On the Pragmatics of Qualitative Assessment

Designing the Process for Content Analysis

Jan Schilling

Department of Industrial and Organizational Psychology, Institute of Psychology,
Technical University Aachen, Germany

Abstract. Using the example of a project on the assessment of implicit leadership theories, this article aims to describe qualitative content analysis as a systematic, rule-based process of analyzing verbal and textual data (e.g., interviews, group discussions, documents). Steps and typical problems in the qualitative assessment process are addressed and guidelines for decision and action presented. The steps include transcription of interview tapes into raw data, condensing and structuring the data, building and applying a category system, displaying data and results for concluding analyses and interpretation. Necessary checks for securing the quality of the assessment are shown for each step. Ideas for the combination of qualitative and quantitative analyses are presented, and applications of qualitative content analysis in the field of psychological assessment discussed.

Keywords: qualitative assessment, assessment process, content analysis, qualitative methods

Introduction

While the use of qualitative methods in psychological assessment is widespread (i.e., structured interviews, action research, in-depth interviews), there often remain uncertainties about their exact application. Of course, books and papers on qualitative research provide many valuable ideas, experiences, and hints regarding these and comparable questions. However, in the face of starting a research or an assessment project they sometimes seem vague and abstract. While it is quite easy to find excellent introductions of theory, methodology, and qualitative data collection (e.g., Neimeyer & Gemignani, 2003), the novice researcher may find it difficult to access what could be called "pragmatic knowledge," especially when it comes to the process of qualitative data analysis and interpretation. Pragmatic knowledge in this context can be understood as established principles, heuristics, and rules guiding the actions and decisions of the researcher during different steps of the assessment process. There are at least two main reasons for the difficulty of defining such pragmatic knowledge in qualitative assessment. First, qualitative research is more or less an umbrella term for many different research traditions (i.e., grounded theory, phenomenology, discourse analysis) with their own theoretical and methodological backgrounds (Creswell, 1998; Denzin & Lincoln, 2000). Hence, it seems not only almost impossible to name common guidelines incorporating all these different approaches but one could also doubt the appropriateness of such an endeavor. Second, many researchers reject the idea that qualitative research can be represented as a linear process of stages and tasks even if circular or iterative procedures are included (e.g., Maxwell, 1998; Miles & Huberman, 1994). In regard to the first argument, this paper will not attempt to develop a process model for all qualitative research traditions but will focus on one special approach: qualitative content analysis (Mostyn, 1985). Qualitative content analysis can be defined as "an approach of empirical, methodological controlled analysis of texts within their context of communication, following content analytic rules and step by step models, without rash quantification" (Mayring, 2000, p. 5). The second argument may be more challenging for the purpose of this paper. Qualitative research requires a flexible, nonsequential approach (Maxwell, 1998). Consequently, its process cannot be pressed into some clear-cut model with distinctive phases but rather has a more complex, idiosyncratic, and fluid structure (Neimeyer & Gemignani, 2003). Of course, it would be misleading to develop a process that could be used like a "recipe." But this does not mean that one should shy away from discussing the assessment process, its steps, and the guiding principles and rules. Following Creswell (1998), the process is described here with the image of a spiral. The process moves in analytic circles from one level to the next (see Figure 1). On the other hand, there is a danger in this argument in favor of flexibility. It could serve as a "protection from rigorous, critical standards that should be applied to any enterprise concerned to sort 'fact' from 'fancy'" (Silverman, 2000, p. 12). Especially the data analysis proce-

European Journal of Psychological Assessment 2006; Vol. 22(1):28–37
DOI 10.1027/1015-5759.22.1.28

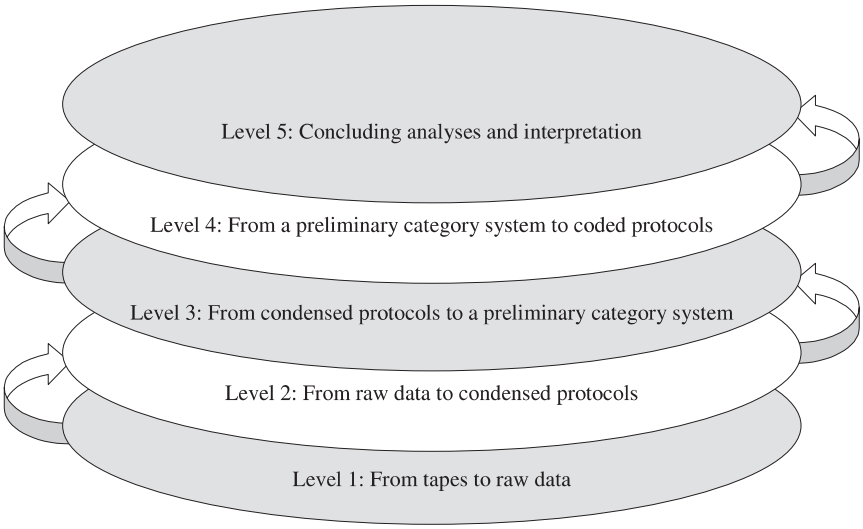

Figure 1. The qualitative content analysis spiral.

dures ("the dirty work;" Conger, 1998) are often missed out or described vaguely in qualitative studies. However, if qualitative research wants to break free from the stigma of being "not scientific," "arbitrary," or "subjective," it has to follow systematic and transparent ways for data collection, analysis, and reporting (Bachiochi & Weiner, 2002; Creswell, 1998). Therefore, the aim of this paper is to give guidance on how to design the process of qualitative content analysis, not as the "one right way" (Tesch, 1990), but as a stimulation for the qualitative researcher.

