The qualitative content analysis process

Satu Elo & Helvi Kyngäs

Accepted for publication 22 November 2007

Correspondence to S. Elo:
e-mail: satu.elo@oulu.fi

Satu Elo PhD RN
Senior Lecturer
Department of Nursing and Health Administration, University of Oulu, Finland

Helvi Kyngäs PhD RN
Professor
Department of Nursing and Health Administration, University of Oulu, Finland

doi: 10.1111/j.1365-2648.2007.04569.x

Abstract
Title. The qualitative content analysis process.
Aim. This paper is a description of inductive and deductive content analysis.
Background. Content analysis is a method that may be used with either qualitative or quantitative data and in an inductive or deductive way. Qualitative content analysis is commonly used in nursing studies but little has been published on the analysis process and many research books generally only provide a short description of this method.
Discussion. When using content analysis, the aim was to build a model to describe the phenomenon in a conceptual form. Both inductive and deductive analysis processes are represented as three main phases: preparation, organizing and reporting. The preparation phase is similar in both approaches. The concepts are derived from the data in inductive content analysis. Deductive content analysis is used when the structure of analysis is operationalized on the basis of previous knowledge.
Conclusion. Inductive content analysis is used in cases where there are no previous studies dealing with the phenomenon or when it is fragmented. A deductive approach is useful if the general aim was to test a previous theory in a different situation or to compare categories at different time periods.

Keywords: concept formation, content analysis, deductive, inductive, nursing, research methods

Introduction
There are numerous approaches for analysing qualitative data. Content analysis is used in many studies in nursing and it has a long history. Even if qualitative content analysis is generally used in nursing studies little has been published on how to apply the method. Information retrieval on the CINAHL and Ovid Medline(r) databases showed that content analysis was often presented as a keyword but not so often in the title. Table 1 shows how the number of papers for which content analysis is given as a keyword has increased, and most of these (70%) have been published in the 21st century. The total number of methodological papers concerning content analysis [CINAHL and Ovid Medline(r) published from 1988 to 2005 was only 15 (Table 2)].

The aim of this paper was to describe inductive and deductive content analysis based on theoretical knowledge and to highlight some examples from studies using content analysis, mainly based on content analysis of the methodological papers listed in Table 1.

Background
Content analysis is a method of analysing written, verbal or visual communication messages (Cole 1988). It was first used as a method for analysing hymns, newspaper and magazine
articles, advertisements and political speeches in the 19th century (Harwood & Garry 2003). Today, content analysis has a long history of use in communication, journalism, sociology, psychology and business, and during the last few decades its use has shown steady growth (Neundorf 2002).

Content analysis as a research method is a systematic and objective means of describing and quantifying phenomena (Krippendorff 1980, Downe-Wamboldt 1992, Sandelowski 1995). It is also known as a method of analysing documents. Content analysis allows the researcher to test theoretical issues to enhance understanding of the data. Through content analysis, it is possible to distil words into fewer content-related categories. It is assumed that when classified into the same categories, words, phrases and the like share the same meaning (Cavanagh 1997).

Content analysis is a research method for making replicable and valid inferences from data to their context, with the purpose of providing knowledge, new insights, a representation of facts and a practical guide to action (Krippendorff 1980). The aim is to attain a condensed and broad description of the phenomenon, and the outcome of the analysis is concepts or categories describing the phenomenon. Usually the purpose of those concepts or categories is to build up a model, conceptual system, conceptual map or categories. The researcher makes a choice between the terms ‘concept’ and ‘category’ and uses one or the other (Kyngäs & Vanhanen 1999). For example, if the purpose of the study is to develop a theory, it is recommended that the term ‘concept’ be used as a proxy for ‘category’. However, in this paper, when describing the analysis process, we use the term ‘category’ because this is mostly used in literature.

