Content Analysis or Thematic Analysis: Similarities, Differences and Applications in Qualitative Research

Niklas Humble and Peter Mozelius
Mid Sweden University, Östersund, Sweden
niklas.humble@miun.se
peter.mozelius@miun.se

Abstract: Research has a long tradition of quantitative research which still dominates many university courses on research methods. Qualitative research is a younger phenomenon that was established in research after the second world war. An emerging research field that needed new analysis methods tailored for qualitative data. Two of the most frequently used approaches in qualitative data analysis are content analysis and thematic analysis. In several aspects content analysis and thematic analysis both share a common approach to analytically examine qualitative data, and the fact that they have been used interchangeably has made it difficult for the more unexperienced researchers to distinguish and choose between them. The aim of this study is to examine doctoral students' perceptions of qualitative analysis with content analysis and thematic analysis. The study had a qualitative approach with data collected from two webinars on qualitative data analysis, where a total of 76 doctoral students participated. Data consists of participant reflections in a Padlet on content analysis and thematic analysis at the two webinars. Webinar participants have given their consent to use their reflections in the Padlet for research. Content analysis with an abductive coding approach was used to analyse the collected data and formulate categories that answer the study's aim and research question. Results show both perceived similarities and differences between content analysis and thematic analysis. Both are perceived to have a similar process in the coding of data, although content analysis has a wider selection of coding approaches and thematic analysis support deeper immersion. Content analysis is also perceived as more practical and straightforward, while thematic analysis is perceived as more intuitive and faster to learn. Both content analysis and thematic analysis are perceived to have individual opportunities and challenges that make them appropriate for different types of research. Findings presented in this study can be used by researchers at any level to explore similarities and differences between content analysis and thematic analysis, and where to apply them in research.

Keywords: Content analysis, Thematic analysis, Qualitative analysis, Qualitative research, Coding

1. Introduction

Research has a long tradition of quantitative research which still has a strong impact on university courses on research methods. Qualitative research is a younger phenomenon that began in the 1920s with the work and advocacy of the psychologist, Paul Felix Lazarsfeld (Jeřábek, 2001; Bailey, 2014). After the Second world war, qualitative methods established itself in research areas such as communication, journalism, sociology, business psychology, and anthropology. However, the development of qualitative methods has far from replaced the use of quantitative methods in these areas (O'Dwyer & Bernauer; Grossman & Cohen, 2017), and with an increased use of mixed method approaches (McCusker & Gunaydin, 2015; Bernard, 2017). In many research areas today, research methods are chosen based on the actual research design (McCusker & Gunaydin, 2015).

In the domain of qualitative research there are several approaches for analysing data (Thorne, 2000; Roller, 2019). Some examples of established methods for analysing qualitative data are content analysis, thematic analysis, open coding, narrative analysis, discourse analysis, and phenomenological analysis (Hsieh & Shannon, 2005; Nowell et al., 2017; Walia, 2015). This study had a focus on content analysis and thematic analysis, and the comparison of the two. In several aspects, content analysis and thematic analysis share a common approach to the analysis of qualitative data, and the fact that they often have been used interchangeably has made it difficult to distinguish and choose between them in research studies. What are doctoral students' perceptions of content analysis and thematic analysis after participating in a webinar where these two approaches were presented, applied and discussed? The aim of this study is to examine doctoral students' perceptions of qualitative analysis with content analysis and thematic analysis.

2. Content analysis and thematic analysis

Quantitative content analysis was used in the early 19th century by Thomas Young in the deciphering of the Rosetta Stone (Larmor, 1934). The first documented use of qualitative content analysis was in 1893 with the aim to find patterns in Shakespeare texts (Sumpter, 2001). Thematic analysis is a younger phenomenon that was more strictly defined by Boyatzis (1998). However, thematic analysis had been used earlier under other names before the term was coined in research fields such as psychology, literature, business study, and sociology (Boyatzis, 1998). In both content analysis and thematic analysis, themes or patterns or codes can be identified

and grouped in mainly two different ways, inductively or deductively (Braun & Clarke, 2006). With the inductive approach, the identified themes or patterns or codes are strongly linked to the data themselves, in what could be described as a bottom-up process (Patton, 1990; Braun & Clarke, 2006). This could be compared to the deductive approach where data is analysed in a top-down process, driven by selected theoretical or analytic interests (Boyatzis, 1998; Braun & Clarke, 2006).

