

Qualitative research evaluation – how to get from first ideas to a final paper

How can qualitative evaluations be carried out? In which contexts is qualitative evaluation appropriate? Which methodological steps need to be considered? Which qualitative methods can researchers use to collect and evaluate data? These are the main questions this article is meant to provide answers to. It gives an overview of the main steps in the different phases of qualitative evaluations. These include planning, implementation, analysis and communication.

Author

Katharina Felbermayr
OeNB, Financial Literacy and
Culture Division
katharina.felbermayr@oenb.at



Qualitative evaluation is iterative and flexible

Qualitative evaluation is an iterative and flexible process. Different phases of research, such as evaluation design and data analysis, do not have to happen in a strict sequence. They can overlap and inform each other.



Qualitative methods need to be aligned with research questions

There are many different ways to collect and analyze data for qualitative evaluation. You should choose your qualitative method based on scientific criteria, prioritizing this question: Which method is best suited for the purposes of my research project, i.e. for answering my research question?



Qualitative evaluation requires expertise

Qualitative evaluation relies on researchers that are experts in collecting and analyzing qualitative data. They need to have the methodological qualifications required, e.g. for adjusting methods to target groups and assessing the informative value of data.

Opinions expressed by the authors of studies do not necessarily reflect the official viewpoint of the Oesterreichische Nationalbank or the Eurosystem.

Abstract

This report provides an overview of the key steps involved in evaluating qualitative research. It is aimed at readers who have little experience in applying qualitative methods and evaluating qualitative research. Readers will learn how to follow an evaluation cycle: it begins by choosing a suitable qualitative inquiry method and ends by analyzing the collected data. Note that I am not going to discuss qualitative methods and methodological approaches in depth here. My goal is to show readers how qualitative evaluation works and to point out its added value. To dive deeper into the topic, I recommend that readers get hold of one of the many qualitative method handbooks around.

1 Introduction: characteristics of qualitative research evaluation

The evaluation process consists of four phases: *planning, implementation, analysis and communication*. In each phase, researchers must go through a series of methodological steps, making informed decisions. I am not going to explain the phases and the methodological steps at length here. This will be covered in one of the publications of the OeNB Financial Literacy Evaluation Series that discusses the evaluation cycle. But first things first: what is an evaluation? In what ways does an evaluation differ from basic research? Hirschauer (2006, p. 405) explains that evaluation research usually refers to applied social research that is characterized by a triangular relationship: (1) A client (2) commissions evaluation research (3) to assess an area of practice. By examining the quality of programs and measures, i.e. their effectiveness, efficiency, acceptance, etc., the research is meant to inform decisions and make decisions more rationally productive.

Researchers must decide for a qualitative evaluation study in the planning phase. Then, they must also decide which type of evaluation they are going to pursue (impact or process evaluation). *Impact evaluation* refers to “a type of evaluation research focused on assessing the effects, outcomes or impacts of a program, intervention or policy. It aims to determine the extent to which desired changes have occurred and the attribution of these changes to the program or intervention” (Lorenz, 2024, p. 17). In contrast, process evaluation refers to a type of evaluation “focused on understanding the implementation, delivery and mechanisms of a program, intervention or policy. It examines how and why the program works (or not), the fidelity of implementation, and the contextual factors influencing outcomes” (ibid.). Qualitative methods are mainly used in the context of *process evaluations*. This has to do with the questions that are of interest in the context of a process evaluation, “such as those about the public acceptability of the intervention and participants’ experiences” (Yoong et al., 2013, p. 63). Such questions are best answered by using qualitative methods such as interviews or focus groups. After all, qualitative methods “tell the *program’s story* by capturing and communicating the *participants’ stories*. [...] They tell what happened when, to whom, and with what consequences” (Patton, 2015, p. 18). Qualitative methods can thus be used to determine whether and in what way or through the actions of which actors interventions have an effect (Kelle and Erzberger, 2006, pp. 291–299). This already addresses a first characteristic of qualitative research. Other central characteristics of qualitative research relate to topics such as context, knowledge generation, openness/diversity of methods and are presented in more detail below.

Context: Qualitative study results cannot be considered in isolation from the individual context and are usually based on a small sample. Although this allows for in-depth and detailed findings (micro-analytical view), it is not possible to make statements that are valid generally. Statements are only ever possible with regard to specific persons, a specific context, etc., and therefore result in a theory with limited scope.

Qualitative research and the interpretative methods used do not aim at generalizations. Instead, the aim is to capture the diversity of perspectives and particularities of the respective voices (deeper understanding).

Deeper understanding: The use of qualitative methods provides valuable information about the actors involved and their context. The added value of qualitative data in evaluation and monitoring is that they complement quantitative data. Qualitative data provide “a depth of contextual understanding and a level of detail that one can’t get with quantitative data alone” (Yoong et al., 2013, p. 113). Especially in the context of a process evaluation, it is important that qualitative methods help gain greater process-specific and detailed knowledge (Kuckartz et al., 2008, pp. 74–75). This allows for arriving at a deeper understanding of the object under investigation. During the implementation of a new measure, qualitative methods as part of a process evaluation can provide valuable information on how the measure is experienced by the actors or what the reasons are why measures are effective or not (Goodrick and Rogers, 2015).

Exploring the inner perspective: Qualitative research is interested in individuals’ inner perspective and subjective opinions and aims to understand these amid various systems or contexts that influence each other. Qualitative methods are therefore particularly suitable for surveying a person’s experience or perception in a specific context. Especially with regard to vulnerable target groups (e.g. people with disabilities), qualitative research can also reflect the diversity and, above all, complexity of experiences (Coons and Watson, 2013). This makes it possible to give vulnerable target groups a voice.

Flexibility and openness of methods: Given its flexible nature, qualitative research allows for making methodological adjustments in advance, e.g. with regard to the interview guide or the interview situation. Moreover, adaptations are possible during the research process, e.g. by compiling new data to follow up on a new hunch (Charmaz, 2014; Felbermayr, 2023). Qualitative research thus offers the methodological flexibility that is necessary to justifiably adapt the research to individual needs.

Discovery of new phenomena: Qualitative research is considered a discovering science (Flick et al., 2012, p. 25). This means it aims at discovering phenomena or detecting topics that have so far stayed under the radar with a view to generating new knowledge. Due to their inherent flexibility, qualitative methods are particularly well suited for investigating fields of research that have been little researched to date. This leads to the discovery of new knowledge that is often not present in the researchers’ prior knowledge and is therefore not taken into account in the construction of standardized (quantitative) instruments (Kelle and Erzberger, 2006).

Interaction and reflection: Qualitative research is characterized by a variety of data collection and evaluation strategies that are applied based on different methodological principles. Despite this diversity, qualitative researchers are united by “the self-image of understanding research as an interaction between the researcher or researchers and the research subjects” (Mensching, 2006, p. 339). Related to this is the question of data independence. While in quantitative research the researcher’s independence from the object under investigation is key, in qualitative research the methodically controlled subjective perception of the researcher is an essential component of gaining knowledge (Flick et al., 2012, pp. 2–25). The researcher therefore consciously adopts a reflexive stance. Qualitative researchers are always part of the reality they are researching and are not neutral, value-free experts. This makes it all the more important to adopt a reflexive stance to be aware of one’s own role in the research process, such as social (power) position, origin, gender or cultural affiliation, and to constantly reflect on this (Charmaz, 2014). All of these aspects can have an influence on data collection, the quality of the data and therefore on the result. This calls for permanent self-reflection in the research process, which is supported, for example, by writing memos (Felbermayr, 2023).

2 Choosing the appropriate qualitative survey method

Deciding on the appropriate qualitative survey method is an important step in planning a qualitative evaluation. The choice must be made in accordance with the research questions and the objectives of the evaluation. According to Patton (2015, p. 248), “getting clear about purpose” is one of the central steps at the beginning of the evaluation. Another thing to be taken into account is the underlying research design. If, for example, an evaluation is meant to find out whether a financial education measure leads to changes in the behavior of students, an observation seems more appropriate than a qualitative interview. However, there is no magic formula for decisions like these. To choose a suitable method, researchers must be familiar with the various qualitative methods, their advantages and disadvantages, fields of application, etc. Here, we therefore first look at ways of using qualitative methods. Next, you will get an overview of the many different qualitative methods available. The following methods will be highlighted in more detail: interviews, focus groups, observation and desk review of documents.

2.1 Ways of using qualitative methods in research evaluation

The World Bank toolkit outlines three fields of application for the use of qualitative methods in evaluation (Yoong et al., 2013). First, qualitative data are used to enhance quantitative research material, such as survey questions. This concerns both the content of such questionnaires and reformulating questions. “For example, evaluators of a financial capability program may conduct a small number of focus groups (qualitative data) to examine how potential program beneficiaries talk about the issues the program addresses (such as savings, budgeting, and so forth), which can then help construct the surveys (quantitative data) using the most appropriate terminology” (Yoong et al., 2013, p. 113). Second, qualitative data can provide a deeper understanding of the development of financial education programs as part of a formative process evaluation in the implementation phase. To this end, for instance interviews are conducted with the people involved. The qualitative data and information collected are meant to help improve the design and implementation of the program. “In particular, qualitative data can shed light on program implementation and operational issues, including questions about the most appropriate mode of delivery, the identification of the target population, and so forth” (ibid., pp. 113–114). Third, qualitative data can provide a valuable input for summative research. This includes both process and impact evaluation. “Although qualitative data can’t by themselves establish causality between the evaluated financial capability program and observed outcomes, they can help produce a richer and more informative picture of the program being evaluated and provide insights that can’t be fully captured through surveys and other quantitative data” (ibid., p. 114). Qualitative methods can, in particular, close gaps in interpretation that can arise through the exclusive use of quantitative methods.

In the various fields of application, qualitative methods can be used in different ways depending on the research interest. The following list provides some key examples of how qualitative methods can be used.

Using qualitative methods as...

- single method
- in a mixed method design
- in a (qualitative) triangulation design
- in a participatory approach
- in a case study
- in a cross-section or longitudinal-study design

The areas of application differ in terms of whether qualitative methods are used as a single method, in combination with a second qualitative method (triangulation) or in combination with quantitative methods (mixed method). Participatory approach and case study refer to further research approaches that allow

using different qualitative methods. In a participatory approach, the research participants should be given an active role in the research process and play an active part in shaping it. At the center of a case study is a case. A case can be defined in different ways, e.g. an institution, a family or several students of a year group, and is examined by applying different methods to gain a better understanding. In addition, qualitative methods can be used in a cross section (one point in time data collection) or longitudinal study design (longitudinal data collection). The individual fields of application are discussed in more detail in separate publications of the OeNB Financial Literacy Evaluation Series.

2.2 Overview of qualitative methods

Numerous methodological handbooks provide a good and in-depth insight into the variety of qualitative research methods (e.g. Denzin and Lincoln, 2018; Monique, 2020; Patton, 2015). However, no specific research methods are used for qualitative evaluation research. Instead, researchers rely on the general canon of methods of qualitative social research (Mensching, 2006, p. 340). Approaches to qualitative research differ in terms of the “type of depth” (Charmaz, 2011, p. 103) that can be achieved with the data collection/evaluation methods. What varies depending on the qualitative research approach are the theoretical point of reference, the understanding of the subject matter and the methodological focus. As a researcher, it is important to select the appropriate qualitative method from the extensive canon of methods (single method). However, there is no universally valid scheme for categorizing the variety of qualitative methods. The methods are structured differently depending on the focus. Patton (2015, p. 14) structures the qualitative methods with reference to the type of qualitative data. Qualitative results are therefore based on three types of qualitative data: (1) interviews, (2) observations and fieldwork and (3) documents, as summarized in table 1. In my opinion, this classification is particularly well suited to categorizing and describing qualitative methods and the different types of collected data. This report presents four qualitative methods in more detail. Each of these methods can be assigned to one of the three categories or types of qualitative data according to Patton (2015, p. 14): *Interviews* and *focus group* (interviews), *observations* (observations and fieldwork) and *desk review of documents* (documents).

