

# **Cognitive Pretesting**

*Timo Lenzner, Patricia Hadler, & Cornelia E. Neuert*

## **Abstract**

Cognitive pretests are used in the development of new or modification of existing survey instruments to ensure that the questions and items are understandable and interpreted as intended, to evaluate how difficult they are, and whether they pose other cognitive problems for respondents. Cognitive pretests are an important component in ensuring high-quality survey data. This GESIS Survey Guideline first provides an introduction to questionnaire pretesting and then goes into detail on the practical implementation of two cognitive pretesting methods: cognitive interviewing and web probing.

## **Citation**

Lenzner, Timo; Hadler, Patricia; Neuert, Cornelia E. (2024). Cognitive Pretesting. Mannheim, GESIS – Leibniz Institute for the Social Sciences (GESIS – Survey Guidelines).

DOI: [10.15465/gesis-sg\\_en\\_049](https://doi.org/10.15465/gesis-sg_en_049)



## 1. What is a pretest, and why should questionnaire pretests be conducted?

The term *questionnaire pretest* broadly refers to the evaluation or testing of survey questions before they are fielded in the actual survey. Pretests are an essential part of the questionnaire design process. The purpose of a pretest is to provide information about some or all of the following things (see Converse & Presser, 1986; Porst, 2014):

- The comprehensibility of the questions  
*Does the meaning that respondents associate with a question correspond to the meaning intended by the researcher? Do different respondents interpret the meaning of a question in the same way?*
- Difficulties that respondents have with their task  
*How difficult is it for respondents to understand and answer the question? Is the subject matter of the question unfamiliar or sensitive?*
- Respondent interest in, and attention to, individual questions  
*Do fatigue effects manifest themselves during the interview/while the questionnaire is being completed? Do respondents think that (individual) questions are redundant or not relevant for them?*
- Frequency distributions of the responses  
*Is the full range of the scale used? How large is the share of respondents who do not provide a substantive response (e.g., leave the question unanswered or select “don’t know”)?*
- Context effects and problems with the question order  
*Do earlier questions influence responses to subsequent questions?*
- The usability of self-administered questionnaires  
*Do respondents understand how to navigate through the questionnaire? Does the questionnaire look and function similarly on different devices (e.g., PCs, smartphones)?*
- Interviewer problems  
*Can interviewers recognize clearly what they are supposed to read out and to whom?*
- Technical problems with the questionnaire (e.g., missing or incorrect filter instructions) and with interview aids (e.g., lists, showcards)
- The duration of the interview/questionnaire completion

All these aspects provide important information about whether the individual questions and the questionnaire as a whole function as they should. The ultimate goal of a pretest is to optimize questions prior to their administration in the actual survey to ensure they gather high-quality (i.e., reliable and valid) data. To ensure high-quality data, desk-based appraisals or expert reviews of a questionnaire are generally not sufficient because, as Sudman and Bradburn (1982: 283) pointed out: “Even after years of experience, no expert can write a perfect questionnaire.” Only with empirical (pre)tests is it possible to check whether survey questions actually measure what they are supposed to measure and whether they yield reliable and valid responses.

## 2. What pretesting methods are available, and which method should you choose?

A diverse range of methods are available for pretesting questionnaires. They include, for example, cognitive interviewing, web probing, eye tracking, usability testing, split ballot experiments, paradata analyses, interviewer debriefing, respondent debriefing, and piloting (Beatty et al., 2019; Campanelli, 2008; Collins, 2015; Presser et al., 2004). Each of these methods has specific strengths and weaknesses and some can be combined within the same data collection (see Collins, 2015, Ch.

2 and sections 2.1, 3.4, and 4.4 of this Survey Guideline). The decision for or against a method depends primarily on the type of information one wants to obtain, the mode of the survey, and time and cost considerations. In what follows, we focus on two families of pretesting methods: (1) piloting, which is recommended for all surveys, and (2) cognitive pretesting, which should ideally be applied during questionnaire development and before piloting. While a pilot is conducted to test the entire questionnaire, cognitive pretesting involves in-depth testing of certain questions only.

## 2.1 Piloting

In a pilot test, the questionnaire is administered under conditions that simulate as realistically as possible those that will prevail in the main survey. In other words, a pilot test is a dress rehearsal for the main survey with a sample of typically 10 to 200 respondents (Prüfer & Rexroth, 1996). Pilot tests should be conducted in the mode (face-to-face, telephone, mail, online, etc.) that will be used in the main survey. To avoid influencing the response process, the respondents are usually not informed that the survey is a test run of the main survey. The objective of a pilot test is to check the feasibility of the survey administration and the functionality of the entire questionnaire. As a rule, it yields reliable information about (1) technical defects in the questionnaire (e.g., defective filter questions or problems that interviewers have administering the questionnaire), (2) frequency distributions of the responses and missing data, and (3) the average duration of the interview/questionnaire completion.

Pilot testing can easily be combined with pretesting methods such as split ballot experiments, in which two or more question versions are randomly presented to subgroups of respondents, paradata analyses (e.g., response times, keystrokes; see GESIS Survey Guideline “[Web Paradata in Survey Research](#)”, Kunz & Hadler, 2020), and interviewer or respondent debriefings. In addition to generating quantitative pilot data, the latter debriefing methods provide information about respondents’ understanding of survey questions and potential question problems. For example, when piloting interviewer-administered questionnaires, the interviewers can be instructed to make a note of, and to report, any difficulties that respondents experienced during the pilot interviews. Alternatively, at the end of a pilot questionnaire, respondents can be asked (some) follow-up questions about what they were thinking when they answered certain questions (respondent debriefing). When testing web questionnaires, these respondent debriefings can be carried out using web probing (see section 4.4 below).