2.1 Qualitative content analysis

Content analysis describes a variety of analytic approaches in a range from impressionistic and intuitive analyses to the side of very strict and systematic text analyses (Rosengren, 1981). This variety has made content analysis useful in several different research fields, where the suitable type of content analysis approach to use depends on the actual problem that is studied, and the specific research purpose (Weber, 1990). In a study by Hsieh and Shannon (2005), three types of qualitative content analysis were defined and presented: Conventional content analysis, Directed content analysis, and Summative content analysis (Table 1).

Table 1: Summary of the three approaches to content analysis (Table from Hsieh & Shannon, 2005)

Type of content analysis	Study starts with	Timing of defining codes or keywords	Source of codes or keywords
Conventional content analysis	Observation	Codes are defined during data analysis	Codes are derived from data
Directed content analysis	Theory	Codes are defined before and during data analysis	Codes are derived from theory or relevant research findings
Summative content analysis	Keywords	Keywords are identified before and during data analysis	Keywords are derived from interest of researchers or review of literature

2.1.1 Conventional content analysis

In some university courses this is seen as the "standard content analysis", but at others not. Conventional content analysis is often used in studies that have the aim to explore a phenomenon where theory or research literature on the studied phenomenon is limited (Hsieh and Shannon, 2005). In this analysis approach the researcher works inductively, avoids the use of pre-defined categories, and instead creates categories that emerges from the analysed data (Kondracki, Wellman & Amundson, 2002). An advantage of the conventional approach to content analysis is to directly generate information from informants without imposing preconceived theoretical perspectives. In a five-step process, the conventional content analysis could be described briefly as: 1) Immerse yourself actively into the data, read all data repeatedly to get a sense of the whole, 2) Derive codes by read word by word while highlighting key thoughts and concepts, 3) Label codes from notes of first impressions and thoughts, labels emerge that reflect multiple key thoughts that are based directly in the data, 4) Creating categories (and subcategories) by sorting and organising the codes into clusters, based on relationship between the codes, and finally 5) Developing and fine-tuning definitions for categories and codes with identified examples from the data. (Hsieh and Shannon, 2005)

2.1.2 Directed content analysis

In a research field with more prior research and existing theory, the qualitative researcher could instead choose the directed content analysis. This approach has been classified as a deductive, top-down and theory driven process (Potter & Levine-Donnerstein, 1999; Hsieh and Shannon, 2005). This approach is often used in studies that aim to extend the existing description of a phenomenon. Moreover, the chosen theory or prior research often guides the discussion on the outcomes of a directed content analysis (Hsieh and Shannon, 2005). The directed approach to content analysis is guided by a more structured process, if compared to the conventional approach (Hickey & Kipping, 1996). With the use of existing theory or earlier research, the first step is to identify important key concepts or variables as initial coding categories (Potter & Levine-Donnerstein, 1999), followed up by the second step where the operational definitions for the categories also are guided by the chosen theory. In a third step, data are coded and sorted into the predetermined categorises. All data rarely fit into the predetermined categories and as a fourth step other potentially relevant data is highlighted, followed up by a fifth step where the highlighted data are analysed and grouped into new categories, or as sub-categories to existing categories (Hsieh and Shannon, 2005).

2.1.3 Summative content analysis

A third approach is the summative content analysis which also has a quantitative touch. Selected words or content in the data are quantified to better understand the contextual use of the words or content, not to infer meaning but, rather to explore usage of them. (Hsieh and Shannon, 2005) However, an analysis that stops there

would be classified as quantitative, if only focusing on the counting or the frequency of certain content in data (Kondracki, Wellman & Amundson, 2002). A summative content analysis must continue beyond the mere frequency counts and comprise a latent content analysis as well. The term latent content analysis refers to the process of interpretation of the content (Holsti, 1969). The initial quantitative approach for deriving keywords could preferably be computer supported, while the analysing of patterns and relationships to understand contextual meanings is carried out manually by the researchers. Boiled down to a five-step process a summative content analysis could be conducted as: 1) Create the initial coding scheme based on theory, previous research or data, 2) Counting the frequency of certain words or concepts, 3) Calculate and compare found frequencies in relation to total length of data and/or to various data sources, 4) Identify and quantify alternatives by counting frequencies of the alternatives to the words or concepts used in step 2, and with the final step of 5) Update the calculation and comparison with the identified alternative words and concepts. (Hsieh and Shannon, 2005)