What the various qualitative methods have in common is that qualitative data are collected, which in turn form the basis for the analysis. But what is the difference between qualitative and quantitative data? The OECD¹/INFE² (2010a, p. 6) states the following in its Guide to Evaluating Financial Education Programmes:

Table 1

Types of qualitative data

1. Interviews	Open-ended questions and probes yield in-depth responses about people's experiences, perceptions, opinions, feelings, and knowledge. Data consist of verbatim quotations with sufficient context to be interpretable.
2. Observations and fieldwork	Fieldwork descriptions of activities, behaviors, actions, conversations, interpersonal interactions, organizational or community processes, or any other aspect of observable human experience are documented. Data consist of field notes: rich, detailed descriptions, including the context within which the observations were made.
3. Documents	Written materials and documents from organizational, clinical, or program records; social media postings of all kinds; memoranda and correspondence; official publications and reports; personal diaries, letters, artistic works, photographs, and memorabilia; and written responses to open-ended surveys are collected. Data consist of excerpts from documents captured in a way that records and preserves the context.

Source: Patton (2015, p. 14).

¹ OECD = Organisation for Economic Co-operation and Development.

² INFE = International Network on Financial Education.

“Data such as written or spoken thoughts and conversations, photographs or drawings is very useful for understanding the experiences of people and exploring questions such as why or how something happened. It is called Qualitative data and is beneficial when you want to describe the variety of experiences, rather than the proportions of people experiencing certain things.

Data that provides you with numbers for analysis is useful when you want to answer questions like how many or how much and is called Quantitative data. However, this type of data does not reveal reasons for not achieving or exceeding programme objectives. You will need to use qualitative data to find reasons, to find strengths and weaknesses of the programme” (OECD/INFE, 2010a, p. 6).

As outlined in the introduction, qualitative methods make it possible to gain a deeper understanding and capture diverse perspectives. Why and how questions are predestined to generate this knowledge, e.g. by means of interviews.

2.2.1 Interviews

Interviews are a popular qualitative method in evaluation and research. An interviewer asks questions on predefined topics to obtain information from the interviewee. Interviews usually take place as individual interviews, i.e. in a one-on-one setting (1 interviewer and 1 interviewee). Conducting pair interviews or group interviews is also possible and common. Individual interviews are preferable for sensitive topics, e.g. debts. Pair or group interviews are suitable should a one-on-one setting be perceived as too stressful or too intimate (Michael, 2022).

Interviews can be conducted in various ways. The following variants can be distinguished (see OECD/INFE, 2010b, p. 11; Michael, 2022):

- **Face-to-face interviews:** The interviewer and interviewee(s) are present in the same room at the same time and usually sit opposite each other. The conversation therefore takes place face-to-face. This is the most common way of collecting qualitative data using interviews. Advantage: In addition to verbal information, the interviewer also gains a deeper insight into the nonverbal behavior of the interviewee. Recommended for discussing sensitive, in-depth topics. Disadvantage: Takes more time to plan and implement.
- **Phone or video interviews:** The conversation takes place via telephone or video. The people are therefore not present in the same room. This form of interviewing has become increasingly important, not least due to the coronavirus pandemic. Advantage: Calls can be arranged more flexibly and are easier to conduct over longer distances and across national borders. Disadvantage: Both parties must have the technical prerequisites and know-how. Conversation and dynamics differ from face-to-face interviews.
- **E-mail interviews:** In this form, the interviewee receives written questions by e-mail. The questions are answered and sent back. This form of survey is only suitable for a structured guideline. Advantage: The interviewees can decide when they want to answer questions. Disadvantage: It is not possible to ask ad hoc questions.

A wide range of qualitative methods can be used in evaluation and research. The same is true for interview forms. Choosing the appropriate interview form depends on the research interest and research questions. The interview form then determines the structure of the guidelines, the degree of openness and structuring as well as the role of the interviewer. Two common interview forms are key informant interviews and in-depth interviews.

Key informant interviews are primarily conversations with experts, which is why this form of interview is often referred to as expert interviews. Key informant interviews are used to gather in-depth knowledge

from experts. The focus is on knowledge about a specific topic and not on the interviewee's biography or person as such. Discussions with experts can contribute to a better understanding of complex topics. Such interviews are often conducted as structured interviews, which is often due to experts' limited time for interviews (Flick, 2006b; OECD/INFE, 2010a).

In-depth interviews focus on interviewees' expectations, experiences, opinions and feelings, which are recorded in detail. This requires an interview format that gives the interviewees enough time and space to reflect and share their insights. The interviews are often conducted in an unstructured manner, which allows for discussing individual topics in greater depth. In-depth interviews are therefore particularly suitable for discussing sensitive topics that people would not address in a group setting (focus group) (Yoong et al., 2013, pp. 114–115). The interviewer's skills and experience are also of central importance when conducting in-depth interviews:

"In qualitative research, the nature of the interaction with the respondent is critically important. Because in-depth interviews call for a high level of skill it is important that those facilitating them have substantial interviewing experience, either in a research setting or some other context involving nondirective interviews (i.e., interviews that are allowed to follow the course the interviewer may set). It is very important that the interviewers do not influence what the respondent says and, above all, that they allow and encourage the respondent to speak at length on the topics to be covered. Interviewers should have well-developed listening skills and be familiar with techniques to probe replies and encourage the respondent to elaborate, such as using neutral prompts like "Why do you say that?" and "Can you tell me more about that?" A good in-depth interviewer will allow respondents to stray from the order of the topics in the topic guide if that is how the respondent wants to tell the story" (ibid., p. 162).

Depending on how the interview guide is structured, a distinction can also be made between (1) structured interviews, (2) unstructured interviews and (3) semi-structured interviews (Michael, 2022).

- **Structured interviews:** In a structured interview, the interviewer follows an interview guide with predefined topics and interview questions. Throughout the interview, the interviewer keeps to the sequence of topics and questions as well as the predetermined structure. Therefore, the conversation leaves little room for flexibility. However, structured interviews are more easily comparable in the analysis. Structured interviews are ideal for obtaining relevant data on a specific topic or for answering a specific research question. This form of structured inquiry is therefore particularly suitable for impact evaluations. (Michael, 2022; Yoong et al., 2013, pp. 121–122).
- **Unstructured interviews:** Unstructured interviews lack a list of interview questions that are asked in a specific order. Here, the interviewer only relies on topics that are of interest and that are used as a conversation starter. Based on the interviewee's statements, the interviewer asks follow-up questions and questions of interest. This form of interviewing corresponds most closely to having a natural dialogue and allows for conducting interviews in a highly flexible way. Unstructured interviews are suitable during exploratory research aimed at collecting new knowledge. However, evaluating and comparing unstructured qualitative data is usually much more complicated than with structured interviews (Michael 2022; Yoong et al., 2013, pp. 121–122).
- **Semi-structured interviews:** Semi-structured interviews are often conducted with a view to evaluating programs. Such interviews comprise some predefined topics and interview questions. In contrast to structured interviews, however, the interviewer does not have to rigidly follow the sequence of questions. This makes it possible to gather new, previously unknown knowledge. In addition, interviewees can address new topics in the interview, which enables deeper insights into their perspectives (Yoong et al., 2013, p. 115). "This allows interviewers to plan questions that are specifically

geared towards their research questions and data that they need while also allowing them to be flexible and reflexive” (Michael, 2022). According to Adams (2015, p. 493), conducting structured interviews requires a special “interviewer sophistication”: “Interviewers need to be smart, sensitive, poised, and nimble, as well as knowledgeable about the relevant substantive issues.”

A decision must also be made about how to record interviews. There is no right or wrong here, but there are advantages and disadvantages that need to be weighed up (see Adams, 2015; Helfferich, 2011).

Audio: In most cases, interviews are recorded with a recording device with the consent of the interviewees. A small recorder with external microphones is recommended for working in the research field, as these provide better sound quality and are often better at filtering out ambient noise. This is particularly necessary when conversations cannot be held in a quiet, enclosed room. Sufficient storage capacity must also be ensured. Recording with a cell phone and storing audio files in a cloud should be viewed critically from a research and data protection perspective. The advantage of tape recording is that it can be listened to repeatedly and the interviewer can concentrate better on the conversation and asking the next questions. At the same time, being recorded can make interviewees feel self-conscious or ill at ease and thus influence the interview. It is advisable to place the recording device in the middle of the table before starting the recording and only start recording the conversation after a warm-up phase.

Video: Recording with video is another option that is particularly popular with focus groups. In addition to the spoken word, people’s behavior is recorded on video, which allows for analyzing nonverbal behavior or group dynamics. Due to their size, video cameras are more likely to be noticed by interviewees than recorders and are not forgotten as quickly. The presence of a camera can therefore influence interviewees’ behavior, which must be reflected in the evaluation.

Paper and pencil: Instead of technical recording options, written notes can also be taken during or after the interview. Taking notes during the interview might bother the interviewees and give the impression that the interviewer is not really listening. At the same time, this presents interviewers with the challenge of having to take notes awhile also keeping the conversation going by asking questions. Taking notes only after the interview requires a high level of concentration and a good memory on the part of the interviewer in order not to forget important points. In both cases, the note-taking is subject to the interviewer’s selective perception. What is noted down is what the interviewer remembers and feels was important. “But notes could not sufficiently preserve the participants’ tone and tempo, silences and statements, and the form and flow of questions and responses” (Charmaz, 2014, p. 91). But when is it advisable to use the paper and pencil technique? Note-taking is useful if the participants refuse audio or video recordings or for particularly sensitive topics.

2.2.2 Focus groups

A focus group is “a moderated discourse procedure in which a small group is encouraged to discuss a specific topic by means of information input or (focused) questions” (Krueger and Casey, 2015; Schulz, 2012). Similar to interviews, there are also different forms of focus groups. According to Morgan (1998, p. 29), the various forms of focus groups have three things in common: “They are a research method for collecting qualitative data, they are focused efforts at data gathering, and they generate data through group discussions.” Focus groups are intended to reflect the diversity of (divergent and controversial) perspectives on a topic. Achieving group consensus is not the goal of focus groups, nor is making decisions (focus groups are not decision-making forums) or promoting disputes among participants. “Focus groups are conducted to gather the range of opinions and experiences” (Krueger and Casey, 2015, p. 509).

A special feature of the focus groups is the dual communication: Communication takes place between the participants, but also between the participants and the moderator. Focus groups can be used to gather both individual opinions and group opinions on a topic or issue. Focus groups make use of the dynamics

of collectives (groups) when collecting data, which proves to be particularly beneficial for collecting attitudes, opinions or taboo topics (Flick, 2009). Just hearing other points of view can be thought-provoking. One advantage of focus groups is therefore that the participants stimulate each other, which leads to open answers and minimizes, to a certain extent, mechanisms such as saying yes or social desirability (Cropley, 2002, p. 110–11). In addition to the verbal statements, one may analyze the interaction between the participants, e.g. by using the method of observation. This also highlights the difference to group interviews, where a question from the moderator is answered by each participant, but there is no discussion among the participants (Yoong et al., 2013, p. 118).