In general, however, pilot testing is a passive pretesting method in which researchers merely observe respondents’ answer behavior without actively asking about their cognitive response processes. Whether respondents have difficulties understanding or answering a particular question is usually only discovered if they point this out themselves. Therefore, a pilot test can only provide comparatively little and unsystematic information about the respondents’ cognitive processes when answering survey questions.

## 2.2. Cognitive pretesting

Cognitive pretesting belongs to the active pretesting methods because the participants’ approach to answering the questions is actively probed and investigated (Beatty & Willis, 2007; Prüfer & Rexroth, 2005; Willis, 2005). Cognitive pretesting is usually conducted early in the questionnaire design phase in order to obtain insights into the cognitive processes that take place when survey questions are being answered:

- How do respondents interpret the questions or specific terms?
- How do they retrieve relevant information and events from memory?

- How do they arrive at a decision as to how they should respond?
- How do they assign their internally determined responses to the response categories provided?

The main objectives of cognitive pretesting are (1) to examine whether survey questions measure what they are supposed to measure (i.e., to evaluate their content-related validity; see GESIS Survey Guideline “Validity in Survey Research”, Repke et al., 2024) and (2) to obtain information about potential question problems. For example, cognitive pretests are particularly suitable for testing the comprehensibility of questions, identifying problems that respondents have answering the questions, establishing the causes of these problems, and generating suggestions for improvement on the basis of these findings.

In contrast to pilot testing, the focus of cognitive pretesting is on evaluating individual questions rather than the questionnaire as a whole. Cognitive pretests are generally suitable for all survey modes. In other words, they can be applied irrespective of whether the subsequent survey is to be administered face-to-face, by telephone, by mail, or online. Conducting a cognitive pretest is particularly recommended when new questions are being developed or existing questions are being adapted. However, it can also make sense to re-test questions for which standard quality criteria (e.g., reliability, validity) have already been established (see GESIS Survey Guideline “[Documenting Measurement Instruments for the Social and Behavioral Sciences](#)”, Schmidt & Lechner, 2020). Thereby one can investigate whether the interpretation of certain questions has changed over time due to changes in the social reality of respondents (Walter, 2018).

In the following, we present two specific cognitive pretesting methods and their practical implementation in more detail, namely cognitive interviewing and web probing.

### **3. Planning and conducting a cognitive interviewing pretest**

Cognitive interviews are semi-structured, in-depth interviews carried out by researchers or specially trained interviewers in individual interview sessions. They have traditionally been conducted in person, but it has also become common to conduct them in remote settings (i.e., over the phone or via video conference; see section 3.2.1 below). In cognitive interviews, participants are asked to answer the survey questions being evaluated and additionally provide information about how they proceeded in answering the questions. The method is characterized by an open interview situation that enables flexible interaction between interviewer and participant. Hence, cognitive interviews lend themselves very well for an intensive exploration of the respondents’ thought processes that led to their answers.

When planning and conducting a cognitive interviewing project, researchers need to consider various aspects. The most important of these are presented below.

#### **3.1. What techniques are (mainly) applied in a cognitive interviewing pretest?**

##### **3.1.1 Think aloud**

The think-aloud technique involves asking participants to verbalize all their thoughts while answering a question. The aim is to reveal their response process, and thus also any problems that participants may have in understanding and answering a question. A think aloud should always be prefaced by an instruction such as the following (Porst, 2014): “While you are answering the following question, can you tell me what you are thinking, or what is going through your mind? Please also mention things that may appear to you to be unimportant. The question is: ....” It is

advisable to repeatedly remind respondents to verbalize their thoughts while answering a questionnaire to prevent them from stopping.

Advantages of this technique include the fact that it is easy to implement and that the risk of interviewers influencing respondents' answers is comparatively low. A critical issue associated with thinking aloud is that most respondents find it difficult to apply this technique and many are not capable of verbalizing the cognitive processes that lead to their answer. Using think aloud also assumes that the process itself does not change respondents' thought processes (reactivity bias, see section 3.1.2 below).

### *3.1.2 Probing*

Probing is a technique that involves asking participants one or more follow-up questions (probes) about how they understood and answered earlier survey questions. Probes can be administered concurrently (i.e., directly after the participant has answered the survey question) or retrospectively (i.e., after he or she has answered a set of questions or the whole questionnaire). Concurrent probing ensures that the thought process is still in short-term memory, while retrospective probing does not disrupt the flow of the questionnaire or related questions (Willis, 2005). Depending on the knowledge interest, probes can, for example, focus on specific cognitive processes (e.g., comprehension probe, information retrieval probe) or ask participants to elaborate on their answer (category-selection probe). Table 1 provides an overview of the types of probes that are most commonly applied in cognitive interviews at GESIS. Probes can be scripted in advance of the interview or formulated spontaneously, for instance, based on participants' answers (emergent probing). In practice, often a combination of scripted and spontaneous probes is used. This ensures a standardized procedure and at the same time enables the interviewers to react flexibly in unforeseen situations.

One of the advantages of the probing technique is that probes are both easy to administer for cognitive interviewers and easy to answer for test participants. Another advantage is that probes are able to uncover "silent misunderstandings," that is, misinterpretations by respondents that might remain unnoticed if, for example, they were only asked to think aloud and not asked directly for their interpretation of a question or a term in a question. Overall, the probing technique reliably provides insights into the participants' cognitive response processes. On the downside, probing is prone to reactivity bias, meaning that respondents may alter their thought processes simply because they are asked to articulate them (Conrad & Blair, 2009). For instance, if an interviewer probes about the interpretation of a particular term in a question, this might prompt the respondent to consider various ways in which the question could be (mis)interpreted, even if they initially found it clear and unambiguous.