2.2 Thematic analysis

The use of thematic analyses has strongly increased in the 21st century after the term was more clearly defined by Boyatzis (1998). Another article that has introduced thematic analysis to a wider audience is the one by Braun and Clarke (2006). Of all research articles published that year, this article has the highest number of citations in Google scholar. Some reasons for the broad use in various research fields are probably the straightforward approach for finding themes that are relevant for the actual research question and the chosen research study design. Without any requirement of frequency counting as in a summative content analysis, to enable the idea that large data sets could be used for several constrained studies with different research questions. In the section above, the three branches of content analysis have been summarised as five-step processes. A thematic analysis was outlined by Braun and Clarke (2006) as a six-step process with the phases of 1) Familiarising yourself with your data, involving transcription, immersion and noting down initial ideas, 2) Generating initial codes from the data that identify a feature of the data, semantic content or latent, that appears interesting, 3) Searching for themes by collating codes into potential themes, 4) Reviewing themes, and checking if the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), 5) Defining and naming themes, and refine the specifics of each theme, and the overall story the analysis tells, and 6) Writing up the presentation of the found theme and fine-tuning the overall story. (Braun and Clarke, 2006) There are also other processes for conducting thematic analysis, such as the slightly extended seven-stage process presented by Norton (2009).

3. Method

The study used a qualitative approach to produce knowledge about the participants' understandings of the studied topic and the meanings they attach to it (Polkinghorne, 2010). Data were collected and analysed from two webinars (online seminars) on content analysis and thematic analysis with a total of 76 participants. The first webinar had 60 international participants, the second webinar had 16 Swedish participants, and both webinars were targeted towards doctoral students. During the webinar, the authors presented material on content analysis (Hsieh & Shannon, 2005; Drisko, 2005; Blair, 2015) and thematic analysis (Braun & Clarke, 2006; Braun & Clarke, 2012; Braun, Clarke & Hayfield, 2019) for qualitative research which were later used for discussions and workshops. As a part of this, a Padlet was presented at the end of the webinars where the participants were asked to share their reflections on similarities and differences between content analysis and thematic analysis, and possibilities and challenges with content analysis and thematic analysis.

In the Padlet, the participants could post their reflections on the topic and comment on other participants' reflections. The Padlet was set to be anonymous, that is no names were shown or required to participate in the Padlet. The authors also informed the participants, before sharing the link for the Padlet, about the intent to use the Padlet reflections for research and that only those that wanted their reflections to be included should post and comment on the Padlet. The participants answered the Padlet either after the webinar or during a break to ensure anonymity from both the authors and the other participants.

Content analysis was used to examine the content and contextual meaning of the collected data (Hsieh & Shannon, 2005). The coding approach was inspired by *conventional content analysis* (Hsieh & Shannon, 2005) and the *abductive approach* for moving between deductive and inductive coding (Graneheim, Lindgren & Lundman, 2017). The coding could be summarised in two parts. In the first part, codes were identified in the Padlet and moved to the spreadsheet by the first author. No predetermined categories were used to identify or organise the codes, instead the categories were formulated during the coding process. With each code added to the spreadsheet either creating a new category, adding to an existing category, or re-formulating an existing category. In the second part, all codes and categories were revised for consistency by both authors.

4. Results and analysis

Through the analysis of the collected data, five main categories have been identified regarding doctoral students' perception of qualitative analysis with content analysis and thematic analysis (Table 2). The first three are considered similarities between content analysis and thematic analysis: 1) similar process of coding and analysing data, 2) similar challenge regarding subjectivity and potential bias, 3) similar challenge in defining and organising data. The last two are considered differences between content analysis and thematic analysis: 4) wide application vs. deep analysis, 5) practical approach vs. intuitive approach.

Table 2: Summary of identified categories

Similarities between content analysis and thematic analysis	*Differences between content analysis and thematic analysis*
1) Similar process of coding and analysing data	4) Wide application vs. deep analysis
2) Similar challenge regarding subjectivity and potential bias	5) Practical approach vs. intuitive approach
3) Similar challenge in defining and organising data	

4.1 Similarities between content analysis and thematic analysis

Content analysis and thematic analysis are considered to have similar processes to conduct qualitative analyses. They both require the researcher to familiarise with the data before coding and analysing. They both approach data as something "out there", content, that can be collected and analysed. Further, both content analysis and thematic analysis can be conducted inductively and deductively.

Content analysis and thematic analysis are considered to have similar challenges regarding subjectivity and potential biases. Both analyses are considered at risk for biased reading when interacting with the collected data, since subjectivity is involved when interpreting collected data. However, thematic analysis seems to be considered more at risk in this regard since it is also considered more "creative" and "open for interpretation" compared to content analysis. While content analysis is considered to have some tools for decreasing the risk of biases in its approach to deconstruct content in smaller parts.