- **Knowledge generation:** Focus groups are particularly well suited to generating (new) knowledge from a group (multiple perspectives). In other words, the focus is on generating knowledge and not on testing hypotheses (Bürki, 2000, p. 101 cited in Schulz, 2012, p. 12).
- **Efficiency:** The opinions of several people can be collected within a short time (a lot of information in a short time, with fewer resources).
- **Naturalness:** Group discussions correspond to natural communication behavior. The atmosphere can contribute to people expressing themselves more spontaneously and freely.
- **Validation:** Opinions or views expressed by individuals are “validated” by the other group members and may have to be justified or revised.
- **Stimulus:** Listening to other opinions from the group can contribute to the formation of new ideas/opinions in the individual (Lamnek, 2005; OECD/INFE, 2010b).

When are interviews preferable to focus groups?

- When sensitive topics are addressed that are difficult to discuss in front of a group, e.g. experiences of violence.
- When individuals are expected to report on their experiences in more detail. In other words, when it is about gaining an in-depth understanding of individuals and less about gauging the diversity of perspectives on a topic or the group opinion. While in an individual interview, the interviewer attempts to evoke all possible aspects, arguments and value judgments of the interviewee on a given topic, in a focus group the moderator presents a few stimuli and these are then discussed in turn by the group participants (Zwick and Schröter, 2012, p. 27).
- When people are asked to talk about a topic at greater length, e.g. their own biographical journey through life in a biographical interview.

What needs to be considered when implementing focus group?

Focus groups take place under controlled and planned conditions; they would not occur in real life in this way (Lamnek, 2005). When conducting focus groups, we distinguish four phases. In the (1) *opening phase*, the focus group begins. The moderator refers to important formal aspects, such as anonymity, voluntariness, audio and/or video recording (Bohnsack and Schäffer, 2001). In the (2) *introduction round*, the participants introduce themselves. People decide for themselves how much information they want to disclose about themselves, e.g. first name only or also surname, profession, etc. The discussion begins with the (3) *stimulus*. The stimulus can be varied (e.g. movie, picture or provocative statement) and should stimulate the conversation or discussion about a topic. This is followed by the (4) *guided discussion*. The moderator uses the guidelines for the discussion as a structure or orientation framework (Lamnek, 2005).

- **Number of participants:** The literature diverges on the recommended sample size for focus groups, which ranges from 4 to 10 people (Yoong et al., 2013, p. 118). The World Bank (2013, p. 145) names 8 people as the ideal number to ensure “a full discussion.” What needs to be taken into account is the size of the basic sample. In the case of vulnerable target groups or a very specific topic, four

people may be the maximum group size that can possibly be reached. The composition of the sample must also be considered. Too much diversity among the participants (heterogeneous group) can lead to individual participants dominating the discussion. Too little diversity (homogeneous group), on the other hand, can hinder a discussion – in the sense that topics are not debated due to similar opinions (Yoong et al., 2013, p. 145).

- **Seating arrangement:** Ideally, the seating arrangement corresponds to an “egalitarian structure,” so that all participants experience equal treatment in their subjective perception (Lamnek, 2005, p. 120). A large, round table is best suited for this; with no one “chairing” or sitting at the head of square table formats. The seating arrangement must also take into account participants’ individual needs. Sign language interpreters must sit opposite deaf people or in the deaf person’s field of vision. Visually impaired people should not sit opposite windows due to the incidence of light, for example, as it is difficult for them to see people sitting in front of windows.
- **Moderator:** The environment in which the focus group is conducted should be comfortable and not frightening. The role of the moderator should not be underestimated. The moderator should be perceived as a friendly person who is open to all perspectives (Krueger and Casey, 2015). “Moderating focus groups requires considerable skill, because a number of issues may arise. It calls for all the skills of a good in-depth interviewer, plus the ability to manage the group dynamic and ensure that everyone contributes more or less equally” (Yoong et al., 2013, p. 150). The challenge as a moderator is to lead the group and, for example, to stop frequent speakers and encourage those who are silent to talk. This is because both types of participants can under certain circumstances make a meaningful discussion impossible (Lamnek, 2005, p. 161). Krueger and Casey (2015, p. 511) put it in a nutshell: “Skillful moderators make facilitation look easy. They are friendly, open, and engage with participants before the group starts, making people feel welcome and comfortable. [...] A focus group is working well when participants begin to build on each other’s comments rather than continually responding directly to the moderator”.

It should be noted at this point that focus groups – as well as interviews and observations – can also be used to generate some quantitative results data. “For instance, focus group leaders can count the number of people who agreed or disagreed with a particular statement, or they can conduct a “ranking” exercise whereby participants rank program elements in a certain way” (Yoong et al., 2013, p. 151).

2.2.3 Observation

The interview and focus group are also explained as scientific methods in the World Bank toolkit (Yoong et al., 2013). One method that is completely lacking in the World Bank toolkit is scientific observation. In social research, observation means directly observing human actions, linguistic utterances, nonverbal reactions (facial expressions, gestures, body language) and other social characteristics (e.g. clothing, symbols, customs, forms of living) (Dieckmann, 2017, p. 548). Observation therefore aims to standardize, document and make observations intersubjectively comparable (Bortz and Döring, 2009, p. 262). As a researcher, the aim is to describe descriptively what you observe and write it down accordingly, e.g. a person raises and lowers their head. The observer’s interpretation follows only in a second step and must take into account the (cultural) context. In German-speaking cultures, for example, raising and lowering the head can be interpreted as nodding. However, it should be borne in mind that other explanations are also possible. A person could try to release tension through movement. Scientific observations therefore require qualified researchers who are trained in applying the method of observation and can observe descriptively. According to Patton (2015, p. 331), the path to a skilled observer comprises the following six points:

- (1) “Learning to pay attention: Seeing what there is to see, and hearing what there is to hear
- (2) Writing descriptively
- (3) Acquiring expertise and discipline in recording field notes
- (4) Knowing how to separate detail from trivia in order to achieve the former without being overwhelmed by the latter
- (5) Using systematic methods to validate and triangulate observations
- (6) Reporting the strengths and limitations of one’s own perspective, which requires both self-knowledge and self-disclosure”

Appropriate training is essential for implementing both qualitative and quantitative methods. In the case of qualitative methods, such as interviewing or observation, it is often assumed that one already “naturally” masters the activity, which is why training is often neglected. “Training to become a skilled observer is a no less rigorous process than the training necessary to become a skilled survey researcher or statistician. People don’t “naturally” know how to do systematic research observations. All forms of scientific inquiry require training and practice” (Patton, 2015, p. 330). An important point in the training is to learn how to concentrate during the observation. It is important not to lose the focus of the observation and not to be distracted by other observation stimuli.

An observation is carried out purposefully and with a clear focus that results from the research interest. At the same time, an observation is always selective, i.e. it is never possible to observe and record the entire abundance. Neither is this possible with a camera, as it only ever has a certain angle of view or section of the action in the picture (Bortz and Döring, 2009). Naturalistic observations take place in the field. The field can be understood differently depending on the research approach. For an organizational researcher, the field is an organization, for ethnographers it is a cultural setting, and for evaluators, the field is the program being evaluated (Patton, 2015). As a researcher, it is important to adopt a neutral position when observing. Especially during longer ethnographic field observations, the tension between closeness and distance to the research participants can be a challenge for researchers.

The following table is intended to provide a better overview of the characteristics of scientific observation. The characteristics are presented using various pairs of opposites (see: Boer and Reh, 2012; Bortz and Döring, 2009; Lüders, 2012; Pauli, 2012) in table 2.

Table 2

Characteristics of an observation

<p>Event sampling The observed events are not structured in time, i.e. the observation is decoupled from temporal information. The focus is on frequencies, i.e. whether or how often the observed events occur (Bortz and Döring, 2009, p. 270).</p>	<p>Time sampling The observed event is divided into fixed time periods. The division of the time interval depends on the subject of examination (e.g. interval of 5 seconds) and requires a high level of concentration. Some 30 minutes at one go are recommended, to be followed by a break (Bortz and Döring, 2009).</p>
<p>Qualitative The focus is on an interpretative approach to observation</p>	<p>Quantitative Quantitative data are produced through observation.</p>
<p>Open The participants know that they are being observed.</p>	<p>Covert The participants do not know that they are being observed (to be reflected on in terms of research ethics!).</p>
<p>Participatory The researcher is part of what is being observed. The role of observer is obvious to everyone.</p>	<p>Non-participatory The researcher is not part of the discussion, conducting the observation as an outsider.</p>
<p>Structured (standardized) The observation is carried out using an observation grid (the degree of structuring may vary).</p>	<p>Unstructured The observation is carried out without using an observation grid. Anything that stands out with regard to the research interest is generally noted down.</p>
<p>With technology The observation is recorded with audio and/or video. Technology, especially audio, can also influence the observation.</p>	<p>Without technology The observation is carried out with paper and pencil, i.e. without any other technical aids.</p>

Source: Boer and Reh (2012); Bortz and Döring (2009); Lüders (2012); Pauli (2012).

We distinguish four types of observation depending on whether the observer participates in the discussion and whether the participants know that they are being observed. In other words, whether the observation is carried out openly or covertly (secretly). Any covert observation must be critically considered from a research ethics perspective and must be carried out in a well-thought-out manner (see the OeNB Financial Literacy Evaluation Series publication on data privacy and research ethics). In agreement with Patton (2015, p. 342), I advocate full disclosure to the participants. This means informing people in advance about the planned observation and obtaining their permission. “Trying to run a ruse or scam is too risky and adds to evaluators stress while holding the possibility of undermining the evaluation if (and usually when) the ruse becomes known” (Patton 2015, p. 342). To better understand the different variants, see the following examples from Bortz and Döring (2009, p. 267) in table 3.

Table 3

Variants of observation

a) Participatory-open: A company psychologist participates openly in employee appraisals to explore group problems.	b) Participatory-covert An official of the authority for the protection of the constitution covertly participates in a demonstration while observing demonstrators' behavior.
c) Non-participatory-open: A soccer coach observes the players' performance on the sidelines.	d) Non-participatory-covert: A developmental psychologist observes an argument between two children behind a one-way mirror.

Source: Bortz and Döring (2009, p. 267).

Example

We may confirm the added value of the qualitative method of observation based on an evaluation study we conducted. The aim of the study was for teachers to evaluate a newly developed educational game for students aged 10 to 14 years as part of a training course. An observation is a good way of gaining a deeper insight into teachers' perspectives when testing the didactic method (educational game). The data collected were anonymous, written observation notes using a structured observation grid. The focus of the observation was on verbal statements and not on nonverbal aspects of communication, e.g. facial expressions or gestures. The participatory open observation took place on two consecutive days. Some 40 teachers were present in each observation setting, who, under the guidance of a moderator, tried out the didactic method in small groups of 6 to 8 teachers. Qualitative observations are used to collect data directly in the situation (in situ) and not retrospectively. This makes it possible to make statements about the teachers' experiences directly when trying out the method, which might not have been discussed in a subsequent interview. The qualitative observation provided valuable information on how the game needs to be changed from the teachers' point of view, e.g. clear and plain language, less text to read.