There are no universal rules as to how varied the cognitive techniques applied in a cognitive interview should be. On the one hand, a certain amount of variety is definitely a good way of avoiding tiring participants. On the other hand, merely for the sake of variety one should not apply techniques that may not yield any new knowledge. Rather, the application of a cognitive technique should always be determined by the knowledge interest of the researcher or by the behavior of the participant.

Table 1. Types of cognitive probing questions

<b>Type of probing question</b>	<b>Description and example</b>
Category-Selection Probing	Questions about the choice of answer category, e.g.: <i>"You have selected [answer] for this question. Please explain your answer in more detail. Why did you choose this answer?"</i>
Comprehension Probing	Questions on understanding, e.g.: <i>"What do you understand by 'a highly responsible professional activity' in this question?"</i>
Confidence Rating	Assessment of the reliability of the response, e.g.: <i>"How sure are you that you've seen a doctor in the last 12 months?"</i>
Difficulty Probing	Questions on the difficulty of answering, e.g.: <i>"How easy or difficult was it for you to answer this question?"</i> <i>If rather/very difficult:</i> <i>"Why did you find it rather/very difficult to answer this question?"</i>
Emergent Probing	Spontaneous questioning in response to an utterance or behavior of the test person, e.g.: <i>"You just frowned and laughed when I read you the answer options. Can you please explain to me why you did that?"</i>
General Probing	Non-specific questions, e.g.: <i>"Do you have any (further) comments on this question?"</i>
Process Probing	Questions on how the answer was formed, e.g.: <i>„How did you arrive at that answer? What was going through your mind?"</i>
Recall Probing	Questions on event recall, e.g.: <i>"How did you remember that you had been to the doctor 6 times in the last 12 months? Did you count or estimate the number of appointments?"</i>
Response Scale Probing	Questions on differentiating between response scale values, e.g.: <i>„Your answer on a scale of 0 to 10 was [answer]. Why did you choose that value rather than the value just above or below it?"</i>
Sensitivity Probing	Questions on the sensitivity of a question: <i>„Do you think that this question asks about things that are too private, or is it ok to ask this?"</i>
Specific Probing	Specific questions, e.g.: <i>"You answered 'yes' in this question. Does this mean that you have already given up on career opportunities for your family, or that you might be willing to give them up but have not yet done so?"</i>

## 3.2 How are cognitive interviews conducted?

### 3.2.1 Interview setting

Cognitive interviews should generally take place in a quiet environment and be audio recorded so that the interviewer can concentrate fully on the conversation during the interview and does not have to write down the participant's answers. Recording the interviews also facilitates subsequent data analysis. Generally, cognitive interviews should begin with a briefing of the respondents in which their task is illustrated. For example, respondents should be instructed to verbalize any problems and difficulties they experience when answering the questions, no matter how trivial they may seem.

When carrying out cognitive interviews remotely, there are two main aspects to consider: (1) which channel to use to communicate with respondents and (2) how to present them with the questions to be tested (see Figure 1 and GESIS Blog Post "[How to conduct cognitive interviews in times of COVID-19](#)", Lenzner et al., 2021). Remote cognitive interviews can take place via video conference or telephone. The advantage of video conferencing is that the interviewer and respondent see each other, and the interviewer can react to non-verbal cues. A disadvantage of this procedure is that it systematically excludes participants who have no access to Internet-connected devices. It is generally advisable to keep the technological burden for participants as low as possible, for example, by choosing a platform that allows users to participate directly in their browser without downloading a program or app and that can be easily accessed via both stationary devices and smartphones.

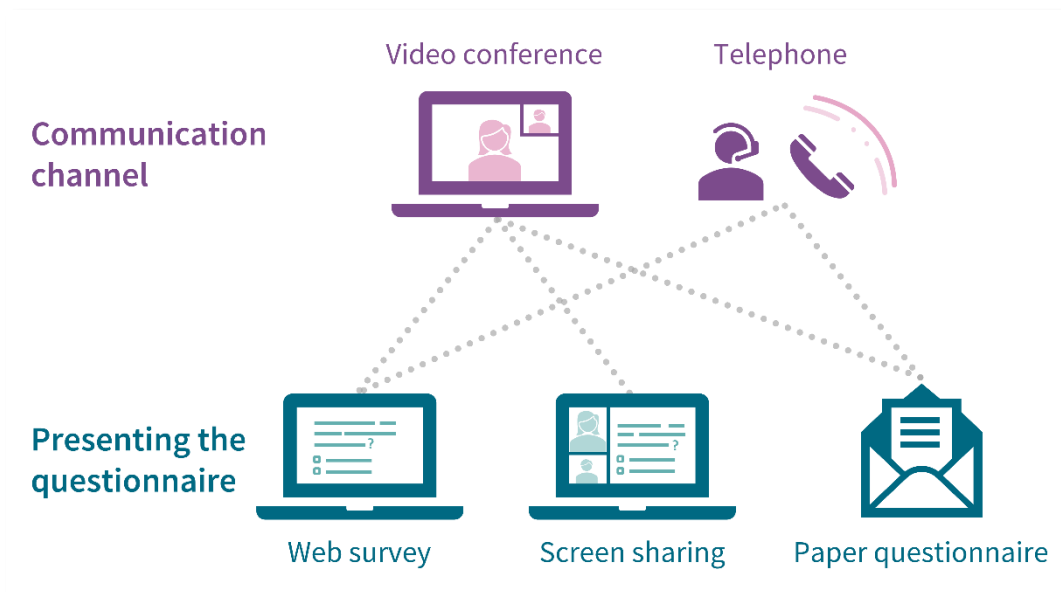


Figure 1. Remote cognitive interviewing settings

Compared to such a setting, carrying out interviews by phone initially appears to be an out-of-date choice of mode, unnecessarily forgoing the opportunities of modern technology. However, the advantage is that a phone call does not impose a technological obstacle to participants. This option is particularly useful for non-tech-savvy populations, but also for participants without a stable Internet connection or flat rate. Finally, respondents with concerns about data protection may be more inclined to participate in a telephone interview than to go online.