Despite criticism, content analysis has an established position in nursing research and offers researchers several major benefits. One of these is that it is a content-sensitive method (Krippendorff 1980), and another is its flexibility in terms of research design (Harwood & Garry 2003). It is also much more than a naive technique that results in a simplistic description of data (Cavanagh 1997) or a counting game (Downe-Wamboldt 1992). Concept analysis can be used to develop an understanding of the meaning of communication (Cavanagh 1997) and to identify critical processes (Lederman

Table 1 Journal articles in CINAHL and Ovid Medline(r) databases using content analysis as a keyword

<table>
<thead>
<tr>
<th>Journal articles</th>
<th>CINAHL n</th>
<th>Ovid Medline(r) n</th>
</tr>
</thead>
<tbody>
<tr>
<td>As a keyword</td>
<td>7902</td>
<td>59,119</td>
</tr>
<tr>
<td>As a keyword and concerning:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychiatry/psychology</td>
<td>713</td>
<td></td>
</tr>
<tr>
<td>Gerontologic care</td>
<td>441</td>
<td></td>
</tr>
<tr>
<td>Public health</td>
<td>389</td>
<td></td>
</tr>
<tr>
<td>Nursing education</td>
<td>314</td>
<td></td>
</tr>
<tr>
<td>Advanced nursing practice</td>
<td>310</td>
<td></td>
</tr>
<tr>
<td>Evidence-based practice</td>
<td>176</td>
<td></td>
</tr>
<tr>
<td>Occupational therapy</td>
<td>165</td>
<td></td>
</tr>
<tr>
<td>Critical care</td>
<td>148</td>
<td></td>
</tr>
<tr>
<td>Quality assurance</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>Informatics</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>Emergency care</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>Home health care</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>Hospice/palliative care</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Nursing administration</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Published in

2000–2007 5582 19,784
2007 (January–August) 278 1312
2006 962 3096
2005 980 3156
2004 868 2911
2003 818 2536
2002 642 2221
2001 540 2242
2000 494 2310
1990–2000 2804 23,400
1990–1995 902 11,038
1995–2000 2180 14,318

*Information not available.

Table 2 Number of journal articles with content analysis in the title

<table>
<thead>
<tr>
<th>Journal articles</th>
<th>CINAHL n</th>
<th>Ovid Medline(r) n</th>
</tr>
</thead>
<tbody>
<tr>
<td>In a title (content anal$)</td>
<td>178</td>
<td>496</td>
</tr>
<tr>
<td>Published in 2000–2007</td>
<td>100</td>
<td>196</td>
</tr>
<tr>
<td>2006</td>
<td>16</td>
<td>30</td>
</tr>
<tr>
<td>2005</td>
<td>15</td>
<td>27</td>
</tr>
<tr>
<td>2004</td>
<td>14</td>
<td>27</td>
</tr>
<tr>
<td>2003</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>2002</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>2001</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>2000</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>Published in 1990–1995</td>
<td>33</td>
<td>112</td>
</tr>
<tr>
<td>Published in 1995–1999</td>
<td>38</td>
<td>70</td>
</tr>
</tbody>
</table>

© 2007 The Authors. Journal compilation © 2007 Blackwell Publishing Ltd
Process of analysis

Two approaches

Content analysis is a method that may be used with either qualitative or quantitative data; furthermore, it may be used in an inductive or deductive way. Which of these is used is determined by the purpose of the study. If there is not enough former knowledge about the phenomenon or if this knowledge is fragmented, the inductive approach is recommended (Lauri & Kyngäs 2005). The categories are derived from the data in inductive content analysis. Deductive content analysis is used when the structure of analysis is operationalized on the basis of previous knowledge and the purpose of the study is theory testing (Kyngäs & Vanhanen 1999). An approach based on inductive data moves from the specific to the general, so that particular instances are observed and then combined into a larger whole or general statement (Chinn & Kramer 1999). A deductive approach is based on an earlier theory or model and therefore it moves from the general to the specific (Burns & Grove 2005). These approaches have similar preparation phases.

Preparation phase

Both inductive and deductive analysis processes are represented as three main phases: preparation, organizing and reporting (Figure 1). Despite this, there are no systematic rules for analysing data; the key feature of all content analysis is that the many words of the text are classified into much smaller content categories (Weber 1990, Burnard 1996). The preparation phase starts with selecting the unit of analysis (McCain 1988, Cavanagh 1997, Guthrie et al. 2004). This can be a word or a theme (Polit & Beck 2004). Deciding on what to analyse in what detail and sampling considerations are important factors before selecting the unit of analysis (Cavanagh 1997). The sample must be representative of the universe from which it is drawn (Duncan 1989). Probability or judgement sampling is necessary when a document is too large to be analysed in its entirety (GAO 1996). The aim is to become immersed in the data, which is why the written material is read through several times (Burnard 1991, Polit & Beck 2004). No insights or theories can spring forth from the data without the researcher becoming completely familiar with them (Polit & Beck 2004). After making sense of the data, analysis is conducted using an inductive or deductive approach (Kyngäs & Vanhanen 1999).