The Example: Assessing Implicit Leadership Theories via Semistructured Interviews

The development of the process model for qualitative content analysis is mainly based on the experiences in a specific research project on implicit leadership theories. Therefore, a brief description of the aims and background of that project seems a necessary basis for the following discussion (the concrete results are of minor importance for the purpose of this paper; see Schilling, 2001). The aim of the study was to analyze the leadership culture in two divisions of a large-scale enterprise. For that purpose, content and structure of the everyday conceptions of leadership of the top- and middle-managers (42 interviewees in total) were analyzed. These implicit leadership theories are a special form of cognitive scheme, which, in analogy to scientific theories, can be understood as a network of concepts and relations. While leadership research has been primarily focused on styles, behaviors, situations, and effectiveness, the study of implicit leadership theories focuses on the inner perspectives of the phenomenon. What aspects of leadership, of its conditions and consequences, are most important for corporate leaders? What typical similarities and differences can be found in personal conceptions of leadership? Are there interrelations between the complexity of implicit leadership theories and personal characteristics of the leaders (e.g., age, leadership experience)? Besides these guiding questions on leadership, its antecedents and consequences, the interviewees were asked about themselves (e.g., vocational education) and their position (e.g., percent of working time used for leadership, leadership experience). The interviews lasted about 1 h 45 min on average and were tape-recorded with the consent of the interview partner. As this example stems from the area of organizational assessment, it should be made clear that qualitative content analysis can also be used in psychotherapy research (e.g., Frommer, Langenbach & Streeck, 2004) or educational research (e.g., Faux, 2000).

A Systematic and Rule-Based Process of Qualitative Content Analysis

Tesch (1990) distinguishes between the so-called linguistic tradition, which treats texts as an object of analysis itself, and the sociological (or at a micro-level: psychological) tradition, which is interested in the human experiences that became manifest in the texts. In this sense, this work is focused on the sociological tradition and aims at presenting phases and rules for qualitative content analysis. It seems important to add one point before the discussion on the process starts. One might criticize that just to look at the data analysis process is insufficient as such a perspective seems to ignore that the choice of method should always depend on what we are trying to find out (Silverman, 2000). A conceptual framework (in our case: the concept of implicit leadership theories) and the research or diagnostic questions derived from it should always be the starting point for the analytic process. Therefore, it will be neces-

sary to discuss each step (organized in different levels of the spiral) in the data analysis process from this perspective. However, as will be shown from a pragmatic point of view, there are concerns and problems in the course of qualitative content analysis that transcend the boundaries of different theoretical perspectives.

First Level: From Tapes via Transcripts to Raw Data

It may not be so common in psychological assessment, but especially for research purposes most qualitative researchers (e.g., Silverman, 2000) recommend tape recording interviews to make sure that their content is exactly retained. The process of transcribing tapes to written texts may seem to be a rather mechanical activity. Nevertheless, it is important to define explicit rules for this phase. First, the rather simple formal aspects of the transcript should be determined (program, font, size, margins). Second, the following questions arise: (1) Should dialects or slips of the tongue be preserved, ignored, or respectively corrected (content-focused)? If dialects are to be transcribed, it should be defined (e.g., by some examples) how the terms should be spelled in order to provide an intelligible transcript. (2) Should observations during the interview (as recorded in a written protocol), sounds (like "uhs" or "ers") as well as audible behavior (like coughing or drumming of the fingers) be transferred or not (speech-focused)? (3) Should all questions of the interviewer or only the main questions from the interview guideline (answer-focused) be transcribed? For example, a specific question-answer-sequence (Q: "How long have you been in this job?" A: "Five years.") could be transcribed in the form of an answer ("I have been in this job for five years.").

The restrictions to content and speech are by no means intended to express that pauses, slips of the tongue, or sounds are not worth analyzing. On the contrary, for instance pauses of the interviewee could add valuable information to better understand the content of his answers (Mostyn, 1985; Silverman, 2000). As always, it depends on the research questions, although it should be noted that these aspects of human communication are not interpreted easily and unambiguously. Focusing on answers has the pragmatic advantage of giving the transcribed texts a more coherent form, making the forthcoming steps of analysis easier. Of course, it should be noted that there is a danger in only transcribing the main guiding questions. For example, it will not be easy to control if the interviewer behavior changed over time (e.g., because of growing experience with the topic), differed between interviewers (in case more than one interviewer is involved), or broke the defined rules (e.g., by posing leading questions). It is recommended to control for these concerns at least by listening to a random sample from the tapes and critically searching for such incidents. Also, if the researcher is interested in the exact diction of the interviewee, then focusing on the answers would pose a problem as the texts will be a mixture of interviewer and interviewee speech. The researcher does not need to make all these decisions in advance. Before writing down the exact transcription rules a general review of the material (listening to some or all of the tapes) might prove helpful to obtain an idea of the overall data (Creswell, 1998; Tesch, 1990). After a first check to secure data quality all interview texts are made anonymous by replacing names of people and institutions with descriptive terms (e.g., "my boss" instead of "Mr. Bauer"). Of course, if the researcher is interested in comparing, for instance, the opinions of different interviewees toward a certain person or institution, a coding scheme for the persons or institutions named needs to be developed and applied here.