2.2.4 Desk review of documents and materials

A systematic document review is a central method that is often used in the evaluation of programs. This method is described in more detail in the World Bank toolkit (see Yoong et al., 2013). The aim is to review documents, records (including archive records) and data in general that are related to and relevant to the program to be evaluated. The World Bank toolkit distinguishes between the following categories of potentially informative information:

- “Official documents and materials describing the program's aims, structure, and so forth (including, perhaps, a program's website)
- Program materials not intended for public circulation (such as meeting minutes, internal progress reports, internal communications about the program, etc.)

- Data gathered in the course of implementing a program (for instance, demographic information about the program beneficiaries, results of specific activities, logs of program activities, etc.)
- Photographs and audio and video recordings
- Nonprogram data (such as financial transactions, school enrollment records, and so forth)” (ibid., p. 122)

Added value of a desk review

- **Source of information:** Documents, records and data of all kinds are a valuable source of information to learn more about the goals, the design, the planned implementation, the process or the stakeholders (donors, employees, customers, etc.).
- **Changes:** In many cases, changes that have occurred during a program can be mapped in the documents. These changes can be significant for the evaluation of programs.
- **Program quality:** Working with documents can make a direct contribution to the evaluation of program quality. “For example, the quality of program materials (inputs) can be reviewed and assessed by skilled peer reviewers, as can an audio recording of the program being delivered (outputs)” (Yoong et al., 2013, p. 122).

2.2.5 Summary

Table 4

Selective qualitative methods for carrying out an evaluation

Qualitative research method	Description	Benefits	Limitations
Key informant interview	One-on-one setting: an expert is interviewed on a specific topic	In-depth expert knowledge is gathered	Access to the field
	Mostly structured interview	Can help better understand complex topics	Limited informative value Can be costly if travel is involved Interviewer bias: requires reflection
In-depth interview	One-on-one setting: one person is asked more in-depth questions on topics	Insight into individual opinions, feelings, expectations	Time-consuming procedure
	Mostly unstructured interview	Person can communicate their own opinion in more detail Suitable for discussing sensitive topics	
Focus groups	A moderated group discussion	Information from several people in a short time	Not suitable for discussing sensitive topics
	A moderator addresses questions to a small group of people who react to the answers	Group dynamics: People’s behavior becomes visible/observable	Group dynamics: individuals dominating the discussion Individual opinion can be influenced by others Moderation bias: Requires reflection
Observation	Persons’ behavior and actions are described by observation	Valuable information to better understand people’s behavior	Observer bias: separate direct observation from interpretation of what is seen
Desk review of documents	Systematic review of documents, records and data in general	Usually more direct and faster access to the field	Gives no insight into the individual perspectives

Source: Author’s compilation based on OECD/INFE (2010a); Yoong et al. (2013).

3 Designing the instrument

Let us assume researchers have both a clear research interest and clear objectives and questions in mind for an evaluation project. They have already selected research design and method in accordance with the objectives. Now, the next step is developing survey instruments.

No single qualitative survey instrument fits all research contexts. Guides for interviews, focus groups and observations must be adapted to the requirements of the research project at hand. Qualitative research offers the methodological flexibility necessary to conduct research in ways that cater to the individual needs of specific target groups. We can make methodological adjustments to the interview guide both *before* the surveying stage and *during* the research process (Charmaz, 2014; Felbermayr, 2023). The time it takes to develop guides should not be underestimated (Flick, 2006b). Therefore, it is important to allow sufficient time for creating, validating and revising guides for evaluation projects. The following explanations can only provide an initial insight into the development of guides. For further details, see the handbook by Patton (2005).

3.1 Developing interview guides and questions

In evaluation research, researchers often work in teams. In qualitative research, this can mean that several people conduct interviews based on an interview guide they developed together. In these cases, it is important to make sure in advance that all researchers understand the interview questions in the same way. They should thus clarify: What exactly does each question mean? What is the aim of asking each question? How should the guide be applied? Otherwise, researchers might interpret individual questions differently and emphasize different things during their interviews. Several factors must be taken into account when developing interview guides: the specific interview method and, depending on this, how structured the interview should be (structured, unstructured, semi-structured) as well as the type of questions. For the design of a structured interview guide, a list of topics is usually drawn up. Questions on the respective topics are written down and prioritized subsequently. In other words, keeping in mind what would be considered an appropriate interview duration, researchers must critically reflect on which questions from the list they should definitely cover (Patton, 2015, p. 256). In a next step, they need to determine a sensible order in which to work through the different topics and questions (Yoong et al., 2013, p. 115). Inefficiencies in the order should be avoided as they have a negative impact on the interview itself and, thereby, on the quality of the survey data (ibid., p. 150). If there are interview questions regarding some sort of evolution over time, for example, it is advisable not to jump back and forth between different points in time. People might get confused if they first have to answer questions about how they are currently doing in implementing a specific measure to promote financial education (present), then about their wishes for the future (future) and finally about prior implementation experiences (past). Instead, it would make sense to cover the different subjects following a chronological order (past, present, future). Asking questions in an effective and logical order can also help interviewers establish rapport with their interviewees by making them feel more at ease during the interview (ibid., p. 150). In structured interviews, the interviewer will stick to a predefined order of questions. Guides for unstructured interviews, on the other hand, can be seen as a point of orientation that supports the interviewer's memory. Interviewers can also deviate from the guide and spontaneously react to new topics raised by interviewees.

An important part of creating interview guides is coming up with questions that are appropriate for the target group. In general, a distinction is made between open-ended and closed questions. How much interviewees reveal varies depending on the type of question.

- *Open-ended questions* are asked to encourage people to narrate, reflect on or make an elaborate statement about an issue. These include questions starting with “why,” “what,” “how,” “when,” “to what extent,” etc. The question “Why are you taking part in the training?” is, for example, meant to make the interviewees give the reasons for their participation. Different people may answer it differently. In this sense, the answer is open, which also explains the term open-ended question. In-depth interviews in particular are characterized by open-ended questions.
- *Closed questions* (yes/no questions), on the other hand, have a limiting character that allows researchers to gather knowledge (quickly) (Michael, 2022). The question “Do you take part in the training?” can only be answered with “Yes” or “No” and results in more superficial knowledge about the situation.

Table 5 uses various topics to illustrate very clearly how both open-ended and closed questions are asked on the same topic, thereby pursuing different research interests and collecting different information.

Table 5

Open-ended and closed questions

Focus of the inquiry	Open-ended inquiry question	Closed question inquiry framing (to be avoided)
Immigration experiences	What are the processes that immigrants experience during immigration? What are the implications of these processes for how they engage where they have immigrated?	Do immigrants’ experiences during immigration affect how they engage in the community after immigration?
Program evaluation	What works for whom in what ways with what results and in what contexts?	Does the program work?
Homeless youth	What are the experiences of homeless youth? How do they perceive and talk about their experience of homelessness?	Are there patterns in the experiences of homeless youth?
Ecology and climate change	How, if at all, is the ecological system of the Great Lakes changing? What factors are contributing to those changes? What are the implications of those changes for the future health of the ecosystem?	Is the social ecological system of the Great Lakes changing? Is climate change causing the ecological system to change? Can the implications for the future be identified?

Source: Patton (2015, p. 253).

What needs to be considered when creating an interview guide?

- **Adequacy for the target group:** The wording of the questions must be adapted to the (linguistic) skills/needs of the respective target group. Possible target groups are experts, parents, young people, children – with/without disabilities, migrant background, etc. The various needs of the target group must be considered when developing a guide as well as later, when conducting the interviews. For example, the wording in an interview should generally be easily comprehensible, not only when interviewing people with an intellectual disability. Researchers should also consider the extent to which specialist or technical vocabulary is necessary (Buchner, 2008). According to Charmaz (2014, p. 96), it is advisable to speak the language of the interviewees: “Following threads in our participants’ everyday language and discourse helps us to form questions from their terms and learn about their lives.”
- **Balance regarding complexity:** It is important to challenge, yet not overwhelm people. Respondents should not feel underestimated by banal wording. At the same time, abstract or complex wording can also be overwhelming (Flick, 2006b). For example, using too many technical terms and specialist vocabulary creates too much of a challenge for children as interviewees.
- **Balance regarding the number of questions:** Participants’ time is precious. Therefore, researchers should only ask as many questions as necessary in order to find out what they need to know. They should avoid questioning people further out of personal interest.

- **One thought – one question:** The guiding principle is that every question should express one thought. Do not ask several questions about different aspects at once, e.g. “How did you use the cash provided to you through the program, and did you get any other type of assistance?” (Yoong et al., 2013, p. 150). Respondents will not know which question they should answer first then – if they even remember all the questions, which is often not the case.
- **Appreciative and neutral wording:** Researchers should avoid wording that can be interpreted as expressing a negative opinion or criticism. An example would be “Why did you do such a terrible thing?” It is also important not to influence respondents with specific wording and ask neutral questions instead. For example, the question “Was the instruction provided by the program effective and suitable?” suggests that the program was effective and suitable. In contrast, open phrases (e.g. “What was your impression of the instruction provided by the program?”) leave more room to respondents for both positive and negative reactions (Yoong et al. 2013, pp. 149–150).

Example: Interview guide

There is no single set of universally applicable standards for the design of interview guides. However, it has become common practice to group questions by topic. Some researchers make lists of topics and corresponding questions below each topic. The following is an example of a slightly different approach: The topics are listed in the left column, with the corresponding questions to the right (see table 6). There are both open-ended questions and combinations of closed questions with open-ended follow-up questions. Open-ended questions are intended to encourage respondents to talk about a given topic, thereby generating a story (Helfferich, 2011, pp. 102–103). Therefore, they are also referred to as story-generating questions. In addition, there are also so-called elaboration probes. The example does not contain any of these. Their aim is not to generate new stories, but rather to keep a story going. Elaboration probes thus do not have any content per se, in the sense that they provide no or as few presuppositions or content-related impulses as possible (ibid., p. 104). An example would be the question “And what happened next?”

Table 6

Example of an interview guide

Topic	Questions
Experience	What is your experience with the new didactic method? (open-ended question) Could you please describe your own role as part of the project? (open-ended question)
Application of the method	Have you exchanged ideas with other colleagues about the application of the method in lessons? If yes/no, why? (closed question followed by open-ended question)
Challenges	Where do you see the biggest challenges? (open-ended question) Have you tried to overcome the challenges? If so, how? (closed question followed by open-ended question) If you could change one thing about the method, what would it be? (open-ended question)

Source: Author’s compilation.

3.2 Developing interview guides and questions for focus groups

When developing guides for focus groups, researchers can follow the four phases that characterize a focus group according to Lamnek (2005): Opening phase, round of introductions, setting a stimulus and guided discussion to conclude. Focus groups are meant to discuss a given topic. The structure and questions in the guide facilitate the discussion process. After the opening phase and introductions, researchers use a stimulus (e.g. picture, video clip, provocative statement) to kick off the discussion about the topic at hand. Ideally, the stimulus introduces the topic and encourages a lively debate. The intro is then followed by the actual discussion based on the guide for focus groups. Similarly to creating an interview guide, it is a good idea to write down a list of topics with questions for focus groups. Here, researchers may choose

between open-ended and closed questions. When researchers develop guides for focus groups, they must consider the same aspects that apply to interviews: balancing complexity and the number of questions, target group adequacy, etc.

Example: Guide for a focus group

The following guide was developed for a focus group with students and teachers of geography and economics. The guide is based on the four phases according to Lamnek (left column in table 7), to which topics and corresponding questions have been assigned. In my experience, this provides a good framework for researchers working with focus groups.

