the reader an immediate idea about the theme.

### **Stage Six: Producing the Report**

The sixth stage begins with the researcher having a set of functional themes and involves writing the research report based on the final analysis. The research report is written to persuade the reader of the complex story derived from the analysis in a valuable and valid manner while staying grounded in the data. The report requires concise, coherent, logical, non-repetitive writing that presents the story told by the data within and framed by the themes. When writing the report, themes should be supported with sufficient evidence. For example, quotations should be included in the report to demonstrate the strength of the themes. The researcher should select particularly strong examples or quotations that reflect the essence of the topic they want to convey, without introducing unnecessary complexity. Quotations should be easily identifiable as examples of the issue at hand.

However, writing the report involves more than just presenting the data. Quotations should be embedded within the analytical narrative based on the researcher's data, going beyond mere description and relating to the research questions. In research using thematic analysis technique (see Özden, Uçansoy Baştürk, & Demir, 2015), examples of the process of coding, sub-theming, and theme development are presented in Table 1:



**Table 1.**

*Relationship Between Code, Category, and Theme*

<b>Codes</b>	<b>Sub-themes</b>	<b>Themes</b>
Not studying for the lesson, challenging, desire for excitement	Personal characteristics	Individual Factors
Passing the course, desire for success, desire to achieve high grades	Seeking academic advantage	
Low academic self-esteem	Inadequate academic self-concept	
Failure to review assignments, proctors, personality traits of instructors, lack of seriousness in proctors, indifference of instructors, lack of objectivity in proctors, insensitivity of proctors, inadequacy of instructors, quality of instructors	Attitudes and actions of instructors	Environmental Factors
Competitive education, difficulty of exams, exam system, emphasis on cognitive assessment, lack of consistency between course content and exam content	Competition-based education and exams	
Minimal or intense course content, rote-learning structure of courses, teaching methods, rote-learning system	Content structure and role of the course	
To know what everyone else is doing, socializing, not wanting to embarrass friends, cooperation	Influence of peers	
Students not being punished	Institutional characteristics	
Not wanting to embarrass the family, risk of losing the student's scholarship	Familial and economic reasons	

In the mentioned study, researchers have generated 228 codes. Examples of these codes include “*avoiding studying, desire for high grades, insensitivity of invigilators, difficulty of exams, excessive course content.*” After completing the coding of the data, the researchers grouped them to create meaningful patterns that could form sub-themes. At this stage, names for sub-themes were determined for groups of codes that could form patterns together. In this context, sub-themes such as “*Personal Characteristics, Inadequate Academic Self-Concept, Instructors’ Attitudes and Actions, Institutional Characteristics*” were created. In other words, the researchers clustered the 228 codes to create nine sub-themes. As a result of these processes, the researchers questioned whether the created sub-themes reflected the overall appearance of the data. Through discussions, it was decided that the sub-themes exhibited characteristics to create new patterns. They believed that sub-themes such as “*Personal Characteristics, Academic Achievement, and Inadequate Academic Self-Concept*” could constitute personal reasons for cheating, and by bringing these sub-themes together, they formed the theme “*Individual Factors.*” They defined the theme of individual factors as a feature that does not explain reasons for cheating related to academic procrastination behavior, academic gains, and overcoming low self-efficacy perception. On the other hand,

it was thought that the sub-themes “*Instructors’ Attitudes and Actions, Competition-Based Education and Exams, Content Structure and Role of the Course, Influence of Peers, Institutional Characteristics, Familial and Economic Reasons*” were related to the institutional, economic, and social environment in which the individual is located. They conceptualized this theme as “*Environmental Factors*,” which helps to contextualize cheating behavior within a framework that facilitates, neutralizes, and rationalizes ethical violations.

## **Conclusion**

The preparation phase of content analysis and the data familiarization phase in thematic analysis exhibit similar characteristics. In both stages, researchers are expected to immerse themselves in the data, read through them several times, and grasp the overall meaning. While thematic analysis primarily advises researchers to consider both latent and manifest content in data analysis, researchers using content analysis can choose between manifest (emerging categories) and latent content (emerging themes) before moving on to the next stage of data analysis. Open coding, grouping codes under possible sub-categories/sub-themes or categories/themes, and comparing and associating developing coding clusters with the entire dataset constitute features of the organizing stage, which is the second stage of data analysis in content analysis. This set of analytical processes used in content analysis is performed under the headings of generating initial codes, naming and defining themes, reviewing themes, and exploring themes in thematic analysis.

As previously mentioned, there are many similarities between the data analysis processes in different stages. The terminology used throughout the data analysis process in both approaches contains closely related explanations. In thematic analysis, terms such as data corpus, dataset, data extract, code, and theme are equivalent to the concepts of analysis unit, unit of meaning, condensed meaning unit, code, and category in content analysis, respectively (Graneheim & Lundman, 2004; Braun & Clarke, 2006; Elo & Kyngäs, 2008).