Second Level: From Raw Data to Condensed Records

Before starting the analysis of the raw data, some definitions and rules guiding the analysis have to be worked out (Mayring, 1994; Schilling, 2001):

1. *Describing the situation of text production.* As a basis for further analysis, the situation of the data generation should be described (e.g., in the form of a contact or document summary sheet; Miles & Huberman, 1994). Basic questions for our example were: Who are the interviewees (e.g., position in the firm)? What is their relationship with the interviewer? When (e.g., during working time) and under which circumstances (e.g., during a process of downsizing in their company) were the interviews done? Where were the interviews done (e.g., in the office of the interviewees)? In which context where the interviews done (e.g., as part of a management learning project)? Were there any disturbances or outstanding reactions from (some of) the interviewees (e.g., comments after the "official" interview ended)? As some or all these questions can be very important to the interpretation of the text, the researcher should answer them carefully and in advance of text analysis.
2. *Directing the analysis.* Based on the communication model of Mayring (1994), the question arises: Is the researcher interested in the topic, the communicator himself, his sociocultural background, the situation of text production, the text itself, or the effect of the message? In the present case, the communicators (the interviewees) were at the center of interest: their experiences, cognitions, and evaluations concerning leadership (i.e., their implicit leadership theories). After that, the researcher has to derive the main dimensions for the categorization of the material from his research questions (Mayring, 1987). In the case of the study on implicit leadership theories those were "Leadership" (with the subdimensions "Leadership in general," "Positive leadership," and "Negative leadership"), "Attributed ante-

cedents of leadership," and "Attributed consequences of leadership." For these dimensions, categories had to be either developed or deductively applied (see Step 3).

3. *Defining the units of analysis.* As the extent of analytic units may differ within the data (from a single word to more than a sentence), the researcher is often left alone with the advice of "using some judgment" in deciding what a meaningful unit is (Locke, 2002). Meaningful unit in this sense would mean a "segment of text that is comprehensible by itself and contains one idea, episode, or piece of information" (Tesch, 1990, p. 116). While this is certainly true, it seems mandatory (for the quality of the assessment) to define at least the boundaries of unitizing. Three kinds of units can be differentiated (cf. Mayring, 1994): (a) what is the smallest text component to be categorized (coding unit: e.g., single word, half-sentence, full sentence, paragraph or complete text; in the present example: word), (b) what is the biggest text component to be categorized (context unit: again single word, half sentence, full sentence, paragraph or complete text; in the present example: full sentence) and (c) in which order should the text components be analyzed (sequencing unit: cross-question, that is interview after interview, or cross-interview, i.e., question after question). The cross-question strategy should be chosen when the guiding questions are overlapping (i.e., answers to one guiding question are not unlikely to occur in the course of another question) and aim at the same topic from different directions (that was the case in the present example). By that, the researcher gets an idea of the full complexity of each interview. If the questions are rather distinct from each other and/or focus on different topics, the cross-interview procedure is helpful to give an impression of the complexity of possible answers from different interviewees.

After these initial definitions, the process of *condensing content analysis* can begin. The aim of this step is to reduce the material while preserving the essential contents. If the strategy of answer-focusing has been applied, the raw data now has the form of texts (one interview corresponds to one text) structured by the guiding questions. The next step is to reduce the material to its basic content (called *"paraphrasing*") by deleting all the words that are not necessary to understand the statement and transforming the sentences into a short form (see Table 1 for an example).

The remaining statements are now *generalized* and *reduced*. This step can best be explained by the examples of our text passage (see Table 2). First of all, especially with regard to possible quantitative analysis later, it is important to make a decision on how to deal with conjunctions (linguistic forms that joins together sentences, clauses, phrases, or words) within the text. Of the different forms of conjunctions, the following are the most important for our discussion: copulative (e.g., "and"), disjunctive (e.g., "or"), restrictive (e.g., "but"), modal (e.g., "by"), temporal (e.g., "after") and causal (e.g., "because"). In the example of our study, it was decided to dissolve all of these relationships in order to get a realistic picture of the complexity of their implicit leadership concepts. For instance, if someone said: "Leadership means setting clear goals. It is also important to set realistic goals," this would result in two statements to be categorized: "setting clear goals" and "setting realistic goals." Now, imagine another interviewee saying: "Leadership to me means setting clear and realistic goals." It would be incorrect to take this as one statement and the other case as two statements. So the generalization here means to break down the statements into their basic parts in order to have a consistent data basis. This may sound easier than it actually is in practice even if the rules are clearly defined. To give a simple example, the sentence "Leadership means setting clear, realistic goals" also implies a copulative conjunction. But a computer-assisted search for "and" would not find it. A more complex problem is the decision if it makes sense to split up all conjunctions. This gets extremely difficult when regarding such frequent conjunctions like "that" or "what" (e.g., "telling that he did a good job"; "telling what was good"). The rule applied here is that these conjunctions were not divided because they are essential to the most basic understanding of the statements (e.g., a statement like "telling" as an aspect of leadership is extremely unspecific and, therefore, not helpful for further analysis). By this example, it becomes clear that this step is one of high importance and difficulty at the same time. The complexity in the use of human language makes it unlikely that a researcher will be able to reach and apply a "perfect" system of rules. But it is important to be as explicit, consistent, and transparent as possible. One might question if this rigorous and time-consuming approach is really necessary at all times. It is, at least in those cases when the researcher also wants to analyze the *total* frequency