Table 7

Example of a focus group guide

Phase	Topic and questions
Opening phase	Welcoming the participants, discussing the consent form, etc.
Round of introductions	Could you briefly introduce yourselves, please, by telling us your name, the school you teach at and how much professional experience you have got?
Stimulus	(Different terms are printed out and put on a table, such as financial education, economics education, economic education, financial literacy, etc.). Understanding financial education Please take a look at these terms: To what extent do they mean the same thing to you? Which term is missing? Which term do you use in class and why? What do you think of the following statement: Terms are subject to trends, change over time, but basically always mean the same thing?
Guided discussion	Providing financial education Which topics do you think are particularly difficult to teach in economics/financial education classes? What makes teaching them so difficult? How do you feel about talking to students about money/pocket money? And which topics are particularly easy to teach in class?

Source: Author's compilation.

3.3 Developing an observation grid

Before developing such a grid, researchers need to decide how to go about observing. Here, it is particularly important to distinguish between structured and unstructured observations. Let us assume the goal is to observe how new teaching material on finance/economics is used in a classroom. Observing how often students raise their hands or teachers call on girls would count as *structured* observations, for example. An *unstructured* observation would be a descriptive record of how girls react to a teacher's response.

For both types of observations, it is important to define categories in advance, i.e. what is to be observed (topic). If observations are carried out by several researchers, it is important to ensure they have a common understanding of the topic. It is advisable to explain the individual subtopics in more detail in a separate document that also includes examples. In the observation grid, researchers usually leave enough space next to the topics to write down observations. Ideally, it also contains an additional column for notes. This should help observers separate the descriptive part from their interpretation of what they see. Writing down that a person raises and lowers their head would be an example of a descriptive observation. Understanding the head movement as consent would be an interpretation. It is possible, for example, that the person has a sore neck and therefore raises and lowers their head. Interpretations must therefore always be contextualized and recognizable as such.

Example: Observation grid

During a training course, two participant researchers observed around 80 teachers while they were trying out a new didactic game. The researchers carried out their observations on two consecutive days. Some 40 teachers were present in each observation setting, who, under the guidance of a moderator, tried out the didactic method in small groups of 6 to 8 teachers. Using a structured observation grid, the researchers collected anonymous data in written observation notes. The observations focused on various aspects of the teachers trying out the new game. Table 8 shows excerpts of an observation grid.

Table 8

Example of an observation guide

Topic	What is being observed?	Notes (interpretation)
Reading the game manual		
Cooperation among participants		

Source: Author's compilation.

3.4 Pretesting

Guides should be tested before using them to collect data in the field. This applies to interviews, focus groups and observations. In pretesting, also known as pilot testing, one or several individuals test the guide. Ideally, they belong to the target group the guide was developed for. The added value of pretesting a guide is often underestimated, and this phase tends to be skipped due to time constraints. However, it is during pretesting that potential weaknesses in a guide are revealed.

How does pretesting work? Let us assume that researchers want to ask teachers about their teaching experience. During pretesting, a researcher will print out and test the guide with a teacher, who will not be interviewed afterward. All the questions in the guide are read out to this person word by word. The researcher encourages the teacher to point out any ambiguities, anomalies, etc. and writes down all the feedback. For example, the interviewed teacher might criticize the wording of certain questions (comprehensibility), the length of the guide or the order in which the questions are presented. Feedback helps improve the guide and ultimately also the quality of future survey data. Furthermore, problems that might otherwise arise in the field can be avoided (Yoong et al., 2013, p. 147). So, before gathering data, at least one person should test the guide.

4 Sampling and access to the field

Sampling refers to “the process of selecting units, such as individuals or organizations, from a larger population” (Yoong et al., 2013, p. 139). This is important as not everyone can take part in an evaluation due to financial, structural or time constraints. Sampling is an essential step in data collection, and researchers need to be very careful when choosing the sample. Getting the sample wrong may make it difficult to answer the research questions.

Qualitative and quantitative research approaches use different sampling strategies. These are explained in more detail below. According to Patton (2015, p. 264), researchers apply different techniques following a logic that belongs to the category of either qualitative or quantitative methods. For *qualitative research*, they select participants through *purposeful sampling*. This means that participants are selected according to predefined criteria. Qualitative research primarily aims at obtaining more in-depth knowledge. Therefore, the sample is rather small. Even a sample size of 1 ($n=1$) is possible. After all, the goal is not to make empirical generalizations or representative statements. Instead, qualitative research aims at generating detailed, information-rich cases which contribute to a better or in-depth understanding of the object of

investigation (ibid.). In contrast, *quantitative research* usually relies on a large sample that is randomly selected and intended to provide information about a certain population as a whole. This sampling strategy is referred to as *random sampling*. The logical structure and power of quantitative methods are grounded in statistical probability theory. “A random and statistically representative sample permits confident generalization from a sample to a larger population. Random sampling also controls for selection bias. The purpose of probability-based random sampling is generalization from the sample to a population and control of selectivity errors” (ibid.).

Sampling and access to the field are often regarded as incidental (Lau and Wolff, 1983, p. 417). However, it is worthwhile exploring questions of field access, not least because the way in which researchers manage to access a field is indicative of certain central field characteristics (Lüders, 2012, p. 392) and opens up another source of potential knowledge (Wolff, 2012, p. 336). Field access is also an interesting subject as there is no one-size-fits-all approach to looking for and getting access to a field (ibid.). Therefore, at the beginning of each research project, researchers have to think about how (best) to access the field in question and to select their sample (Felbermayr, 2023, p. 66).

How large should the sample be?

The first step is to define the target group, the selection criteria and the size of the sample. Contract-based research projects that receive external funding tend to be subject to predefined sampling criteria (Kuckartz, 2006, p. 277). For example, if an institution wants to evaluate whether teachers use its financial literacy program, the target group is implicitly defined by the question. Research interests are usually decisive for choices regarding the fine-tuning of the sample (e.g. teaching subject, teachers’ age or gender). There are no widely accepted rules for determining the adequate sample size. It varies from project to project and is largely determined by the research interests, the qualitative method of choice, the size of the target population and the desired heterogeneity. Compared to quantitative research, the sample size is significantly smaller because the aim is to generate in-depth, contextual knowledge and not to be able to make statements about the general population (Yoong et al., 2013, pp. 143 and 145). Instead of a concrete sample size in the form of a number, qualitative research aims at a state called theoretical saturation, a term coined by Grounded Theory³. Theoretical saturation is reached when additional surveys do not reveal, or cannot be expected to reveal, any additional knowledge (Charmaz, 2014). This point marks the ideal sample size (Yoong et al., 2014).

Which sampling strategy should be pursued?

The first question to clarify is whether the participants for qualitative surveys should be recruited from a larger quantitative sample or whether the sample should be selected directly from the population (Yoong et al., 2013, p. 144). Yoong et al. (2013) distinguish between different strategies for qualitative sampling.

- **Purposeful sampling:** Evaluations pursue a clear objective or research interest. In most cases, the selection of participants is purposeful in order to obtain the information necessary to answer the research questions. The aim of this type of sampling is to increase the credibility and depth, but not the representativeness of the results (Yoong et al. 2013, p. 143). There are many different types of purposeful sampling. Patton (2015, pp. 264–272) provides a comprehensive list of 40 purposeful sampling strategies including detailed descriptions.

³ Grounded Theory refers to a theory that is developed directly from the qualitative data collected (Charmaz, 2014, p. 6), i.e. the theory is grounded in the subject matter. There are specific methods for this, such as initial sampling or intensive interviewing for data collection, as well as initial and focused coding for data analysis. The foundations of Grounded Theory were established in the 1960s by the sociologists Barney G. Glaser and Anselm L. Strauss.

- **Maximum variation sampling:** Here, the goal is to capture a maximum of heterogeneity in the sample. Research can aim at diversity in accordance with predefined criteria (such as educational qualifications, migration background or age) in order to take different perspectives and experiences into account (ibid.). In this context, Kelle and Erzberger (2006, p. 293) point out that capturing and describing heterogeneity or the greatest possible diversity often requires a larger number of cases to carry out a contrasting analysis. Researchers must take this into account when planning evaluations in terms of the resources they require. After all, a larger number of cases also means more time for analysis.
- **Outlier sampling:** During sampling, it is also possible to specifically select outliers, i.e. cases that are unusual or outside the norm in one respect or another (Yoong et al., 2013, p. 143). If, for example, researchers test students at a school for their competencies in certain areas of economics, outlier sampling would mean to select the students with the best and worst results (outliers on both ends of the spectrum).
- **Snowball sampling:** Snowball sampling is a very popular sampling strategy. It means starting with a small group of people and recruiting new participants from their contacts or networks (ibid., p. 143). This strategy can be particularly effective for sensitive topics or target groups that are difficult to reach. If you are interested in what young people know about debt or how they feel about it, for example, you would first discuss the subject with a few of them. In a next step, you could then ask them to tell their friends about the study and the opportunity to participate. The idea is that motivated participants will hopefully be able to inspire others.
- **Theoretical sampling:** I would like to add to the list by Yoong et al. (2013) strategies that researchers rely on in combination with specific methods or research designs. Theoretical sampling is a well-known approach to sampling for recruiting participants for Grounded Theory studies. *Theoretical sampling* differs from conventional sampling strategies in that the first steps in accessing the research field have to be seen as provisional. Based on an initial, analytical knowledge base and first theoretical insights, additional participants have to be recruited throughout a research project. This allows for participants to join at a later stage. The open character of theoretical sampling is in line with Wolff's (2012, p. 336) claim that researchers should understand paving the way into the field as an ongoing (and active) task. Recruiting additional participants along the way is difficult to achieve for longitudinal studies where researchers should interview the same people several times over an extended period of time (Felbermayr, 2023, p. 67). This also applies to evaluations, which are often subject to time pressure, which makes it impossible to add participants later on.

How to get in touch with your sample?

Identifying and overcoming gatekeepers is a key step in accessing the research field. Gatekeepers are usually in key positions and can enable or prevent researchers' access to the field (Helfferrich, 2011). As intermediaries, Lewis and Porter (2004, p. 192) consider researchers to be in a position of power because at the interface they are "shaping what is researched and whose voices are heard," which significantly influences our perception of the field. Therefore, it is important to consider in advance *who* the relevant gatekeepers are and *how* one can get past them. When interviewing teachers, for example, the school principals and authorities in charge usually also have to give their permission. If minors are involved, their parents' consent will always be required. Obtaining the necessary permits and getting past gatekeepers can cause delays. The evaluation design should thus allow enough time for these steps.

There are various ways of contacting the sample. Let us assume researchers are planning to evaluate financial education activities in some schools. To this end, they will run focus groups with pupils. One option is to make contact in person on site, i.e. the researchers go to these schools and present their project

in class with the schools' consent. This has the advantage that there is no selection by third parties. A second variant would be making contact via a third party. Researchers entering schools are always subject to institutional restrictions. Teachers, acting as gatekeepers, are likely to specifically address certain students more so than others. Gatekeepers' preselection, assumptions and prejudices would then largely predetermine the sample (Felbermayr, 2023, p. 69). Preselection could, for example, exclude people with intellectual disabilities from an interview if gatekeepers considered their verbal skills not advanced enough to express themselves adequately (Buchner, 2008, p. 518). Following Lau and Wolff (1983, p. 418), researchers need to ask themselves: What can gatekeepers do? Who influences them? Do they receive instructions? And consequently, whom are they allowed to let past? A third option is to contact potential participants online. Here too, researchers can either contact them directly (if, for example, they know their e-mail addresses) or gatekeepers can forward e-mails to them.