Content analysis and thematic analysis are considered to have similar challenges to define and organise data. Defining categories is considered a challenge in both types of analyses. It is also considered a challenge to distinguish between codes and themes and how to handle potential overlap of codes and themes. It is also considered a risk that the analyses will end up coding everything and produce too many categories. Sticking to the research questions in the analyses is considered important to prevent this from happening.

4.2 Differences between content analysis and thematic analysis

A difference between content analysis and thematic analysis is that they are considered to be a choice between wide application and deep analysis. Content analysis is considered wider in its application, compared to thematic analysis, due to its connection to quantitative research. Because of this, content analysis can be used for both qualitative research and quantitative research, and anything between. It is further considered to be better suited for handling larger sets of data, compared to thematic analysis. Thematic analysis on the other hand is considered to support deeper analysis, or immersion, compared to content analysis, which allows it to bring "more in-depth understanding" of the studied phenomenon. This is connected to that thematic analysis is considered a "purely qualitative" analysis compared to content analysis. Yet within qualitative research, thematic analysis is considered a "flexible method for qualitative analysis" that can be "applied to many research designs".

A difference between content analysis and thematic analysis is that they are considered to be a choice between a practical approach and an intuitive approach. Content analysis is considered straightforward for data categorisation and therefore "faster to get started with", compared to thematic analysis. Content analysis is considered to have a "clear epistemology", providing the researcher with a route to follow on how to do research. Related to this, content analysis is considered more "positivistic", compared to thematic analysis, which could be problematic on "an ontological standpoint". Thematic analysis on the other hand is considered intuitive in its approach to data. Related to this, thematic analysis is considered more inductive in its approach by looking at data "from scratch"; and a fast way to learn how to do qualitative analysis for a "developing researcher". While content analysis is considered more time consuming in choosing between all variations and approaches.

5. Discussion

As pointed out by Hsieh and Shannon (2005), a challenge with the conventional content analysis "is that it can easily be confused with other qualitative methods such as grounded theory method (GTM) or phenomenology", and the same must be concluded if compared to inductive thematic analysis. However, a difference with thematic analysis if compared to grounded theory method, where there is a clear resemblance in the analysis process, is that thematic analysis does not have the goal of generating a theory (McLeod, 2001; Braun & Clarke, 2006).

The perception presented above that thematic analysis is considered more inductive in its approach could be compared to how thematic analysis is presented by Braun and Clarke (2012, p. 58): "TA has the ability to straddle three main continua along which qualitative research approaches can be located", with the first continuum set to "inductive versus deductive". This might depend on that the workshop activity on thematic analysis was introduced with instructions for an inductive thematic analysis. However, the main introduction of thematic analysis in the seminar presented the choice of induction or deduction in thematic analysis, and also involved a slide with:

> *"In reality, coding and analysis often uses a combination of both approaches. It is impossible to be purely inductive, as we always bring something to the data when we analyze it, and we rarely completely ignore the semantic content of the data when we code for a particular theoretical construct"* (Braun & Clarke, 2012, p 58.)

Regarding that content analysis is considered more positivistic compared to thematic analysis could probably be explained by the presentation by Braun and Clarke (2006) as a qualitative approach, while content analysis was presented with an origin in quantitative research. On the other hand, the seminar presentation also comprised Boyatzis (1998) idea that thematic analysis also could be used to transform qualitative data into a quantitative form for further statistical analyses. The reversed transformation order than the one from quantitative to qualitative described in the process of a summative content analysis.

What could be added for a richer discussion on philosophical stances in future seminars is to elaborate on the inductive vs. deductive continuum. Moreover, this should be related to the two other continua brought up by Braun and Clarke (2012), with an experiential versus critical orientation to data, and an essentialist versus constructionist theoretical perspective in a thematic analysis. According to Braun and Clarke (2012, p. 59), "Deductive TA is often critical in its orientation and constructionist in its theoretical framework, examining how the world is put together (i.e., constructed) and the ideas and assumptions that inform the data gathered.", while an inductive thematic analysis relates more to the other ends of the continua, and with the less constructivist assumption of a knowable world.