Table 1. Example of paraphrasing texts

Original	Paraphrases
I have to think about this. Motivating people, of course. What else? Well, that is, you know, one thing is, if someone did a good job, you have to tell him, really tell him, that was good. And tell him what exactly was good about it. Then he has, he is really, you surely know that, satisfied with his work and will do, perform better in the future. And if possible do it in front of others, so they will be encouraged too. I have been working in this company for some time and I can tell you: unfortunately, all that is not common practice.	Motivating people. If someone did good job, tell him that and what was good. He is satisfied with his job and will perform better in the future. Tell him in front of others. Others will be encouraged too. All that is not common practice in the company.

of statements that fall into one category and is not only interested in how many of the research subjects used a certain category *at all* (no matter how often). The second aspect of generalization is the decision if the statements should be transformed into one grammatical form. For the example of leadership behaviors, this would mean to standardize the statements into a verb form (e.g., "motivating people") or substantival mode (e.g., "motivation of people"). This standardization of statements proves helpful for coding but should be avoided if one is interested in the exact language of the subjects.

The next step of reducing the material involves two decisions. First of all, all the statements that are not important in regard to the questions to be answered are deleted. In our example, the statement "All that is not common practice in the company" is not selected for further analysis as it is not part of the implicit leadership theory (i.e., what aspects of leadership, its conditions, and consequences are most important for the managers?). As the researcher might be easily criticized for neglecting important aspects of the text, the selection should be carefully based on theoretical considerations. In case of doubt, only statements explicitly connected to the research or assessment interest should be left in the analysis process. The second decision in this selection process concerns deleting those statements that are either recurrent and identical in meaning. For the study on implicit leadership, it was determined to only delete statements if they recurred in the same paragraph and with the same words. A reappearance at a later point of the interview was regarded as an indicator of the importance of the statement. This decision might be criticized as, to some degree, arbitrary, but most important in this process is the transparency and explicitness. A control check completes this second part of the analysis process. The full material should be checked against the original (any relevant statement missing?), both by the researcher (or the research team) and a second person (or team) who should be trained in the process but not involved in the previous steps. If the researcher decided to split up conjunctions in the text, it has to be controlled that the defined rules were kept accurately. By my own experience, this can be an especially time-consuming task.

Third Level: From Condensed to Structured Protocols and a Preliminary Category System

With regard to amount and form, the text material can now be submitted to *structuring content analysis*. Structuring means that each statement is attached to one of the defined dimensions (in the example of implicit leadership theories: "Leadership," "Attributed antecedents of leadership," "Attributed consequences of leadership"). Often, the process of qualitative analysis is only associated with a situation where the researcher has no theory and will generate dimensions and categories from the data. This might be true in some rare cases, but mostly researchers have at least preliminary models guiding their "data-driven" approach. To improve the transparency of the analysis, this preliminary model should be made explicit and used for structuring the material. The strength of the qualitative approach lies in the fact that such a model can be elaborated or changed within the course of the analysis. In some cases, it might even be important to develop a system of dimensions and categories that can be used in future studies and, thereby, be used to compare different studies (see Feixas, Geldschläger, & Neimeyer, 2002, for an example of a category system to analyze personal constructs).

As before, this step has to be controlled in order to ensure the accuracy of the structuring. There are different ways of controlling the data quality. At best, independent raters should be used to check the interrater reliability of the structuring (either all the material or randomly selected samples; Conger, 1998). Alternatively, the statements could be coded and then coded again by the same person (intrarater reliability) after a significant period of time has passed (Erdener & Dunn, 1990). Also possible would be that a pair or a team of researchers discusses samples of the material together to develop a common understanding of the dimensions. Afterwards, they could structure the material independently from each other, compare their results, document and resolve all cases of doubt. At the least, samples of the structuring results of one rater should be checked by another. Cases of doubt should always be checked in regard to the original material and if necessary result in a revision of the dimension definitions (if they are not clear or specific enough).

Table 2. Example of generalizing and reducing the paraphrases

Paraphrases	Generalization und reduction
Motivating people. If someone did good job, tell him that and what was good. He is satisfied with his job and will perform better in the future. Tell him in front of others. Others will be encouraged too. All that is not common practice in the company.	Motivating people. Someone did a good job. Telling that he did a good job. Telling what was good. He is satisfied. He will perform better in the future. Telling him in front of others. Others will be encouraged too.

Fourth Level: From a Preliminary Category System to Coded Protocols

The development and application of a category system lies at the heart of qualitative content analysis (Mostyn, 1985). According to Holsti (1969), there are certain criteria to keep in mind when building or selecting categories. They should reflect the purpose of the research, be exhaustive, and mutually exclusive. Keeping this in mind, there are two basic approaches in the development of a category system (inductive and deductive), which should not be regarded as mutually exclusive but rather as complementary (Tesch, 1990). Within the framework of qualitative approaches, it is often of central interest to build the categories as near to the material as possible. Such inductive category development implies the following steps (cp. Mayring, 2000): (a) defining criteria of selection (i.e., which aspects should be taken into account) and levels of abstraction (i.e., how far away from the text material) for the categories; (b) step-by-step formulation of inductive categories, if necessary subsuming old or formulating new categories (Conger, 1998); (c) revising the categories after 10 to 50 percent of the material has been coded (formative check of reliability): checking the agreement of different raters, discussing cases of doubt and problems with the scope and overlapping of the categories within the research team; (d) checking the coding about two-thirds of the way through the material to prevent drifting into an idiosyncratic sense of what the codes mean (Miles & Huberman, 1994); (e) final working through the material (summative check of reliability: e.g., by checking interrater reliability); (f) interpretation of results and quantitative steps of analysis.