How can people express interest in participation?

To gain access to the field, it is essential to produce information material that is appropriate for the target group. Information about the project (process, objectives, handling of data, etc.) can be provided in the form of handouts or folders, for example. In addition, researchers need to prepare consent forms and ideally send them to participants before they start gathering data.

Information material should always contain the researchers' contact details, so that interested people can get in touch with them directly. From a research ethics perspective, forwarding contact information to other scientists is a touchy issue. Griffin und Balandin (2004, p. 73) alternatively refer to "information drop" as a more ethical practice. In this case, potential participants receive the researchers' contact information and may decide whether they want to get in touch with them or not. This, in turn, can result in a rather specific sample of highly motivated and committed participants, which must be taken into account when analyzing the data.

5 Collecting data

Collecting too much data is a major risk as it might be impossible to analyze the data adequately then, due to a lack of time or human resources. This risk applies to the collection of both qualitative and quantitative data. People tend to think that collected data do not need to be analyzed later. "However, once collected, data pile up, with a pyramiding effect in terms of data processing and analysis effort (as well as adding to the cost of data collection)" (Hatry and Newcomer, 2015, p. 710). In this section, I am providing methodological advice for the collection of qualitative data. The following remarks will always remain incomplete, as researchers can come across countless different challenges, given the diversity of research contexts, methods or target groups they may work with.

Let us assume that, in accordance with their research interests, researchers have already selected an appropriate method and developed, validated and revised a survey guide. Once access to the research field has been gained and the sample has been determined, the actual data can be collected. For qualitative evaluation research projects, tensions arise due to the different demands of qualitative research and evaluation. Qualitative research is characterized by a methodological openness in the research process, which is hardly given when conducting evaluations and externally funded, contract-based research projects. In qualitative basic research, data are often collected in an iterative process. Researchers thus still get to revise their research questions or add new participants to the sample while already collecting data. This openness does not apply to research that is funded by third parties and evaluations, where clients have a specific research question in mind. As a result, the circularity of the research process that characterizes qualitative research is also lost (Kuckartz, 2006, p. 276).

Conducting the interview: Establishing rapport plays an important role in qualitative research, where data are often collected through direct contact with people. It is important to reflect on one's own role as a researcher as well as one's relationship with the participants and to ensure, for example, that participants do not confuse a professional relationship with friendship (for such ethical challenges, see Felbermayr and Lorenz, 2024). In addition to getting the relationship dynamics right, the actual conduct of interviews and focus groups is also demanding for interviewers. "It is very important that the interviewers do not influence what the respondent says and, above all, that they allow and encourage the respondent to speak at length on the topics to be covered. Interviewers should have well-developed listening skills and be familiar with techniques to probe replies and encourage the respondent to elaborate [...]" (Yoong et al., 2013, p. 148). Regarding wording choices during the interview, the same aspects are relevant as when creating guides. Researchers need to find the right balance in terms of complexity (not too challenging, but challenging enough), ask an appropriate number of questions, and choose terminology that is adequate for the target group as well as appreciative and neutral.

Technical equipment: Researchers often make audio or video recordings of interviews and focus groups. They should check how their equipment works ahead of conducting the interview, in order to solve potential problems in advance. Before recording, they should also make sure that there is enough memory left on their recording devices and bring spare memory cards as well as batteries on the day. If, unexpectedly, the recording does not work or a person does not agree to a recording, researchers should write down what was said during the interview right after. However, analyzing interview content as remembered by an interviewer is not the same as analyzing a transcript. After all, the former depends on the interviewer's memory capacity, which can result in (non)arbitrary omissions or subjective distortions (Kelle and Erzberger, 2006, p. 295).

5.1 Notes

Notes should be written immediately after data collection. This documentation should contain information about both interviewer and interviewee as well as the interview frame (such as location, duration, etc.) (Helfferich, 2011, p. 193). The notes must be treated confidentially. Researchers use them to document the research process. According to Helfferich (2011, p. 193), research notes consist of shorthand notes about the interview atmosphere and specific aspects of the interviewer's personal relationship with the interviewee. This can help prompt the interviewer's memory of a specific interview, particularly after a longer period of time. Notes thus serve to reconstruct the interview situation as well as to self-reflect (Charmaz, 2014).

Example of notes

There are no strict rules for writing notes. The notes should be adapted to the method used, the context in which data were collected and the needs of the researcher. In my own work, for example, I have used Helfferich's (2011, p. 201) template and included additional aspects, such as the weather, the time of day and seating arrangements (Felbermayr, 2023, p. 82).

Table 9

Example of notes

Background information	Interview scenario
Interview code:	Location:
Name of the interviewer:	Before the interview:
Date and time:	Seating arrangements:
Weather on the day of observation:	Particular incidents during the survey:
Attendants:	Perceived atmosphere:
	Interaction/difficult parts:
	After the conversation:

Source: Helfferich (2011, p. 201); Felbermayr (2023, p. 82).

6 Analysis

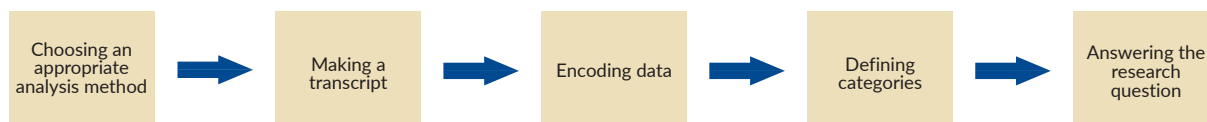
Once the data have been collected using an appropriate method, they must be analyzed. The logic and steps involved in analyzing qualitative data differ from those involved in analyzing quantitative data. What both have in common, though, is that, metaphorically speaking at least, data collection never takes place in a vacuum. This means that data are always collected with a specific research interest in mind and should help answer a specific question (Yoong et al., 2013, p. 65). When analyzing their data, researchers should not lose sight of the goal of their evaluation. Furthermore, the analysis of both qualitative and quantitative data must be rigorous, systematic and intersubjectively comprehensible (ibid., p. 172).

Just as making the right methodological choices at the beginning of the evaluation cycle, researchers must also select an appropriate analysis method (*choosing an appropriate analysis method*). Ideally, they already pick an analysis method at the beginning of the research process as the methods they use for data collection and analysis must be a good fit. For example, relying on a method that focuses on group dynamics (as is the case with focus groups) would hardly work when analyzing an individual interview. After interviews, focus groups and observations, audio recordings or observation notes are used to make transcripts (*making a transcript*). These transcripts will then be subject to qualitative analysis. The different strategies for analyzing qualitative data follow a similar structure: The *encoding of data* precedes the *definition of (more abstract) categories*. The methodological background of the selected analysis method determines how the individual coding steps are designed and how the categories are defined. At the end of the analysis process, researchers must *answer the research question* and write a final report.

Figure 1 illustrates the steps described above, which underly many methods used to analyze qualitative data. Once they have selected a specific method (e.g. content analysis according to Mayring, 2018; Grounded Theory, Charmaz 2014), researchers should still thoroughly examine the literature on the selected method. For example, many qualitative data analysis methods require researchers to encode the data, but the coding procedure differs depending on the underlying methodological approach. Coding is not a uniform process. It can vary greatly. Furthermore, figure 1 suggests that analysis is a linear process. However, in many cases this does not correspond to the iterative and procedural character of qualitative research in reality.

Figure 1

Main steps of qualitative data analysis



Source: Author's compilation.

6.1 Choosing an appropriate analysis method

A wide range of methods for qualitative data analysis corresponds to the diversity of qualitative methods for gathering data. In many cases, more than one method can be used to analyze qualitative data. For example, no standardized procedure applies to analyzing focus groups. Which analysis method to choose depends on one’s research interests. If the focus is primarily on content (What is being said?), researchers are more likely to choose a method for content analysis. If researchers are more interested in developing subjects or in learning about the opinions of a group (What do people say and how do they say it?), they will choose methods that allow them to focus on the group in their analysis. As a point of orientation, Goodrick and Rogers (2015, p. 566) have created a useful heuristic to classify the various qualitative analysis methods. They distinguish between four groups of methods; each group has a different primary focus and objective: enumerative, descriptive, hermeneutic and explanatory. Table 10 provides an overview of these methods. A more detailed description follows below.

Table 10

Four groups of methods for analyzing qualitative data

Primary purpose	Description	Examples of methods
Enumerative	Summarizing data in terms of discrete and often a priori categories that can be displayed and analyzed quantitatively	Classical content analysis (Krippendorff, 2013) Word count Cultural domain analysis (pile sorts, free lists) (Spradley, 1980) Ethnographic decision models (Gladwin, 1989)
Descriptive	Describing how concepts and issues are related	Matrix displays (Miles et al., 2014) Timelines Concept maps/mind maps (Trochim, 1989) Template/framework analysis (Crabtree and Miller, 1999; Ritchie and Spencer, 1994)
Hermeneutic	Identifying or eliciting meanings, patterns and themes	Thematic analysis (Boyatzis, 1998) Constant comparative method (Strauss and Corbin, 1998) Thematic narrative analysis (Riessman, 2008) Discourse analysis (Wetherell et al., 2001) Qualitative content analysis (Schreier, 2012)
Explanatory	Generating and testing causal explanations	Qualitative comparative analysis (Ragin, 1987) Process tracing (Collier, 2011)

Source: Goodrick and Rogers (2015, p. 567).

- Enumerative methods:** As the name suggests, enumerative methods convert qualitative data into numbers. Qualitative data are prepared in such a way that they can later be analyzed quantitatively. This is done “by sorting the data into coding frameworks, tallying the data, and then developing categories” (Goodrick and Rogers, 2015, p. 566). Enumerative methods can be used to analyze open-ended interview questions or to depict patterns in data, such as frequencies. An enumerative method is not suitable for analyzing rich data (e.g. in-depth interviews) “as it [the method] tends to be overly reductionist and can decontextualize or distort the meaning” (ibid., p. 567).
- Descriptive methods:** Descriptive methods are used to illustrate how different entities (e.g. ideas, stakeholders, aspects of a certain problem) are related to each other. For this purpose, the data are usually presented in tables or diagrams to allow for comparisons and contrasting (ibid., p. 566). “While enumerative methods are descriptive, our classification of these methods as descriptive indicates that the intended focus and product is not enumeration” (ibid., p. 572).

- **Hermeneutic methods:** Hermeneutic methods are characterized by (a) an iterative process between the collection and analysis of qualitative data and (b) the recognition of the role of researchers in producing data (ibid., p. 576). The focus is on identifying or determining manifest and latent meanings in the data (ibid., p. 566). “Manifest meanings are visible, descriptive labels that appear in the data. Latent meanings are underlying meanings, gleaned from an iterative process of examining the material, looking for similarities and differences, and identifying themes” (ibid., p. 576). Hermeneutic methods are suitable for analyzing rich data and provide in-depth insights. However, hermeneutic analyses are significantly more time-consuming than enumerative or descriptive analyses (ibid., p. 579–580).
- **Explanatory methods:** Here, the focus is on testing causal hypotheses. To do so, researchers establish and analyze theories about the relationship between a cause and its effect. Explanatory methods are suitable for evaluations when, for example, clients want to know whether a newly developed measure works and how it has helped achieve the desired results (ibid., pp. 566 and 580).