Once the channel of communicating with participants is clarified, researchers must consider how to present the questions to be tested to the participants. Interviewer-administered questionnaires do



not require adjustment to a remote setting, as the interviewer can read the questions and response options aloud and record the respondent's answer. Self-administered questionnaires, in contrast, must be delivered to the respondents in some way. One option is to program an online questionnaire with the questions to be tested. At the beginning of the cognitive interview, the participant is sent the link to the questionnaire and is asked to fill it out during the interview. The benefit of this procedure is that the questions are presented in real-time. On the downside, respondents must have access to a device with an Internet connection, which may not be the case for all potential participants. If the interview takes place via video conference, screen sharing presents itself as a second option. In this case, the interviewer shares the questionnaire with the participant during the interview. The advantage of this procedure is that the interviewer can control that the respondent is presented the right question at the right time. The disadvantage is that this method only works as long as the video connection is of high quality. The third option is to use a paper questionnaire, circumventing potential problems due to the respondent's technical equipment. On the downside, researchers must mail the questionnaire to respondents in advance, at best several days before the interview, which carries the risk that respondents read the questions before the interview or even lose the questionnaire.

In sum, the channel of communicating with participants and the way the questionnaire is presented in a remote cognitive interview can be mixed and matched in many ways. Even though remote cognitive interviewing poses some unique challenges in comparison to face-to-face interviewing, it also has some unique advantages. For example, interviewing respondents from and at home means less effort in terms of time and mobility on the side of interviewers and participants, and generally more flexible interview times. Moreover, remote cognitive interviews can be carried out with geographically dispersed participants much more easily, without imposing additional costs for travel or accommodation.

### *3.2.2 Number of interviews and sampling*

There are different practices and recommendations with regard to the number of interviews that should be conducted in the context of a cognitive interviewing project. As a rule, between five and 30 interviews are carried out per pretest (round) because the most serious question problems can usually be identified on the basis of a relatively small number of interviews (Willis, 2005). On the other hand, Blair and Conrad (2011) demonstrated that conducting more cognitive interviews than are typically carried out increases the probability of uncovering further significant question problems. However, when one considers the large volume of verbal data produced in the context of cognitive interviews, and the fact that these data must be analyzed, conducting more than 30 interviews per pretest (round) appears quite impracticable. Because of such cost-benefit considerations, researchers tend to work with 10 to 20 participants per pretest (round). If sufficient resources are available, it is recommended that the questions that have been revised based on the findings of the cognitive interviews are tested in a further round. This iterative approach enables the effectiveness of the revisions to be evaluated (Lenzner et al., 2023; Willis, 2005).

The pretest participants should generally have the same characteristics as the later respondents in the main survey. For example, if the main survey is aimed exclusively at pensioners, then only pensioners should take part in the cognitive interviews. If, on the other hand, it is a survey of the general population, then there should be some variation in terms of participants' gender, age, highest level of education and any other study-relevant characteristic. Typically, a quota sample is selected for cognitive interviewing. It is not necessary to draw a random sample because the main objective of cognitive interviewing is to uncover problems with the questions rather than provide as precise and estimate as possible of the frequency with which these problems occur in the

population. It is advisable to list the relevant participant criteria in a sampling plan to ensure that enough people with the relevant characteristics are recruited (see Figure 2).

<b>Age</b>	<b>Education</b>	<b>Female / Non-binary</b>	<b>Male</b>	<b>Σ</b>
<b>18 – 45 years</b>	B-	2	1	<b>3</b>
	B+	1	2	<b>3</b>
<b>46 – 75 years</b>	B-	1	2	<b>3</b>
	B+	2	1	<b>3</b>
<b>Σ</b>		<b>6</b>	<b>6</b>	<b>12</b>

(B- = no university entrance qualification; B+ = university entrance qualification)

Figure 2. Example of a cognitive interviewing quota plan

*3.2.3 Use of interview protocols*

Regarding the use of interview protocols, current practices range from an (almost) completely unstructured to an (almost) completely standardized approach. In the former, only the aims of testing are pre-formulated in the interview protocol and the interviewers decide spontaneously during the interview how and what to ask. In the latter approach, the cognitive techniques to be used are scripted in advance in the interview protocol, and the interviewers are instructed to administer the protocol as in a normal standardized interview, without asking additional questions or engaging in conversation with the participants. We recommend a mixture of the standardized and the non-standardized approaches, in which the cognitive techniques for testing questions are defined in advance, but the interviewers are also given the opportunity to ask additional follow-up questions until they feel they have obtained all necessary information from the respondents. An interview protocol should contain the questions to be tested, the aims of testing, the specific cognitive techniques to be used, and space for the interviewer’s notes and comments. Examples of interview protocols used at GESIS can be found in the GESIS Pretest Database ([pretest.gesis.org](http://pretest.gesis.org)) under the “Downloads” tab in the project view.