6. Conclusion

As discussed above there are both similarities and differences between content analysis and thematic analysis. The chosen division of content analysis into three different approaches can explain some of the participants' perceptions in comparisons and discussions. There is a clear similarity between the conventional content analysis and an inductive thematic analysis, while the comparison of a thematic analysis and a summative content analysis shows a fundamental difference. However, there is at the same time such a rudimentary difference between the conventional content analysis and the summative content analysis that could explain the ambiguity in the use of the term content analysis. To more thoroughly present these different approaches to qualitative analysis would facilitate for doctoral students in the inception phase of their doctoral studies. This could be considered an important early step in a qualitative research career, considering that both thematic analysis and content analysis are seen as the common foundation for the use of more complex qualitative analysis methods.

7. Limitations and future research

Some of the limitations of this study is that it only examines the doctoral student perspective, and that the data collection occurs after presentation, discussion, and workshops on the studied methods. As discussed in the paper, participant reflections are certainly affected by the presentations, discussions and workshops conducted at the webinars. With that said, it would be interesting to examine how doctoral students apply different methods for qualitative analysis in research papers. Which are the characteristics of different methods for

qualitative analysis? And do they differ from the use by more senior researchers? The authors further plan to examine the impact of different analyses approaches and theory on the results of a study, by conducting two separate analyses (deductive content analysis and inductive thematic analysis) on the same data set.

References

Bailey, L. F. (2014). The origin and success of qualitative research. *International Journal of Market Research*, *56*(2), 167-184.

Bernard, H. R. (2017). *Research methods in anthropology: Qualitative and quantitative approaches*. Rowman & Littlefield.

Blair, E. (2015). A reflexive exploration of two qualitative data coding techniques. *Journal of Methods and Measurement in the Social Sciences, 6*(1), 14-29.

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*(2), 77-101.

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 57–71). American Psychological Association. https://doi.org/10.1037/13620-004

Braun, V., Clarke, V., & Hayfield, N. (2019). 'A starting point for your journey, not a map': Nikki Hayfield in conversation with Virginia Braun and Victoria Clarke about thematic analysis. Qualitative Research in Psychology, 1-22.

Drisko, J. W. (2005). Writing up qualitative research. *Families in Society, 86*(4), 589-593.

Graneheim, U. H., Lindgren, B. M., & Lundman, B. (2017). Methodological challenges in qualitative content analysis: A discussion paper. *Nurse education today*, *56*, 29-34.

Grossman, S., & Cohen, A. (2017). Contributions to a gestalt quantitative research tradition: Establishing the Gestalt mental status exam. *Gestalt Journal of Australia and New Zealand*, *13*(2), 29-44.

Holsti, O. R. (1969). Content analysis for the social sciences and humanities. *Reading. MA: Addison-Wesley*

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, *15*(9), 1277-1288.

Jeřábek, H. (2001). Paul Lazarsfeld—The founder of modern empirical sociology: A research biography. *International Journal of Public Opinion Research*, *13*(3), 229-244.

Kondracki, N. L., Wellman, N. S., & Amundson, D. R. (2002). Content analysis: Review of methods and their applications in nutrition education. *Journal of nutrition education and behavior*, *34*(4), 224-230.

Larmor, J. (1934). Thomas Young. *Nature*, *133*(3356), 276-279.

McCusker, K., & Gunaydin, S. (2015). Research using qualitative, quantitative or mixed methods and choice based on the research. Perfusion, 30(7), 537-542.

McLeod, J. (2001). *Qualitative research in counselling and psychotherapy*. Sage.

Norton, L. (2009). *Action research in teaching and learning: A practical guide to conducting pedagogical research in universities*. Routledge.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International journal of qualitative methods, 16*(1

O'Dwyer, L. M., & Bernauer, J. A. (2013). *Quantitative research for the qualitative researcher*. SAGE publications.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. SAGE Publications, inc.

Polkinghorne, D. E. (2010). Qualitative research. In *Handbook of clinical psychology competencies*.

Potter, W. J., & Levine-Donnerstein, D. (1999). Rethinking validity and reliability in content analysis.

Roller, M. R. (2019). *A quality approach to qualitative content analysis: Similarities and differences compared to other qualitative methods*. SSOAR-Social Science Open Access Repository.

Rosengren, K. E. (1981). Advances in Scandinavia content analysis: An introduction. *Advances in content analysis*, *9*, 9-19.

Sumpter, R. S. (2001). News about news: John G. Speed and the first newspaper content analysis. *Journalism History*, *27*(2), 64-72.

Thorne, S. (2000). Data analysis in qualitative research. *Evidence-based nursing*, *3*(3), 68-70.

Walia, R. (2015). A saga of qualitative research. *Social Crimonol*, *5*(2), 124.