Qualitative methods for analyzing data are often characterized by a high level of accuracy, which is associated with being very time-consuming. This often conflicts with the limited time resources available for conducting evaluations (Flick, 2006a, p. 21). This is particularly the case with time-consuming qualitative methods (e.g. Grounded Theory), which do not fit into the tight time constraints of evaluation designs and often require a pragmatic approach (Kuckartz, 2006, pp. 278–279). One possible way out of this dilemma is to use so-called shortcut strategies. What is meant here are pragmatic ways of analyzing data (Kuckartz, 2006, p. 278) or justified deviations from maximum requirements for completeness and accuracy (Flick 2006a, p. 21). In line with the requirement for intersubjective comprehensibility, it is important for researchers to be transparent about how they apply shortcut strategies, as these must also meet scientific quality standards (Kelle and Erzberger, 2006, p. 285; Kuckartz, 2006, p. 280). Examples of shortcut strategies:

- **Transcript:** An appropriate strategy could be to transcribe only relevant parts of an interview, as opposed to the entire conversation, which is usually the state of the art in qualitative basic research (Kelle and Erzberger, 2006, p. 294). However, the parts of the interview that are not transcribed should then be paraphrased accordingly (Flick, 2006a, p. 21).
- **Coding:** Similarly, it is possible not to encode an entire transcript. The focus can be on segments that are relevant for answering the research question.

6.2 Making a transcript

Making a transcript, i.e. writing down what was said in the course of an audio recording, is an essential step in the research process. Here, it is important to keep in mind that a transcript is already a first interpretation. After all, the researchers decide, for example, whether an expression is transcribed as emphasized or not (Felbermayr, 2023). Researchers can make their own transcripts or use contract agencies. Agencies must be provided with a detailed set of transcription rules. Aspects of data protection, such as ensuring secure data transfer, must also be taken into account. Today, a wide variety of free and paid software is available to produce transcripts. Software programs have the advantage that certain keyboard shortcuts can be used to transcribe much faster. However, such programs must be used carefully, and researchers should consider questions like: Where are the (original) data uploaded to? Where are the data stored (e.g. in a cloud)? Who has access to the data? The same approach applies to using AI-supported transcription software.

Depending on the purpose of a research project, different rules for transcribing apply, which leads to varying levels of depth and detail. A rough distinction can be made between simple and in-depth transcription rules. When deciding which to apply, it is important that researchers consider the objective of their analysis and the time available. An in-depth transcript serves to reconstruct the interview situation in the best way possible. Making such a transcript is very time-consuming. Therefore, it should only be

made if the research question requires a more in-depth analysis of the data; or if the in-depth transcript adds analytical value (Kuckartz, 2006, p. 274). Analyzing a biographical interview using reconstructive methods may require a more in-depth transcript. If the way of speaking does not need to be reproduced in detail in order to answer the research question, it is preferable to opt for a transcript that is based on simple transcription rules and focuses on the content. Applying simple transcription rules means getting a transcript of an audio file in a relatively simple form. The audio file gets transcribed word by word while spelling and syntax errors do not get corrected. However, pauses, for example, are only indicated approximately (duration as felt) by a hyphen in the transcript, e.g. “-” for a short pause. In an in-depth transcript, however, pauses would be indicated in brackets, e.g. “(1 sec.)” to mark a pause of one second. Table 11 contains further examples of simple and in-depth transcripts.

While transcribing, it is also necessary to determine whether and, if so, how expressions in a specific dialect are to be rendered. To ensure anonymity, it may be appropriate to translate local dialect into standard language. Otherwise, dialect, as a personal characteristic, might help identify the speaker. From a research ethics and data protection perspective, this aspect must be taken into account, particularly for research involving small and vulnerable groups (Felbermayr, 2023, pp. 81–83).

Table 11

Examples of simple vs. in-depth transcription rules

Examples of simple transcription rules	Examples of in-depth transcription rules
Indicate pauses approximately, e.g. “-” = for a pause of medium length.	Specify pauses in the transcript, e.g. “(1)” = for a pause lasting one second.
Write down filler words (e.g. hm, ah).	Indicate a switch of speakers using empty lines as well as exact time stamps.
Indicate a switch of speakers using empty lines, but no time stamps.	Write down or indicate when speakers talk at the same time.
Underline words that speakers emphasize, e.g. I rode the bike.	Include comments about expressions regarding the speaker’s emotional state (e.g. laughing) or about non-verbal expressions (e.g. coughing, clearing their throat).
	Reproduce unfinished words, e.g. Good mor_.
	Reproduce phonetic extensions, e.g. Go=od morning.

Source: Author’s compilation.

6.3 Encoding data

Coding is “a central part of all qualitative data analysis” (Yoong, 2013, p. 174). A code is “a descriptive word or phrase that is intended to describe a fragment of data” (Goodrick and Rogers, 2015, p. 564). The type of coding depends on a researcher’s method of choice, how findings shall be obtained and the role ascribed to (theoretical) assumptions. First, a *deductive* approach to coding requires that codes and categories are developed in advance that are deemed relevant for answering the research question. Once obtained, data are then sorted according to the previously developed codes and categories. During the analysis, new codes and categories can and should always be added. Second, an *inductive* approach requires that codes are developed based on the collected data, as opposed to developing codes and categories before collecting data. Third, an *abductive* approach refers to a creative, process-based type of analysis that is typical of constructivist Grounded Theory (Charmaz, 2014).

The extent of coding depends on various factors, such as type of data, level of abstraction, aim of the study or stage of the research process (Charmaz, 2014, p. 128). Three popular forms of coding are word-by-word, line-by-line and incident-with-incident coding (ibid., 133). A researcher must be transparent about and justify their choice of one way of encoding over another, because “[t]he size of the unit of data to code matters” (ibid., p. 124). The way of coding can be changed while analyzing data. However, for transparency reasons, changes must also be disclosed. Word-by-word coding means that there is a code for each word, while line-by-line coding focuses on a larger data unit. Line does not mean a complete and

grammatically correct sentence, but one line in the transcript. Accordingly, researchers analyze the transcript line by line (Charmaz, 2014, p. 124; Felbermayr, 2023, p. 84). The two types of coding described (word, line) are probably too detailed and time-consuming for analyses in the context of evaluations. Incident-by-incident coding is recommended, where entire paragraphs or events in the transcript are assigned a code. The codes can operate at different levels. In a first step, they are mostly descriptive. As the analysis moves forward, the codes become increasingly analytical and abstract (Yoong et al., 2013, p. 174). So-called *in vivo* codes are a way of directly representing the perspective of participants. These are codes taken word by word from a transcript. As *in vivo* codes represent the language of the interviewees, they serve as “symbolic markers of participants’ speech and meaning” (Charmaz, 2014, p. 134).

Researchers can use software for coding or do it manually (paper-pencil technique). Various software programs support the process of encoding and analyzing qualitative data. Goodrick and Rogers (2015, p. 585) divide these into the following four groups:

- (1) “Word processing and spreadsheet software, such as Word and Excel, which can be used for simple coding and analysis;
- (2) specialized qualitative data analysis software, such as NVivo, HyperResearch, MaxQDA, and Atlas-TI;
- (3) machine learning software, such as MonkeyLearn;
- (4) visual analysis software that produces word cloud or network diagrams of text.”

Using software for data analysis can be helpful in the research process. Programs for analysis “merely enable evaluators to store, organize, search, categorize, and group large volumes of mostly text-based data” (Yoong et al., 2013, p. 175). However, coding is still done by the researchers, which is why the software itself, despite all its advantages, is not an analytic solution (Birks and Mills, 2011, p. 101). There is also a risk that software will generate too much code and thus an unmanageable amount of data (Goodrick and Rogers, 2015, p. 586). At the same time, computer programs can help (better) organize and analyze the amount of data involved in large research projects with several researchers. However, sometimes researchers still prefer manual data analysis, as it gives them a more holistic view of the data. This may well be the case for in-depth analyses of individual interviews (Yoong, 2013, p. 175).

6.4 Defining categories

Coding means labeling text segments that can be smaller or larger. In a next step, the individual codes are grouped into categories. A category thus comprises a varying number of codes on a similar topic, problem, etc. in the data (Goodrick and Rogers, 2015, p. 564). For example, if researchers want to evaluate the implementation of a new financial education program, they can group codes that deal with teachers’ experiences. This group of codes then needs to be further analyzed and categorized, e.g. by creating a category with negative and a category with positive experiences (*ibid.*). In the course of coding, researchers must find a way to systematically abstract away from or summarize the abundance of codes obtained (Nightingale et al., 2015, p. 470).

6.5 Answering the research question

Answering the research question is the final step in the research process. Usually, a final report summarizes the main results of the evaluation through the lens of the research questions. The requirements for such a final report must be clarified in advance and often depend on client expectations, the target group, disciplinary requirements, etc. According to Mensching (2006, p. 340), clients expect not only a descriptive presentation of the results, but also practical evaluations and recommendations for action. There is no standardized way of presenting results. Researchers can visualize qualitative results in different ways, also using a variety of software programs. In many cases, the visualization will depend on the evaluation method and the research question. If a research project investigates the relationship between different categories (how aspects are connected), the visualization should reflect this.

In a scientific context, research projects are assessed based on quality criteria. According to Mayring (2018, p. 21), quantitative quality standards such as objectivity, reliability and validity can be a point of reference for qualitative research but are not sufficient on their own. They must therefore be supplemented with a view to assessing or evaluating the quality of qualitative research processes. Qualitative evaluations must adhere to the quality standards of qualitative research in general. These include, for example, intersubjective comprehensibility, reflective subjectivity or empirical anchoring (Mensching, 2006, p. 340). In this context, the literature discusses the extent to which there can be uniform standards at all, given the diversity of approaches and methods for qualitative research projects (Flick, 2006c). Flick (*ibid.*, p. 439) wonders, for example, how to measure how much knowledge was actually gained in an exploratory study or how to evaluate the appropriateness of the methods used for a specific research field and question. Assessing the originality of a researcher's access to the field or their methodological approach is equally challenging (*ibid.*). The same holds true for assessing creativity in dealing with data as well as the relationship between the single steps involved in the research process and the process as a whole (*ibid.*).

Leaving these questions aside, stakeholders agree that there need to be criteria for assessing research projects. In addition to standards for qualitative and quantitative research, specific criteria for evaluation studies can also be used, such as those of the Joint Committee on Standards (JCS) of the American Evaluation Association (AEA), the SEVAL standards by the Swiss evaluation society SEVAL or the standards of the German evaluation society DeGeVAL (Caspari, 2006).

7 Summary and concluding remarks

How can qualitative evaluations be carried out? Which are the methodological steps that need to be considered? Which qualitative methods can researchers use to collect and evaluate data? What are the advantages and disadvantages of each method? In which contexts is qualitative evaluation actually appropriate?

These are the questions this report is meant to provide answers to. It gives an overview of the main steps in the different phases of qualitative evaluations. These include planning, implementation, analysis and communication. In section 1, I introduced the characteristics of qualitative research. An important step in the planning phase is selecting adequate methods for collecting and analyzing data. Advantages and disadvantages of common qualitative surveying methods – interviews, focus groups, observations and desk reviews of documents – are mentioned in section 2. I discussed the development of appropriate survey instruments for each method in more detail in section 3. Next, section 4 dealt with a topic that is often strongly underrepresented in the literature, although it is essential for implementing evaluations or studies successfully: sampling and access to the research field. In section 5, I discussed how to apply the different methods, while section 6 focused on the analysis of qualitative data.