*3.2.4 Other practical aspects to consider*

Cognitive interviews should be scheduled to last between 60 and a maximum of 90 minutes. If they take longer, the concentration and motivation of the participants (and the interviewers) may decrease significantly. Depending on the number of probing questions applied, around 20 to 25 questions, or items, can be tested in this time.

When cognitive interviews are conducted at an early stage of the questionnaire development process, it is not absolutely necessary to test the questions in the mode of the actual survey. Ultimately, the central aim is to gather information about respondents’ cognitive processes when answering the questions and not to test the usability of the questionnaire for the respondents or interviewers. However, since aspects such as the communication mode of the survey or the layout

of questions can also have an influence on respondents' cognitive processes, it is recommended to stay as close as possible to the later survey mode in cognitive interviews.

### 3.3 How are cognitive interview data analyzed?

Before analyzing cognitive interview data, the individual recordings (and, if applicable, the interviewers' notes) need to be transcribed into a data entry template (see Figure 3). Such a template should contain the following information for each of the tested questions: (1) the question itself, (2) the responses to the tested question, (3) participants' spontaneous utterances on the question, (4) participants' answers to the probing questions (or other cognitive techniques), and (5) remarks by the cognitive interviewer. Data entry templates are matrix-based, meaning that they contain a series of grids in which each row represents a single participant and each column represents an area of inquiry. When analyzing these data, researchers review the grids vertically and compare the data across participants.

Question 1: How would you rate your general health?					
Respondent ID	Survey response	Spontaneous reaction / remark	Category selection probe	Comprehension probe	Interviewer notes
TP01	1 - very good	-	"I thought about my fitness and whether I was sick, and that wasn't the case, so very good"	R considers aspects of her physical health, including fitness and diet and whether she was ill.	-
TP02	4 - poor	"I've been ill so often this past year!"	R includes an operation, illnesses and anxiety he experienced in the aftermath of the operation.	INT: So you also considered your mental health when answering the question? R: Yes, absolutely.	-
TP03	3 - medium	Respondent hesitated	R chooses the middle category as a mean value: "I would say, I'm generally in a good state, like, I don't have cancer, but my allergies are terrible this year."	R considers his general physical health in the past years, but also his current state.	R hesitated, but this was not the result of comprehension difficulties, but of him considering many aspects of his health
...	1 - very good 2 - good <b>3 - medium</b> 4 - poor 5 - very poor				
<b>Probe wordings</b>					
<b>Category selection probe:</b> Can you tell me why you chose [ANSWER] as an answer to this question?					
<b>Comprehension probe:</b> How do understand the term "health" in this question?					
<b>Summary:</b>					
Most respondents answered the question without hesitation.					
Their responses to the probes corresponded with their survey response.					
All respondents considered aspects of their physical health; some respondents (TP02, 05) also considered their mental health.					

Figure 3. Example of a cognitive interviewing data entry template

Cognitive interview data are analyzed on a question-by-question basis (Willis, 2015). The analysis is guided by the key research questions formulated a priori (e.g., What do participants understand by a given term? How sensitive do the respondents perceive the question?). The data are commonly analyzed using inductive, qualitative procedures such as the constant comparative method (CCM). The CCM is particularly suitable for explorative data analysis and hypothesis generation. It comprises three steps:

1. Open coding: The verbal data of the participants are openly coded by topic, and initial categories are created (for example, for different interpretations or answer strategies).
2. Axial coding: The initial categories are compared and refined, and the dimensions that distinguish them from each other are identified.
3. Selective coding: In the last step, a hypothesis or theory is formulated that describes the patterns found in the data, the phenomena that a survey question captures and/or the mechanisms by which problems occur.

A detailed description of how this method is applied when analyzing cognitive interview data can be found in Ridolfo and Schoua-Glusberg (2011).

### 3.4 How can cognitive interviews be enriched with other pretesting methods?

Cognitive interviews are very well suited for being combined with usability testing. Usability testing involves observing how respondents interact with self-administered survey instruments, especially web questionnaires. It provides insights into how respondents navigate through a questionnaire, whether they can use the survey software as intended, and whether the questions look and function similarly on different devices (e.g., personal computers, tablets and smartphones). When supplementing cognitive interviewing with usability testing, the interview sessions are divided into two consecutive parts: first, the respondent fills out the questionnaire on a computer device or on paper, while the interviewer observes their response behavior and takes notes on any conspicuous features. Optionally, respondents can be asked to think aloud during this exercise. Afterwards, the interviewer conducts a cognitive interview and inquires about any previously noted issues, in addition to using predefined cognitive techniques to evaluate individual survey questions.

Usability testing can utilize eye-tracking technology (i.e., the recording of participants' eye movements) to discover where respondents look, for how long and in what order (i.e., how they process survey questions). However, the use of eye-tracking technology in usability testing is more of an add-on than a must. It is usually sufficient if the interviewer can observe the respondent's interaction with the questionnaire. This can be done either by sharing the screen (e.g., in remote interview settings) or by using a video camera that captures the respondent's interaction with the questionnaire and displays it live on a separate computer screen. In both cases, the interviewer can observe the respondent's answering behavior from a distance, which is unobtrusive and guarantees a comfortable atmosphere for the participant. When testing questions on mobile devices (i.e., tablets or smartphones), special cameras can be mounted on the devices (see Figure 4).