A rule-of-thumb in building categories is to write a formal definition of the working category label when a category contains more than six and fewer than twelve data fragments (Locke, 2002). To enhance the interrater reliability of the categories, independent coders should identify categories separately and then come to an agreement about the final categories and subcategories with the research team (Bachiochi & Weiner, 2002). Whenever possible, one should also test the reliability of the coding instructions with a fresh set of independent raters who should be able to work with the coding instructions as their sole guide (Krippendorff, 1980). Any difficulties with applying the coding instructions should be regarded as valuable hints for problems within the coding scheme or the description of the categories.

To accompany the codes, a list of acronyms for the various categories is established (Conger, 1998). Often, the researcher does not start from scratch but rather has a rough category system (derived from theory and/or prior research on the topic) that he wants to test and refine (e.g., by building subcategories) in confrontation with the text material. In this way, deductive and inductive strategies in the development of a category system are often combined. As Mayring (2000) states, the exact step of *deductive category application*, that is connecting a category system with the object of research, is often poorly described. Starting with a theoretical discussion and explanation of the system, the researcher has to define main and – if necessary – subcategories as well as formulate anchor examples (prototypes) and coding rules. The definition of coding rules aims at demarcating the different categories. An example from our study to delimit the category "Delegating" from the category "Planning and Organizing" would be: If a statement implies the term "delegating" only with "task" (i.e., delegating a task to a subordinate), it should be categorized as "Planning and Organizing." This rule is derived from the definition of the category "Delegating": "Allowing subordinates to have substantial responsibility and discretion in carrying out work activities, handling problems, and making important decisions." Therefore, the essence of the delegation is not to give a task (which would be an example of organizing the work in the own department) but to grant freedom for decision and action.

While it is analytically helpful to divide the development and application of the category system, the discussion should have made clear that in reality these two activities are not sequential, but iterative steps. When it comes to coding the statements, it is of great importance to be careful with superficial resemblance. But coding does not only mean assigning each statement to one content category. That would mean disregarding much information within the text material. For the purpose of further analyses, each statement can also be coded with regard to its context, ranking from simple aspects like the guiding question to which the statement was made (necessary for analyses of typical interview courses and of the usefulness of interview questions; cp. Schilling, 2001) to codes to analyze the argumentation structure of the material (e.g., the co-occurrence of statements in one paragraph).

Also, some checks should be made to secure the quality of the data. Concerning the category codes, some possible ways were already named in the discussion of Step 3. In a formative check of reliability, a pair or team of researchers discusses the material together in order to develop (a common understanding of) the categories. They document and resolve cases of doubt, i.e., all statements that cannot be categorized easily. The rule applied in the current example is to first check the context of the statement: if all other statements in the same argument block fall into the same category, the doubtful statement is also coded that way. Second, if the first rule can not be applied (e.g., because the other statements fall into different categories), the original text is checked to get the sense of the statement. Afterwards, samples of the material are reviewed by independent raters and a first check of interrater reliability is made (formative evaluation of reliability).

Finally, all the statements are coded by independent raters and the interrater reliability of the coding is checked (summative check of reliability). Another very important step is to look over those statements that could not be categorized ("*misfit analysis*"). The residual statements

should be analyzed carefully for their frequency (as an indicator of the inclusiveness of the category system), content (what aspects are possibly missing in the category system), and if there are any systematic regularities (e.g., that all statements came from one or a special group of interviewees). In the example of the study on implicit leadership theories, the misfit analysis showed that only 33 of the statements (0.6%) could not be categorized. Concerning the content, the majority of these statements were formulated in a very abstract way (e.g., "leadership situation" as an antecedent). However, there were also a very small number of statements concerning "leadership of the own person" (self-leadership), a category that proved useful in a later study on implicit leadership theories in corporate mission statements.

Fifth Level: Concluding Analyses and Interpretation

With some exceptions, the basic patterns of qualitative content analysis described in Steps one to four can be applied to many different studies and topics. However, the concluding analyses (both qualitative and quantitative) and their interpretation are particularly dependent on the research or assessment questions. Hence, the following discussion can only provide some general considerations on how to conclude, present, and interpret the generated data. As the major focus of this article is on qualitative analyses, only some general statements about the role of quantification and statistical analyses of the content data will be presented here. Descriptive numerical analyses in the context of content analysis are sometimes devalued as "simple" and uninspired counting of words or themes (Mostyn, 1985). From experience with the example study on implicit leadership theories, even the rather basic measures of *absolute topic frequency* (i.e., total number of times a topic is addressed by the research subjects), *relative topic frequency* (i.e., average percentage of a topic by person), and *person frequency* (i.e., how many of the research subjects address a certain theme) can yield insightful results that complement the qualitative analysis. In particular, quantitative analyses may help the researcher avoid weighing single comments too heavily and generalizing findings too quickly. Although it is tempting to include the most vivid or surprising quotes (Bachiochi & Weiner, 2002), the mentioned frequency analyses can help to critically evaluate how representative these statements are for the whole sample. Finally, beyond the commonly used frequency analyses, there exists a great variety of more complex methods for the analysis of categorical data (e.g., Agresti, 1990) that can prove helpful in answering certain questions the researcher may have (e.g., comparison of subgroups, analysis of latent classes, search for types or configurations). In the example of the leadership study, multidimensional scaling was used to analyze the argumentation structure of the implicit leadership theories (see Schilling, 2001).