Inductive content analysis

If the researcher has chosen to use inductive content analysis, the next step is to organize the qualitative data. This process includes open coding, creating categories and abstraction. Open coding means that notes and headings are written in the text while reading it. The written material is read through again, and as many headings as necessary are written down in the margins to describe all aspects of the content (Burnard 1991, 1996, Hsieh & Shannon 2005). The headings are...
Figure 1 Preparation, organizing and resulting phases in the content analysis process.
collected from the margins on to coding sheets (Cole 1988, Downe-Wamboldt 1992, Dey 1993) and categories are freely generated at this stage (Burnard 1991).

After this open coding, the lists of categories are grouped under higher order headings (McCain 1988, Burnard 1991). The aim of grouping data was to reduce the number of categories by collapsing those that are similar or dissimilar into broader higher order categories (Burnard 1991, Downe-Wamboldt 1992, Dey 1993). However, Dey (1993) points out that creating categories is not simply bringing together observations that are similar or related; instead, data are being classified as ‘belonging’ to a particular group and this implies a comparison between these data and other observations that do not belong to the same category. The purpose of creating categories is to provide a means of describing the phenomenon, to increase understanding and to generate knowledge (Cavanagh 1997). When formulating categories by inductive content analysis, the researcher comes to a decision, through interpretation, as to which things to put in the same category (Dey 1993).

Abstraction means formulating a general description of the research topic through generating categories (Robson 1993, Burnard 1996, Polit & Beck 2004). Each category is named using content-characteristic words. Subcategories with similar events and incidents are grouped together as categories and categories are grouped as main categories (Dey 1993, Robson 1993, Kyngäs & Vanhanen 1999). The abstraction process continues as far as is reasonable and possible. An example of the abstraction process is shown in Figure 2. Inductive content analysis has been used, for example, in studies of the environment that supports well-being in older people (Juvani et al. 2005), support networks of adolescents with chronic disease (Kyngäs 2004) and men’s experiences of heart failure (Europe & Tyni-Lenne 2004).

**Deductive content analysis**

Deductive content analysis is often used in cases where the researcher wishes to retest existing data in a new context (Catanzaro 1988). This may also involve testing categories, concepts, models or hypotheses (Marshall & Rossman 1995). If a deductive content analysis is chosen, the next step is to develop a categorization matrix (Table 3) and to code the data according to the categories (Table 4). In deductive content analysis, either a structured or unconstrained matrix of analysis can be used, depending on the aim of the study (Kyngäs & Vanhanen 1999). It is generally based on earlier work such as theories, models, mind maps and literature reviews (Sandelowski 1995, Polit & Beck 2004, Hsieh & Shannon 2005).

After a categorization matrix has been developed, all the data are reviewed for content and coded for correspondence with or exemplification of the identified categories (Polit & Beck 2004). When using an unconstrained matrix, different categories are created within its bounds, following the principles of inductive content analysis. If the matrix is

---

**Figure 2** An example of the abstraction process.
structured, only aspects that fit the matrix of analysis are chosen from the data (Patton 1990, Sandelowski 1993, 1995). This can also be called testing categories, concepts, models or hypotheses (Marshall & Rossman 1995). When using a structured matrix of analysis, it is possible to choose either only the aspects from the data that fit the categorization frame or, alternatively, to choose those that do not. In this way, aspects that do not fit the categorization frame can be used to create their own concepts, based on the principles of inductive content analysis.

The choice of method depends on the aim of the study (Catanzaro 1988, Robson 1993, Marshall & Rossman 1995). For example, categories used to describe the different meanings that the concept of illness has for teenagers with diabetes could be used as a frame of categorization when analysing interviews with teenagers suffering from rheumatoid arthritis. In this case, the aim would be to test whether illness has the same meaning for teenagers suffering from rheumatoid arthritis as for those with diabetes (Kyngäs & Vanhanen 1999). Deductive content analysis has been used for example in studies of patients’ readiness for dietary change (Kasila et al. 2003), self-care of older people (Backman & Hentinen 2001) and mental health care (Latvala et al. 2000).