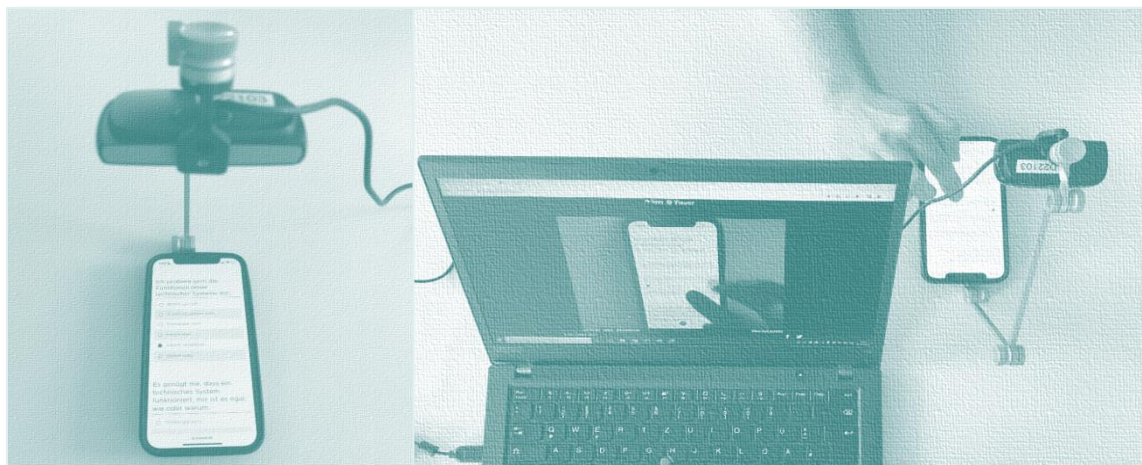


Figure 4. Example of hardware setup for mobile device testing

Supplementing cognitive interviewing with eye tracking is a valuable addition that goes beyond just usability testing. Respondents' reading patterns are indicative of difficulties they may have in answering questions, and so eye tracking can uncover question problems that are unconscious to the participants. This allows additional question problems to be identified that would remain undetected in a purely cognitive interview (Neuert & Lenzner, 2016). Moreover, eye tracking is a non-

reactive method which rules out interviewer effects. However, since eye-tracking data can only point to possible problems in answering questions but cannot identify the causes of these problems, eye-tracking sessions should always be accompanied by a cognitive interview, in which the interviewer addresses the observations made during the eye-tracking session and asks predefined probes on respondents' cognitive processes when answering the questions. Details on how to combine eye tracking and cognitive interviewing can be found in the Survey Guideline "[Use of Eye Tracking in Cognitive Pretests](#)" (Neuert & Lenzner, 2019).

#### **4. Planning and conducting a web probing pretest**

Web probing refers to the method of implementing cognitive probing questions in (self-administered) online questionnaires. Similar to cognitive interviews, respondents first answer a target question (or set of target questions) and afterwards receive one or more probing questions about their answering processes (usually on the next survey page; see section 4.2.2).

The main difference between cognitive interviewing and web probing is that no interviewer is present in the latter method. The self-administered nature of web probing has several advantages, but also some disadvantages (Behr et al., 2020; Lenzner & Neuert, 2017; Meitinger & Behr, 2016; Neuert et al., 2023; Hadler, 2023). On the positive side, interviewer effects are eliminated, social desirability bias is reduced, and there is no need to transcribe interview data. It also enables the time- and cost-efficient recruitment of large and geographically dispersed samples, which in turn allows the quantification of results and makes cross-cultural comparisons relatively easy (see GESIS Survey Guideline "[Web probing](#)", Behr et al., 2017). On the negative side, the absence of an interviewer means that the interactivity and flexibility of follow-up questions as well as the possibilities of motivating respondents to give detailed answers are limited. Furthermore, in contrast to closed questions, responding to open-ended questions imposes a greater burden on respondents. Compared to cognitive interviewing pretests, this often leads to lower response quality (e.g., higher rates of probe nonresponse and uninterpretable answers; Fowler & Willis, 2020; Lenzner & Neuert, 2017; Meitinger & Behr, 2016). In addition, since web surveys should generally be kept rather short (i.e., lasting around 10 to 15 minutes), only a reduced number of questions can usually be tested in web probing pretests.

Overall, web probing is particularly suitable as a cognitive pretesting method when (1) evaluating questions that will later be used in self-administered web surveys, (2) the test objectives are very clear and targeted probes can be anticipated (i.e., interactivity and flexibility in probing are not of primary importance), (3) a certain geographical spread of participants is required, and (4) statistical analyses of the pretest data are to be conducted in addition to the qualitative ones (e.g., analyses of internal consistency and item-total correlations of item batteries, analyses of response times; see section 4.3).

Studies comparing web probing and cognitive interviewing have shown that they produce similar results in terms of problems identified and item revisions suggested (Lenzner & Neuert, 2017; Meitinger & Behr, 2016). When choosing between the two methods, researchers are advised to weigh up the specific advantages and disadvantages mentioned above and to decide on a project-by-project basis which method is most likely to provide the desired insights. Depending on the research purpose, a (sequential or parallel) combination of cognitive interviewing and web probing may also be suitable (e.g., Hadler et al., 2022).

## 4.1 What techniques are (mainly) applied in a web probing pretest?

As the method's name suggests, web probing relies on implementing cognitive probing questions in online questionnaires. The most commonly used types of probes in web probing research studies to date are comprehension probes, category-selection probes and specific probes (see Table 1 above for a description of these probe types). However, in our web probing pretests at GESIS, we have also effectively used other probe types, such as process probes, difficulty probes, recall probes, and sensitivity probes (see [pretest.gesis.org](http://pretest.gesis.org)). In our experience, probes that work well in cognitive interviews are also suitable for web probing pretests.