The way findings are presented has been widely discussed in both content and thematic analysis. Simply put, the emergence of theme(s) can be considered the result or final product of data analysis. The term “theme” has been associated with many definitions and used interchangeably with other concepts such as category, domain, unit of analysis, stage, process, outcome, and strategy (DeSantis & Noel Ugarriza, 2000). A theme is a structure that integrates different data points into a coherent whole (Sandelowski & Leeman, 2012). It captures some key points in the data related to the research question and may indicate some meanings or patterns that could answer the research question in the dataset (Braun & Clarke, 2006). A category primarily refers to a descriptive level of content and can be seen as an expression of the explicitly evident content of the text. On the other hand, a theme can be defined as the interpretation of latent content within the text (Graneheim & Lundman, 2004). Particularly in thematic analysis, themes are often challenging to define as they tend to be abstract (DeSantis & Noel Ugarriza, 2000; Spencer et al., 2003). Additionally, in thematic analysis, a theme does not necessarily have to be tied to quantifiable criteria; instead, it depends on whether something significant related to the overarching research question is found (Spencer et al., 2003; Braun & Clarke, 2006). In content analysis, however, it is possible

to access a theme based on frequency within the text. Content analysis is objective and systematic, focusing on apparent meanings rather than hidden meanings in the document (Bloor & Wood, 2006).

When conducting content analysis, one of the first decisions to be made is whether the analysis will focus on explicit or implicit content. Content analysis deals with interpreting both explicit and implicit content, but the depth of interpretation and level of inference may vary (Graneheim & Lundman, 2004; Powers & Knapp, 2006). In contrast, thematic analysis integrates both the explicit and implicit aspects of the data. This means that analyzing the hidden aspects of the data is an integral part of the open-ended approach to analysis (Braun & Clarke, 2006).

Another characteristic of data analysis in thematic analysis is the creation of a thematic map. This map includes a visual presentation of themes, codes, their relationships, detailed descriptions and explanations of each theme, their criteria, examples, counterexamples, and other relevant details. Additionally, as part of the data analysis process, it helps to review and identify consistent yet distinct themes (Ryan & Bernard, 2000; Braun & Clarke, 2006). In both approaches, data analysis processes are not linear but iterative. Nonetheless, regular reviews are essential. The conclusion should take the form of a narrative describing the data in relation to the research question or questions.

The final stage of data analysis in both approaches is reporting research findings. This stage is particularly emphasized as the culmination of the data analysis process in thematic analysis. Additionally, in both approaches, researchers are encouraged to use their creativity to present findings in the form of a narrative, play, film, or other narrative forms, as well as through maps or models. It is emphasized that high-quality data analysis depends on collecting high-quality data. It is the researcher's responsibility to collect quality data and present complex data in an engaging manner that yields interesting findings. After collecting and transcribing data, analysts are advised to immerse themselves in the data through reading and re-reading, paying special attention to participants' behaviors and emotions, in order to capture the overall meaning (Polit & Beck, 2003).

### **Declaration of Conflicting Interests**

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

### **Funding**

The author declared that this study has received no financial support.

## References

- Altheide, D. L. (1987). Ethnographic content analysis. *Qualitative Sociology*, 10(1), 65-77.
- Ballinger, C., Yardley, L., & Payne, S. (2004). Observation and action research. In D.F. Marks & L. Yardley (Eds), *Research methods for clinical and health psychology* (pp.102-121). London: Sage.
- Bloor, M., & Wood, F. (2006). *Keywords in qualitative methods: A vocabulary of research concepts*. London: Sage.
- Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101.
- Cavanagh, S. (1997). Content analysis: Concepts, methods and applications. *Nurse Researcher*, 4, 5-16.
- DeSantis, L., & Noel Ugarriza, D. (2000). The concept of theme as used in qualitative nursing research. *Western Journal of Nursing Research*, 22, 351-372.
- Dey, I. (1993). *Qualitative data analysis: A user-friendly guide for social scientists*. London: Routledge.
- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62, 107-115.
- Ersoy, A. (2013). Türk öğretmen adaylarının kültürlerarası deneyimlerinde karşılaştıkları sorunlar: Erasmus değişim programı örneği. *Eğitim ve Bilim*, 38(168), 154-166.
- Flick, U. (2014). Mapping the field. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp.3-18). Los Angeles: Sage.
- Gbrich, C. (2013). *Qualitative data analysis: An introduction (2<sup>nd</sup> ed)*. London: Sage.
- Gibson, W. J., & Brown, A. (2009). *Working with qualitative data*. Los Angeles: Sage.
- Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today*, 24, 105-112.
- Graneheim, U. H., Lindgren, B.-M., & Lundman, B. (2017). Methodological challenges in qualitative content analysis: A discussion paper. *Nurse Education Today*, 56, 29-34.
- Green, J., & Thorogood, N. (2008). Analysing qualitative data. In Silverman D (Ed.), *Qualitative methods for health research* (pp.173-200). London: Sage.