Trustworthiness

The analysis process and the results should be described in sufficient detail so that readers have a clear understanding of how the analysis was carried out and its strengths and limitations (GAO 1996). This means dissection of the analysis process and the validity of results. Elements of validity in content analysis are universal to any qualitative research design there are additional factors to take into consideration when reporting the process of analysis and the results.

The results are described contents of the categories, i.e. the meanings of the categories. The content of the categories is described through subcategories (Marshall & Rossman 1995). Creating categories is both an empirical and a conceptual challenge, as categories must be conceptually and empirically grounded (Dey 1993). Successful content analysis requires that the researcher can analyse and simplify the data and form categories that reflect the subject of study in a reliable manner (Kyngäs & Vanhanen 1999). Credibility of research findings also deals with how well the categories cover the data (Graneheim & Lundman 2004).

It is important to make defensible inferences based on the collection of valid and reliable data (Weber 1990). To increase the reliability of the study, it is necessary to demonstrate a link between the results and the data (Polit & Beck 2004). This is why the researcher must aim at describing the analysing process in as much detail as possible when reporting the results. Appendices and tables may be used to demonstrate links between the data and results. To facilitate transferability, the researcher should give a clear description of the context, selection and characteristics of participants, data collection and process of analysis (Graneheim & Lundman 2004). Demonstration is needed of the reliability of the findings and interpretations to enable someone else to follow the process and procedures of the inquiry.

Authentic citations could also be used to increase the trustworthiness of the research and to point out to readers from where or from what kinds of original data categories are formulated (Patton 1990, Sandelowski 1993). The researcher should make sure that informants are not identified by quotes from the data (Ford & Reutter 1990). There has been some debate about the suitable amount of authentic citation. In our experience, if there are more citations than authorial text, then the analysis process is usually incomplete.

The internal validity of content analysis can be assessed as face validity or by using agreement coefficients (Weber 1990).

<table>
<thead>
<tr>
<th>Table 3 An example from a categorization matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>What kind of mental well-being threats does diabetes have for adolescents?</td>
</tr>
<tr>
<td>Dependence</td>
</tr>
<tr>
<td>Dependence on parents</td>
</tr>
<tr>
<td>Dependence on insulin</td>
</tr>
<tr>
<td>Dependence on nurses</td>
</tr>
<tr>
<td>Dependence on physicians</td>
</tr>
<tr>
<td>Dependence of regular daily life</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4 An example of coding the data to the categorization matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>What kind of mental well-being threats does diabetes have for adolescents?</td>
</tr>
<tr>
<td>Dependence</td>
</tr>
<tr>
<td>Dependence on parents</td>
</tr>
<tr>
<td>Dependence on insulin</td>
</tr>
<tr>
<td>Dependence on nurses</td>
</tr>
<tr>
<td>Dependence on physicians</td>
</tr>
<tr>
<td>Dependence of regular daily life</td>
</tr>
</tbody>
</table>
The qualitative content analysis process

What is already known about this topic

- Content analysis is used in many studies in nursing and it has a long history, but there has been little discussion of its use.
- Content analysis may be used with either qualitative or quantitative data and in an inductive or deductive way.
- Content analysis is a flexible method and there are no simple guidelines for data analysis, which makes it challenging for the researcher.

What this paper adds

- Content analysis is extremely well-suited to analysing data on the multifaceted, sensitive phenomena characteristic of nursing.
- The use of inductive content analysis is recommended when there are no previous studies dealing with the phenomenon or when knowledge is fragmented.
- A deductive approach is useful if the aim is to test an earlier theory in a different situation or to compare categories at different time periods.

However, there are various opinions about seeking agreement (Graneheim & Lundman 2004), because each researcher interpret the data according to their subjective perspective and co-researchers could come up with an alternative interpretation (Sandelowski 1995). Content validation requires the use of a panel of experts to support concept production or coding issues. Graneheim and Lundman (2004) defend the value of dialogue among co-researchers to agree the way in which the data are labelled.

Discussion

Content analysis does not proceed in a linear fashion and is more complex and difficult than quantitative analysis because it is less standardized and formulaic (Polit & Beck 2004). There are no simple guidelines for data analysis: each inquiry is distinctive, and the results depend on the skills, insights, analytic abilities and style of the investigator (Hoskins & Mariano 2004). One challenge of content analysis is the fact that it is very flexible and there is no simple, ‘right’ way of doing it. Researchers must judge what variations are most appropriate for their particular problems (Weber 1990), and this makes the analysis process most challenging and interesting. An enormous amount of work is required during the process (Polit & Beck 2004).