The goal of the concluding qualitative content analyses is not to produce counts, but rather to fracture the data, rearrange it to facilitate the comparison of objects within and between categories (Maxwell, 1998), and to draw and verify conclusions. For that purpose, it is necessary to find adequate forms to display the data and the results. A display can be understood as an "organized, compressed assembly of information that permits conclusion drawing and action" (Miles & Huberman, 1994, p. 11). While quantitative results can be displayed parsimoniously in the form of tables, distributions, and statistical values, the display of qualitative results is still an unsolved problem (Flick, 1991). In the past, the most frequent, but cumbersome form for qualitative data has been *extended text* (in the form of transcripts; Miles & Huberman, 1994), while the typical, but unsatisfying solution for the display of results was to justify conclusions by reporting "typical" quotations (*selective justification*; Flick, 1991). Both forms of display are obviously prone to mistakes and bias. Miles and Huberman (1994) give an illustrative and informed overview on possible forms of qualitative displays including matrices, graphs, charts, and networks as well as some ideas creating displays. The exact form of the display is highly dependent on the questions the researcher wants to answer. One example from the study on implicit leadership theories will now be presented to illustrate this point. The *concept map* in Figure 2 is a helpful device to give an overview on the major themes given by one, a group, or – in this case – all the interviewees.

Based on the coded statements of all interviewees, dominant (recurrent) themes were identified within each category by sorting the statements for each category based on their semantic similarity. In this way the exact content of the categories is extracted, graphically displayed, and complements the quantitative results for the categories. The sequence in the presentation of the themes (e.g., "analyzing problems" and "generating solutions" for the category "Problem solving") can be used to express the rank order in the importance of the different themes (e.g., how often a certain theme is addressed in total or by how many subjects). Interpreting this map might imply looking for patterns (e.g., are there certain similarities concerning the views on "good" leadership transcending the boundaries of the categories?), to cluster objects (e.g., which categories have many, which only one or two main themes?), to make contrasts (e.g., how do categories with many main themes differ from those with only few?), and comparisons (e.g., if used for an individual or a group: comparing different views on a topic) and also looking for the unsaid (Mostyn, 1985). Not only the presence, but also the absence of statements can yield interesting results. In our example, it would be possible to compare the description and anchoring examples of the different categories with the actual statements of the interviewees. This example should give some ideas on how displays can be used to reach conclusions and