- Guthrie, J., Yongvanich, K., & Ricceri, F. (2004). Using content analysis as a research method to inquire into intellectual capital reporting. *Journal of Intellectual Capital*, 5, 282-293.
- Holley, R. M. (2020). *Content analysis of combat veterans' experiences with posttraumatic growth*. (Dissertation Manuscript). ProQuest Dissertations and Theses. UMI 28028851.
- Horsburgh, D. (2003). Evaluation of qualitative research. *Journal of Clinical Nursing*, 12, 307-312.
- Hsieh, H.F. & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15, 1277-1288.
- Lincoln, Y. & Guba, E. (1985). *Naturalistic inquiry*. London: Sage.
- Lindgren, B.-M., Lundman, B. & Graneheim, U. H. (2020). Abstraction and interpretation during the qualitative content analysis process. *International Journal of Nursing Studies*, 108. DOI: 10.1016/j.ijnurstu.2020.103632
- Loffe, H., & Yardley, L. (2004). Content and thematic analysis. In D.F. Marks & L. Yardley (Eds), *Research methods for clinical and health psychology* (pp. 56-69). London: Sage.
- McCain, G.C. (1988). Content analysis: A method for studying clinical nursing problems. *Applied Nursing Research*, 1(3), 146-150.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. London: Sage.
- Morse J. M., & Field P. A. (1995). *Qualitative research methods for health professionals*. Thousands Oaks: Sage.
- Özden, M., & Saban, A. (2017). Nitel arařtırmalarda paradigma ve teorik temeller. A. Saban & A. Ersoy (Edt.), *Eđitimde nitel arařtırma desenleri içinde* (ss.1-30). Ankara: Anı.
- Özden, M., Uçansoy Bařtürk, A., & Demir, M. (2015). Kopya çektim, çünkü...: Bir olgubilim çalıřması. *Turkish Online Journal of Qualitative Inquiry (TOJQI)*, 6(4), 57-89.
- Polit, D., & Beck, C. (2003). *Nursing research: Principles and methods*. Philadelphia: Lippincott Williams & Wilkins.
- Powers, B., & Knapp, T. (2006). *Dictionary of nursing theory and research*. New York: Springer.
- Robson, C. (1993). *Real world research: A resource for social scientists and practitioner-researchers*. Oxford: Blackwell Publishers,
- Rolfe, G. (2006). Validity, trustworthiness and rigour: Quality and the idea of qualitative research. *Journal of Advanced Nursing*, 53, 304-310.

- Roller, M. R., & Lavrakas, P. J. (2015). *Applied qualitative research design: A total quality framework approach*. New York: The Guilford Press.
- Ryan, G., & Bernard, H. (2000). Data management and analysis methods. In N. K. Denzin & Y. S. Lincoln (Eds), *Handbook of qualitative research* (pp. 769-802). London: Sage.
- Saldaña, J., & Omasta, M. (2021). *Qualitative research: Analyzing life*. Sage.
- Sandelowski, M. (2010). What's in a name? Qualitative description revisited. *Research in Nursing & Health*, 33(1), 77-84.
- Sandelowski, M., & Barroso, J. (2003a). Writing the proposal for a qualitative research methodology project. *Qualitative Health Research*, 13, 781-820.
- Sandelowski, M., & Leeman, J. (2012). Writing usable qualitative health research findings. *Qualitative Health Research*, 22, 1404-1413.
- Schreier, M. (2013). Qualitative content analysis. In U. Flick. (Ed.), *The Sage handbook of qualitative data analysis* (pp. 170-183). Los Angeles: Sage.
- Shannon, Lyle W. (1954). The opinions of the little orphan Annie and her friends. *Public Opinion Quarterly*, 18, 169-79.
- Sparker, A. (2005). Narrative analysis: Exploring the whats and hows of personal stories. In I. Holloway (Ed.), *Qualitative research in health care* (pp. 191-208). Berkshire: Open University Press.
- Spencer, L., Ritchie, J., O'Connor, W. (2003). Analysis: Practices, principles and processes. In J. Ritchie & J. Lewis (Eds), *Qualitative research practice: A guide for social sciences students and researchers* (pp.199-218). London: Sage.
- Tashakkori, A., & Teddlie, C. (2003). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks: Sage
- Tuckett, A. G. (2005). Applying thematic analysis theory to practice: A researcher's experience. *Contemporary Nurse*, 19, 75-87.
- Twining, P., Heller, R. S., Nussbaum, M., & Tsai. C.-C. (2017). Some guidance on conducting and reporting qualitative studies. *Computers & Education*, 106, A1-A9. <https://doi.org/10.1016/j.compedu.2016.12.002>.
- Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & Health Sciences*, 15, 398-405.
- Willig, C. (2013). Interpretation and analysis. In U. Flick. (Ed.), *The Sage handbook of qualitative data analysis* (pp. 136-149). Los Angeles: Sage.