With regard to probe wording, it is important to be as precise and specific as possible, particularly since there is no interviewer present to clarify the intention of the probe (i.e., the kind of information sought from respondents). It is crucial to use probes that encourage respondents to provide detailed answers, avoiding those that lead only to yes/no responses. Ideally, the probes should also communicate the expected depth and length of respondents' answers. For example, the probe in example (2) below is likely to produce richer responses when used to evaluate the survey question on respondents' net household income than the probe in example (1):

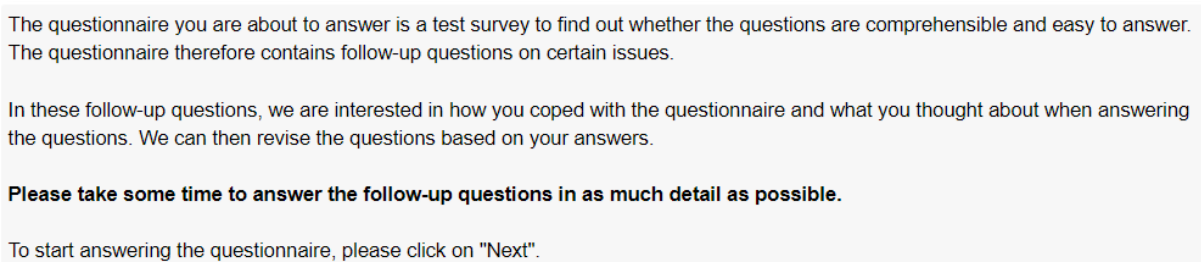
Question: Taking all incomes together: What was the net income of your household last month?

Example probe (1): How did you arrive at your answer?

Example probe (2): What was going through your mind when answering this question? Please explain which types of income from which persons you have taken into account in your answer.

## 4.2. How are web probing pretests conducted?

Similar to the briefing used in cognitive interviews, web probing pretests should begin with a note about the purpose of the study and an introduction of the probing procedure applied. In particular, respondents should be informed about the fact that the online questionnaire contains open-ended questions about some of the closed survey questions. It is also recommended to stress that detailed answers from respondents are crucial for the study's success. Figure 5 shows an exemplary introduction page of a web probing pretest.



The questionnaire you are about to answer is a test survey to find out whether the questions are comprehensible and easy to answer. The questionnaire therefore contains follow-up questions on certain issues.

In these follow-up questions, we are interested in how you coped with the questionnaire and what you thought about when answering the questions. We can then revise the questions based on your answers.

**Please take some time to answer the follow-up questions in as much detail as possible.**

To start answering the questionnaire, please click on "Next".

Figure 5. Example of an introductory web probing page

### 4.2.1 Probe layout

Answering open-ended questions in general and probes in particular is relatively strenuous for respondents, at least compared to answering closed survey questions. To reduce this burden and improve the quality of the answers to the probes, it is advisable to

- (1) preface probes with introductory sentences explaining that the probe relates to the previous survey question and that additional information is requested,

- (2) repeat the question to be tested,
- (3) repeat the respondent's answer to the question (if relevant, e.g., when asking a category-selection probe), and
- (4) in the case of numerical response scales, to repeat the end labels of the scale.

In this way, it is easier for respondents to remember their cognitive processes and the flow of the questionnaire is improved (cf. Behr et al., 2012). Figure 6 shows examples of this probe layout for category-selection probes about questions with fully labelled verbal answer scales and numerical answer scales.

We would like to receive more information on one of the previous statements.

The statement was: "I am concerned that my own standard of living will decline because of climate policy."

Your response was: **"Rather disagree"**

---

**Please explain your answer in more detail. Why did you choose this answer?**

We would like to get more information on the previous question.

The question was: "To what extent do you feel affected by infrastructure (e.g., highways, railroad tracks, power lines) where you live?"

Your answer on a scale from 0 (not at all affected) to 10 (very affected) was: "7".

---

**Please explain your answer in more detail. Why did you choose this answer?**

Figure 6. Examples of two layouts for category-selection probes

Research has shown that the size of the text box used in open-ended probes influences the depth, length, and format of the answers (Behr et al., 2014; Meitinger & Kunz, 2024). While larger text boxes convey the message to respondents that they are expected to provide longer, explanatory answers, smaller text boxes suggest that keyword-like examples or short definitions are desired. Consequently, category-selection probes asking for elaborations of survey answers, for example,

should be accompanied by a large, multi-line text box, while specific probes asking respondents to provide examples should be presented with one or more single-line answer boxes.

#### *4.2.2 Probe placement*

Concerning the placement of probing questions in the questionnaire, researchers must decide (1) when to ask which probes within the overall questionnaire and (2) in which order to ask probes that relate to the same survey question. As for the first decision, researchers should refrain from asking the same probes repeatedly (e.g., several category-selection probes with the same layout and text box size without any other probe type in between) as this may lead to habituation effects and to respondents skimming or not reading the probe text at all. This, in turn, can lead to respondents overlooking a sudden change in the type of probe and consequently giving answers that do not fit the probe (“mismatching” answers; Behr et al., 2014). Similar to cognitive interviews, web probes should ideally vary to some degree, but at the same time, only those types of probes should be asked that yield the desired insights.

With regard to the second decision, that is, when several probes are asked on the same survey question (e.g., a category-selection probe, a comprehension probe, and a specific probe), initial evidence suggests that a category-selection probe should be asked before all other probes, as this reduces the amount of probe nonresponse, for instance (Meitinger et al., 2018). Asking multiple probes on one and the same survey question also begs the question of whether the probes should be presented on the same survey page as the question to be tested (embedded design) or on one or more separate pages (scrolling and paging design). First results indicate that probe nonresponse is reduced in both the scrolling and paging design compared to the embedded design, but overall, the three designs seem to have only a small impact on the quality of probe answers (Neuert & Lenzner, 2023).