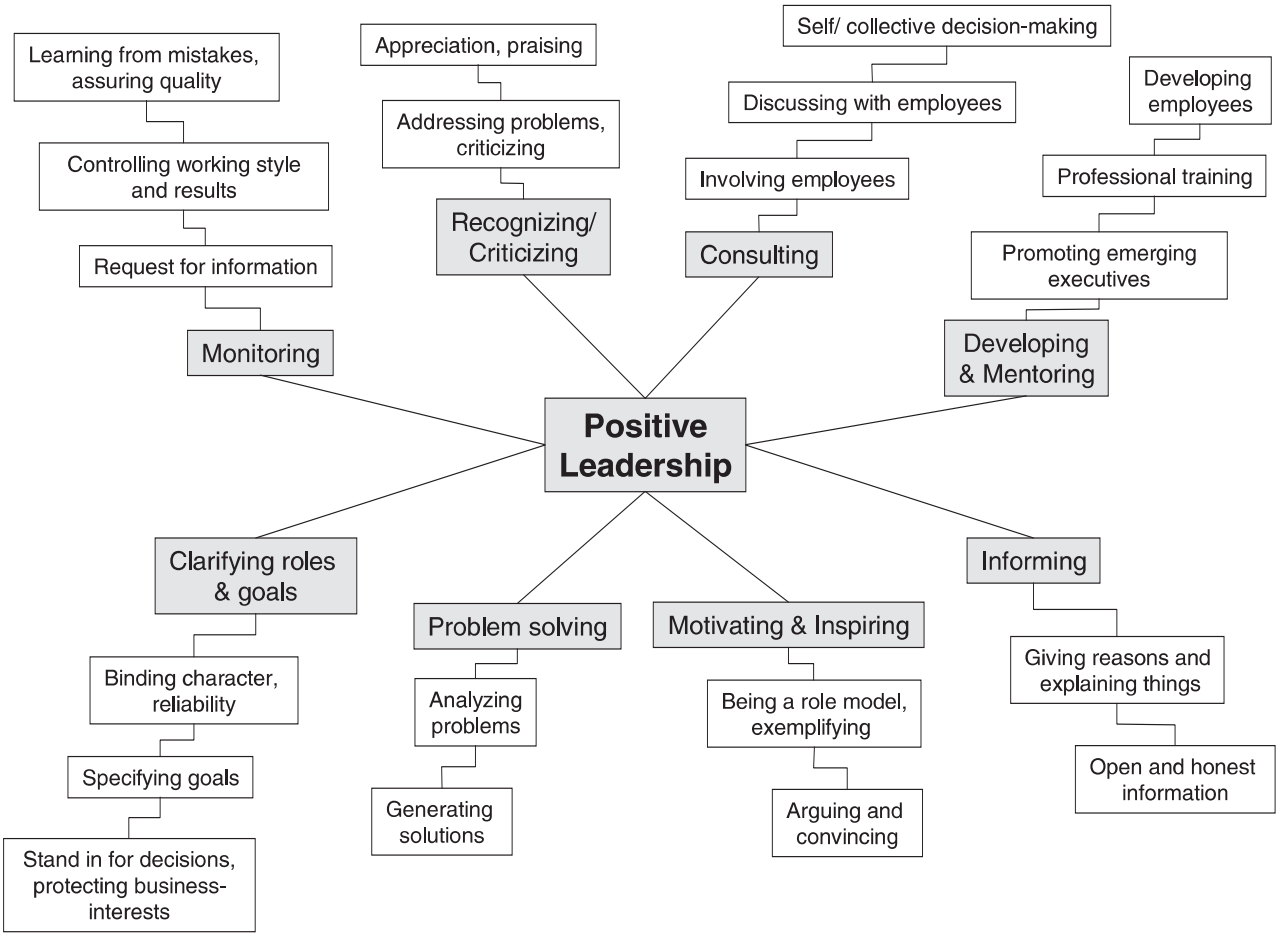

Figure 2. Example of a concept map.

– in the end – answer the research questions. For more details, the overview of Miles and Huberman (1994) is recommended.

Discussion and Conclusions

This article aimed at describing specific steps and problems in the course of qualitative content analysis. Guidelines have been formulated for the assessment process in general (guidelines for the assessment process (GAP); cp. Fernández-Ballesteros et al., 2001), but are still missing for the field of qualitative assessment. This article was intended to help fill this gap, at least for the field of qualitative content analysis. As for GAP, the intention was not to formulate rigid, inflexible rules, but rather to give recommendations for professional behavior (De Bruyn, 2003). Possible applications of content analytical procedures within the field of psychological assessment are manifold. At the individual level, the discussed process might prove helpful to analyze interview data generated in the course of clinical assessment (e.g., the use of articulated thoughts; Davidson, Vogel, & Coffman, 1997, or the mapping of cognitive structures; Cacioppo, von Hippel, & Ernst, 1997), personality assessment (e.g., the analysis of personal constructs; Bannister & Fransella, 1981), or personnel selection (e.g., the employment interview; Jelf, 1999). At the organizational level, the assessment of organizational culture is especially associated with the use of qualitative methods (Kluge, 2003; Schein, 1990). Despite their variety, verbal or textual data is generated in all of these areas, hence, the discussed process could be used as a guideline for systematic and rule-based data analysis.

However, four important points should be mentioned. First, the use of a method should always depend on the questions one wants to answer. Qualitative methods have their special strengths in the discovery and generation of hypotheses, but also to get a more in-depth understanding of the ideas and views of a person. Therefore, qualitative designs are often the first step to analyze a topic and should be complemented and/or followed by quantitative approaches when necessary. For the example of implicit leadership theories, the next steps should include qualitative (e.g., testing the results in another setting) as well as quantitative designs (e.g., developing and testing a standardized

instrument for implicit leadership theories based on the results). Second, many researchers point out that qualitative data analysis does not come after data gathering (e.g., Maxwell, 1998; Silverman, 2000; Tesch, 1990), but should be intertwined activities in the assessment process. In this sense, the assessor will be able to progressively focus his inquiry and test emerging conclusions. For reasons of clear presentation, the above discussion on qualitative content analysis might have given the impression of a straightforward process that starts after data collection has finished. In reality, it might better resemble a complex, circular process in which the researcher develops and changes his proceedings, generates and discards his ideas. As long as the decisions are made explicit and, therefore, intelligible to the reader, this should not be regarded as a flaw but as a sign of securing the quality of the assessment. As Silverman (2000, p. 121) puts it, "in most qualitative research, sticking to your original research design can be a sign of inadequate data analysis rather than demonstrating a welcome consistency." Third, another point not mentioned in the discussion concerns the use of computer programs for qualitative analysis (e.g., Textpack, Atlas.ti, MaxQDA; see Alexa & Zuell, 1999). As Creswell (1998) points out these programs can be especially useful for studies with large and/or diverse databases. Typically, they do not give assistance in the process of qualitative data analysis as they are intended to be open to different kinds of qualitative approaches (e.g., content analysis, grounded theory, phenomenology). In this sense, these programs might be helpful in managing, retrieving, and analyzing the data but they do not release the assessor from making the decisions on how to shape the analytic process. Fourth, the accuracy and transparency of the analysis process is achieved at the expense of time and labor on the part of the researcher (team). In our example, analyzing one interview from level one to four took about 24 h (3 working days) per person. While for research purposes this might be accepted to secure the accuracy of the results, it is a serious problem for practical applications, especially in organizations. Therefore, it seems highly important to find and discuss ways to improve the economy of the qualitative content analysis process. One of the most time consuming aspects of the content analysis is the transcription of the tapes. Instead of transcribing the tapes, they could rather be used to control for the accuracy of and if necessary correct written protocols. With this reduced body of text material, it would also be possible to combine paraphrasing, generalizing, and reducing as one step instead of three. Finally, as mentioned before, the rather difficult step of defining the coding units is only important if one is interested in the total number of statements (absolute topic frequency). This step could be skipped if the researcher is just focusing on the number of interviewees who mentioned a certain aspect at all (person frequency). As these different measures produced rather similar results in the present study it could be a another approach to make the process of qualitative content analysis more much economic.