Note that I only provided an overview of the individual methodological steps in a qualitative evaluation cycle. It is advisable to explore these methods further, e.g. when making decisions about methods for gathering and analyzing data. An essential step here is adapting methods to the respective target group. Developing a guide to interview children, for example, poses different challenges for researchers than developing a guide for adults. To ensure that research in the field of inclusion and disability is not research about, but research with the target population, people with disabilities must take part in relevant research projects, regardless of their disabilities, and researchers must make methodological adaptations catering to their individual needs (Coons and Watson, 2013; Felbermayr, 2023; Unger, 2014). Researchers can face numerous challenges throughout evaluations. Hatry and Newcomer (2015) provide a list of 27 pitfalls to be avoided during the evaluation process (from planning to implementation and the dissemination of results). I recommend reading their article in the planning phase of an evaluation project to avoid methodological pitfalls. It is also important to consider challenges regarding research ethics and data protection regulations, which are covered in another publication of the OeNB Financial Literacy Evaluation Series.

References

- Adams, W. C. 2015.** Semi-structured interviews. In: Newcomer, K. E., H. P. Hatry and J. S. Wholey (eds.). Handbook of practical program evaluation. 4th edition. New Jersey: Jossey bass. 492–505.
- Birks, M. and J. Mill. 2011.** Grounded Theory. A practical guide. Thousand Oaks: SAGE.
- Bohnsack, R. and B. Schäffer. 2001.** Gruppendiskussionsverfahren. In: Hug, T. (ed.). Einführung in die Forschungsmethodik und Forschungspraxis. Baltmannsweiler: Schneider-Verlag. 324–341.
- Boer, H. and S. Reh. 2012.** (eds.). Beobachtung in der Schule – Beobachten lernen. Wiesbaden: Springer VS.
- Bortz, J. and N. Döring. 2009.** Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. 4th revised edition. Heidelberg: Springer.
- Buchner, T. 2008.** Das qualitative Interview mit Menschen mit so genannter geistiger Behinderung – Ethische, methodologische und praktische Aspekte. In: Biewer, G., M. Luciak and M. Schwinge (eds.). Begegnung und Differenz. Menschen – Länder – Kulturen. Beiträge zur Heil- und Sonderpädagogik. Bad Heilbrunn: Klinkhardt. 516–528.
- Caspari, A. 2006.** Partizipative Evaluationsmethoden. Zur Entmystifizierung eines Begriffs in der Entwicklungszusammenarbeit. In: Uwe, F. (ed.). Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen. Reinbek bei Hamburg: Rohwohlt. 365–384.
- Charmaz, K. 2011.** Grounded Theory konstruieren. Kathy C. Charmaz im Gespräch mit Antony J. Puddephatt. In: Mey, G. and K. Mruck (eds.). Grounded Theory Reader. Wiesbaden: VS Verlag für Sozialwissenschaften. 89–107.
- Charmaz, K. 2014.** Constructing grounded theory. 2nd Edition. Thousand Oaks: SAGE.
- Charmaz, K. and A. Bryant. 2008.** Grounded Theory. In: Given, L. M. (ed.). The Sage encyclopedia of qualitative research methods. Volume 1. A – L. Los Angeles u. a.: SAGE. 374–377.
- Coons, K. D. and S. L. Watson. 2013.** Conducting research with individuals who have intellectual disabilities: Ethical and practical implications for qualitative research. In: Journal on developmental disabilities 19(2). 14–24.
- Cropley, A. 2002.** Qualitative Forschungsmethoden. Eine praxisnahe Einführung. Frankfurt am Main: Verlag Dietmar Glotz.
- Denzin, N. and Y. S. Lincoln. 2018.** The Sage handbook of qualitative research. 5th edition. Thousand Oaks: SAGE.
- Dieckmann, A. 2017.** Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. 11th edition. Reinbek bei Hamburg: Rohwohlt Taschenbuch Verlag.
- Felbermayr, K. 2023.** Entscheidungsprozesse am inklusiven Übergang. Eine Grounded Theory Studie im Längsschnitt. Bad Heilbrunn: Julius Klinkhardt. https://www.pedocs.de/frontdoor.php?source_opus=26629
- Flick, U. 2006a.** Qualitative Evaluationsforschung zwischen Methodik und Pragmatik. Einleitung und Überblick. In: Flick, U. (ed.). Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen. Reinbeck bei Hamburg: Rohwohlt. 9–29.
- Flick, U. 2006b.** Interviews in der qualitativen Evaluationsforschung. In: Flick, U. (ed.). Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen. Reinbek bei Hamburg: Rohwohlt. 214–232.
- Flick, U. 2006c.** Qualität in der Qualitativen Evaluationsforschung. In: Flick, U. (ed.). Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen. Reinbeck bei Hamburg: Rohwohlt. 424–443.
- Flick, U. 2009.** Qualitative Sozialforschung. Eine Einführung. 2nd edition. Reinbeck bei Hamburg: Rohwohlt Taschenbuch Verlag.
- Flick, U., E. v. Kardorff and I. Steinke 2012.** Was ist qualitative Forschung? Einleitung und Überblick. In: Flick, U., E. v. Kardorff and I. Steinke (eds.). Qualitative Forschung. Ein Handbuch. Reinbek bei Hamburg: Rohwohlt Taschenbuch. 13–29.

- Goodrick, D. and P. J. Rogers. 2015.** Qualitative Data Analysis. In: Newcomer, K. E., H. P. Hatry and J. S. Wholey Handbook of practical program evaluation. 4th edition. New Jersey: Jossey bass. 561–595.
- Griffin, T. and S. Balandin. 2004.** Chapter 3. Ethical research involving people with intellectual disabilities. In: Emerson, E., C. Hatton, T. Thompson and T. R. Parmenter (eds.). The international handbook of applied research in intellectual disabilities. Chichester: John Wiley & Sons Ltd. 61–82.
- Hatry, H. P. and K. E. Newcomer. 2015.** Pitfalls in evaluation. In: Newcomer, K. E., H. P. Hatry and J. S. Wholey Handbook of practical program evaluation. 4th edition. New Jersey: Jossey bass. 701–724.
- Helfferich, C. 2011.** Die Qualität qualitativer Daten. Manual für die Durchführung qualitativer Interviews. Wiesbaden: Springer Fachmedien.
- Hirschauer, S. 2006.** Wie geht Bewerten? Zu einer anderen Evaluationsforschung. In: Flick, U. (ed.). Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen. Reinbek bei Hamburg: Rohwohlt. 405–423.
- Kelle, U. and C. Erzberger. 2006.** Stärken und Probleme qualitativer Evaluationsstudien. Ein empirisches Beispiel aus der Jugendhilfeforschung. In: Flick, U. (ed.). Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen. Reinbek bei Hamburg: Rohwohlt. 284–300.
- Krueger, R. A. and M. A. Casey. 2015.** Focus group interviewing. In: Newcomer, K. E., H. P. Hatry and J. S. Wholey (eds.). Handbook of practical program evaluation. 4th edition. New Jersey: Jossey bass. 506–534.
- Kuckartz, U. 2006.** Quick and dirty? Qualitative Methoden der drittmittelfinanzierten Evaluation in der Umweltforschung. In: Flick, U. (ed.). Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen. Reinbek bei Hamburg: Rohwohlt. 267–283.
- Kuckartz, U., T. Dresing, S. Rädiker and C. Stefer. 2008.** Qualitative Evaluation. Der Einstieg in die Praxis. 2nd updated edition. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lamnek, S. 2005.** Gruppendiskussion. Theorie und Praxis. 2nd revised and expanded edition. Weinheim und Basel: Beltz Verlag.
- Lau, T. and S. Wolff. 1983.** Der Einstieg in das Untersuchungsfeld als soziologischer Lernprozess. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 35(3). 417–437.
- Lewis, A. and J. Porter. 2004.** Interviewing children and young people with learning disabilities: Guidelines for researchers and multi-professional practice. In: British journal of learning disabilities 32(4). 191–197.
- Lorenz, T. 2024.** Mixed methods. OeNB Financial Literacy Evaluation Series.
- Lorenz, T. and K. Felbermayr. 2024.** Data privacy and research ethics. OeNB Financial Literacy Evaluation Series.
- Lüders, C. 2012.** Beobachten im Feld und Ethnographie. In: Flick, U., E. v. Kardorff and I. Steinke (eds.). Qualitative Forschung. Ein Handbuch. 9th edition. Reinbek bei Hamburg: Rowohlt Taschenbuch. 384–401.
- Mayring, P. 2018.** Gütekriterien qualitativer Evaluationsforschung. In: Zeitschrift für Evaluation 17(1). 11–24.
- Mensching, A. 2006.** Zwischen Überforderung und Banalisierung. Zu den Schwierigkeiten der Vermittlungsarbeit im Rahmen qualitativer Evaluationsforschung. In: Flick, U. (ed.). Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen. Reinbek bei Hamburg: Rohwohlt. 339–362.
- Michael, V. 2022.** Interviews [https://core-evidence.eu/posts/methods-toolkit-qual-interviews\(24/11/2023\)](https://core-evidence.eu/posts/methods-toolkit-qual-interviews(24/11/2023)).
- Monique, H. M. 2020.** Qualitative research methods. 2nd edition. London: SAGE.
- Morgan, D. L. 1998.** The Focus Group Guidebook. Thousand Oaks et al: SAGE.
- Nightengale, D. S. and S. B. Rossmann. 2015.** Collecting data in the field. In: Newcomer, K. E., H. P. Hatry and J. S. Wholey. Handbook of practical program evaluation. 4th edition. New Jersey: Jossey bass. 445–473.
- OECD/INFE. 2010a.** Guide to Evaluating Financial Education Programmes.
- OECD/INFE. 2010b** Detailed Guide to Evaluating Financial Education Programs.
- Patton, M. Q. 2015.** Qualitative Research & Evaluation Methods. 4th edition. Los Angeles et al: SAGE.
- Pauli, C. 2012.** Kodierende Beobachtung. In: Boer, H. d. and S. Reh (eds.). Beobachtung in der Schule – Beobachten lernen. Wiesbaden: Springer VS. 45–63.

- Schulz, M. 2012.** Quick and easy!? Focus groups in applied social science. Schulz, M., B. Mack and O. Renn. (eds.). Fokusgruppen in der empirischen Sozialwissenschaft. Von der Konzeption bis zur Auswertung. Wiesbaden: Springer Fachmedien. 9–22.
- Unger, H. v. 2014.** Partizipative Forschung. Einführung in die Forschungspraxis. Wiesbaden: Springer.
- World Bank. 2013.** A toolkit for the evaluation of financial capability programs in low- and middle-income countries. Washington, DC: The World Bank.
- Wolff, S. 2012.** Wege ins Feld und ihre Varianten. In: Flick, U., E. v. Kardorff and I. Steinke (eds.). Qualitative Forschung. Ein Handbuch. 9th edition. Reinbek bei Hamburg: Rowohlt Taschenbuch. 334–349.
- Yoong, J., K. Mihaly, S. Bauhoff, L. Rabinovich and A. Hung. 2013.** A Toolkit for the Evaluation of Financial Capability Programs in Low-, and Middle-Income Countries. Washington, DC: The World Bank.
- Zwick, M. M. and R. Schröter. 2012.** Konzeption und Durchführung von Fokusgruppen am Beispiel des BMBW-Projekts Übergewicht und Adipositas bei Kindern, Jugendlichen und jungen Erwachsenen als systemisches Risiko. In: Schulz, M., B. Mack and O. Renn. (eds.). Fokusgruppen in der empirischen Sozialwissenschaft. Von der Konzeption bis zur Auswertung. Wiesbaden: Springer Fachmedien. 24–48.