The analysis process has been little discussed in literature, possibly because content analysis has in the past been criticized for being an overly simple method. The most paralysing moment in conducting content analysis is simply getting started. Researchers should allow themselves simply to read through each interview as many times as necessary to apprehend its essential features, without feeling pressured to move forward analytically (Sandelowski 1995). They often see the beginning of the categorization phase as chaotic, because at that point they possess several, seemingly unconnected, pieces of information (Backman & Kyngäs 1998).

Another problem may be that narrative material is generally not linear, and paragraphs from transcribed interviews may contain elements relating to several categories (Dey 1993, Polit & Beck 2004). According to Glaser (1978), tolerance of feeling uncertain is required. To resolve the situation, it is also necessary to be prepared to go back to the data to check the reliability of the categories. The assumption that content analysis is an easy method can mislead researchers, and unexpected difficulties may arise during the analysing process (Glaser 1978).

Regardless of the ‘quality’ of qualitative data, its sheer quantity can be daunting, if not overwhelming. Depending on the level of detail, six interviews may easily run to 50–100 single-spaced pages of transcribed text (Miles & Huberman 1994). Hundreds of pages of data can lead the researcher to think that it cannot be managed. In addition, many interesting points that are not related to the topic under study often come up when analysing the data. In that case, keeping the research question in mind is an essential aspect of content analysis. Even in the middle of chaos, the researcher must always be able to go back to the research tasks and only look for units of analysis that have relevance to them.

Reporting the study and presenting its results can also be challenging, as the results are formed through a process comprising a number of phases. Researchers often wish for more detailed instructions on how to carry out content analysis, but those who have already gone through the process know that describing the analysis is often one of the most challenging phases of the study. It is possible to describe some parts of this process in great detail, but other parts – such as the researcher’s own actions and insights – may be difficult to put into words (Backman & Kyngäs 1998). Qualitative data analysis software programs can be used to make content analysis more manageable and ordered, and may facilitate new levels of analysis (Gerbic & Stacey 2005).

If qualitative data are compressed too much, the very point of maintaining the integrity of narrative materials during the analysis phase becomes lost. If the conclusions are merely
summarized without including numerous supporting excerpts, the richness of the original data disappears. It is sometimes difficult to give a thorough presentation of the results of qualitative research in a format that is compatible with the space and word limitations in professional journals (Polit & Beck 2004).

Study results developed through content analysis often display signs of an incomplete analysing process. The researcher has not abstracted the data, or has included too many different things in a single category (Dey 1993, Hickey & Kipping 1996). However, an abundance of categories is usually a sign of being unable to categorize the data. When the abstracting process is still ongoing, the results may include categories that do not cancel each other out. In addition, when the saturation of the data is incomplete, it may be difficult to link data items to each other (Patton 1990, Cavanagh 1997). Seemingly simple study results are a sign of incomplete analysis (Weber 1990).

Conclusion

Content analysis is extremely well-suited to analysing the multifaceted, sensitive phenomena characteristic of nursing. An advantage of the method is that large volumes of textual data and different textual sources can be dealt with and used in corroborating evidence. Especially in nursing research, content analysis has been an important way of providing evidence for a phenomenon where the qualitative approach used to be the only way to do this, particularly for sensitive topics. The disadvantage of content analysis relates to research questions that are ambiguous or too extensive. In addition, excessive interpretation on the part of the researcher poses a threat to successful content analysis. However, this applies to all qualitative methods of analysis.

In nursing studies, deductive content analysis has been used less than inductive approach. It is likely that in the future, the use of deductive content analysis will become more common, because inductively built models or concept systems can be complemented, tested and developed further with the aid of deductive analysis.

Author contributions

SE and HK were responsible for the study conception and design and the drafting of the manuscript. SE and HK performed the data collection and data analysis. SE obtained funding and SE and HK provided administrative support. SE and HK made critical revisions to the paper. SE provided statistical expertise. HK supervised the study.

References


JAN: RESEARCH METHODOLOGY

The qualitative content analysis process