References

Agresti, A. (1990). *Categorical data analysis.* New York: Wiley.

Alexa, M., & Züll, C. (1999). Commonalities, differences, and limitations of text analysis software: The results of a review. *ZUMA Working Paper, 6*, 1–29.

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

Bannister, D., & Fransella, F. (1981). *Inquiring man: The psychology of personal constructs*. London: Croom Helm.

Cacioppo, J., von Hippel, W., & Ernst, J. (1997). Mapping cognitive structures and processes through verbal content: The thought-listing technique. *Journal of Consulting and Clinical Psychology, 65*, 928–940.

Conger, J.A. (1998). Qualitative research as the cornerstone methodology for understanding leadership. *Leadership Quarterly, 9*, 107–121.

Creswell, J.W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.

Davidson, G., Vogel, R., & Coffman, S. (1997). Think-aloud approaches to cognitive assessment and the articulated thought paradigm. *Journal of Consulting and Clinical Psychology, 65*, 950–958.

De Bruyn, E.E.J. (2003). Assessment process. In R. Fernández-Ballesteros (Ed.), *Encyclopedia of psychological assessment* (pp. 93–97). London: Sage.

Denzin, N.K., & Lincoln, Y.S. (2000). Introduction: The discipline and practice of qualitative research. In N.K. Denzin & Y.S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 1–29). Thousand Oaks, CA: Sage.

Erdener, C.B., & Dunn, C.P. (1990). Content analysis. In A.S. Huff (Ed.), *Mapping strategic thought* (pp. 291–300). Chichester: Wiley.

Feixas, G., Geldschläger, H., & Neimeyer, R.A. (2002). Content analysis of personal constructs. *Journal of Constructivist Psychology, 15*, 1–19.

Fernández-Ballesteros, R., De Bruyn, E.E.J., Godoy, A., Hornke, L.F., Vizcarro, C., Westhoff, K., Westmeyer, H., & Zaccagnini, J.L. (2001). Guidelines for the assessment process (GAP): A proposal for discussion. *European Journal of Psychological Assessment, 17*, 187–200.

Flick, U. (1991). Stationen des qualitativen Forschungsprozesses [Steps of the qualitative research process]. In U. Flick, E. von Kardorff, H. Keupp, L. von Rosenstiel, & S. Wolff (Eds.), *Handbuch Qualitative Sozialforschung – Grundlagen, Konzepte, Methoden und Anwendungen* (pp. 147–179). München, Germany: Psychologie Verlags Union.

Frommer, J., Langenbach, M., & Streeck, U. (2004). Qualitative psychotherapy research in German-speaking countries. *Psychotherapy Research, 14*, 57–75.

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

Jelf, G.S. (1999). A narrative review of post-1989 employment interview research. *Journal of Business and Psychology, 14*, 25–58.

Kluge, A. (2003). Assessment of organizational culture. In R. Fernandez-Ballesteros (Ed.), *Encyclopedia of psychological assessment* (pp. 649–657). London: Sage.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage.

Locke, K. (2002). The grounded theory approach to qualitative research. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations – Advances in measurement and data analysis* (pp. 17–43). San Francisco: Jossey-Bass.

Maxwell, J.A. (1998). Designing a qualitative study. In L. Bickman & D.J. Rog (Eds.), *Handbook of applied social research methods* (pp. 69–100). Thousand Oaks, CA: Sage.

Mayring, P. (1987, August). *New approaches to qualitative research in German psychology: An attempt at integration*. Paper presented at the Third Symposium of the International Association for Qualitative Research in Psychology, Perugia, Italy.

Mayring, P. (1994). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* [Qualitative content analysis: Fundamentals and techniques]. Weinheim, Germany: Deutscher Studien Verlag.

Mayring, P. (2000). Qualitative content analysis. *Forum: Qualitative Social Research, 2*, 1–28. Retrieved March, 31, 2001, from http://www.qualitative-research.net/fqs-texte/2-00/2-00mayring-e.htm.

Miles, M.B., & Huberman, A.M. (1994). *Qualitative data analysis: An expanded sourcebook*. London: Sage.

Mostyn, B. (1985). The content analysis of qualitative research data: A dynamic approach. In M. Brenner, J. Brown, & D. Cauter (Eds.), *The research interview* (pp. 115–145). London: Academic Press.

Neimeyer, G.J., & Gemignani, M. (2003). Qualitative methods. In R. Fernández-Ballesteros (Ed.), *Encyclopedia of psychological assessment* (pp. 785–800). Sage: London

Schein, E.H. (1990). Organizational culture. *American Psychologist, 45*, 109–119.

Schilling, J. (2001). *Wovon sprechen Führungskräfte, wenn sie über Führung sprechen: Eine Analyse subjektiver Führungstheorien* [What do leaders talk about when they are talking about leadership: An analysis of implicit leadership theories]. Hamburg: Dr. Kovac.

Silverman, D. (2000). *Doing qualitative research – A practical handbook*. London: Sage.

Tesch, R. (1990). *Qualitative research: Analysis types and software tools*. New York: Farmer Press.

Jan Schilling

Department for Industrial and Organizational Psychology
Institute for Psychology of the Technical University Aachen
Jägerstr. 17/19
D-52066 Aachen
Tel. +49 241 8096127
Fax +49 241 8092526
E-mail Jan.Schilling@psych.rwth-aachen.de