#### *4.2.3 Number of respondents and sampling*

With regard to the number of respondents interviewed in web probing pretests, sample sizes of 120 to 500 respondents are recommended. Larger samples make it possible to carry out statistical analyses in addition to the qualitative ones or to test different versions of questions (i.e., to randomly assign respondents to different question versions; see section 4.3 below). For such experimental designs, it is advisable to have at least 60 respondents per experimental group.

Respondents in web probing pretests are usually recruited from online access panels. These are non-representative groups of respondents who have agreed to take part in online surveys at regular intervals. Online access panels enable the recruitment of specific target groups and large samples in a comparatively cost- and time-efficient manner. Similar to cognitive interviews, web probing pretests typically use quota sampling to achieve a certain level of heterogeneity in the sample composition (see section 3.2.2 above).

#### *4.2.4 Other practical aspects to consider*

Following the general recommendation to keep the length of online surveys under 15 minutes (Revilla & Höhne, 2020), the web probing studies and pretests conducted at GESIS usually last between 5 and 15 minutes, with up to 8 or 9 probes being asked (Behr et al., 2017; Lenzner et al., 2022; Hadler et al., 2024). However, experimental findings indicate that web probing questionnaires can be even longer and contain more probes. Neuert and Lenzner (2023) compared a questionnaire with 26 items that contained 13 probes with a version that contained 21 probes and found that



increasing the number of probes did not negatively affect response quality across various indicators, such as probe nonresponse rates, number of uninterpretable answers, and respondents' satisfaction with the survey. Only the dropout rate was slightly higher in the condition with more probes, which suggests that a larger sample of respondents should be drawn when asking a higher number of probes.

As mentioned above, web probing studies often suffer from high probe nonresponse rates. To mitigate this problem, researchers should consider using automated prompts available in common online survey software solutions that display a (motivational) message whenever respondents leave a text box blank. These should be realized as so-called "soft prompts", which do not force respondents to type in an answer but allow them to skip a question and move on to the next survey page (e.g., by clicking on a corresponding answer box; see Figure 7). Alternatively, researchers may want to consider integrating open-access tools such as EvalAnswer (Kaczmirek et al., 2017) into their online surveys, which automatically detect different forms of probe nonresponse and immediately present respondents with tailored motivational statement aimed at persuading them to put more effort into answering the probes (e.g., "You seem to be in a hurry! Please take another moment to answer the question in as much detail as possible.").

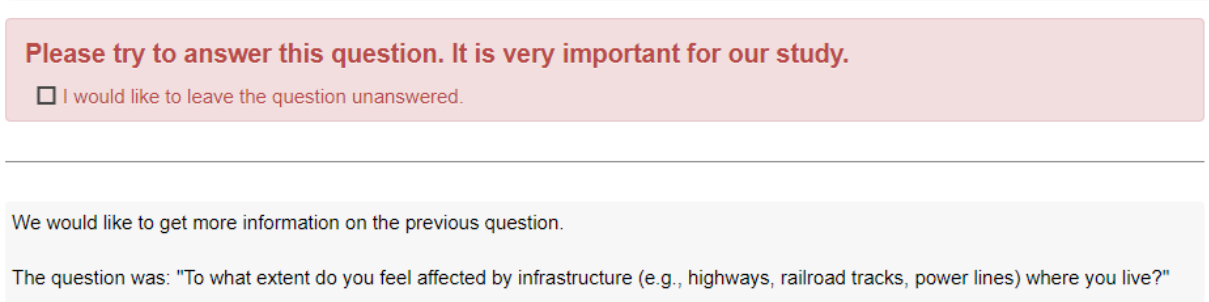


Figure 7. Example of a soft prompt reminding respondents to answer a probing question

#### 4.3 How are web probing data analyzed?

When analyzing web probing data, the same methods are used as when analyzing data from cognitive interviews (see section 3.3 above). The data exported from online survey software is usually already in a matrix format, so the only data pre-processing step researchers need to perform is to insert blank columns into the dataset into which the codes for the themes in the probe answers can later be inserted during the qualitative analysis.

In addition to conducting qualitative analyses of the answers to open-ended probes, the larger sample sizes in web probing pretests enable researchers to carry out statistical analyses of the tested survey questions as well. These include, for example, looking at frequency distributions of the responses to these questions (including analyses of item nonresponse) and determining the internal consistency and item-total correlations in item batteries. Whenever different question versions are tested, it also makes sense to examine differences in the response times for these questions and the answer distributions (for example, by means of Chi<sup>2</sup>-tests or independent-samples t-tests), because longer response times can be indicative of cognitively or demanding questions (Lenzner et al., 2010). Examples of such statistical analyses can be found in the web probing pretest reports that have been published since the year 2022 in the GESIS Pretest Database ([pretest.gesis.org](http://pretest.gesis.org); e.g., [Hadler et al., 2024](#); [Schick et al., 2022](#), [2023](#)).

#### 4.4 How can web probing be enriched with other pretesting methods within the same data collection?

As mentioned above, the larger sample sizes used in web probing pretests make it possible to include randomized experiments in the online questionnaire. In this way, two or more different versions of a question (differing, for example, in the wording of the question or the response scale) can be tested simultaneously and be compared. Such experiments can be helpful in determining what differences it makes, if any, to phrase or order questions in alternative ways. Based on the results, researchers can make an informed decision about which version to use in their later survey. For example, if one of two question version receives significantly fewer “don’t know” responses, it indicates that this question version is clearer to respondents and should thus be favored over the other version.

The fact that web probing relies on online questionnaires means that a considerable amount of paradata can easily be gathered as a by-product of the data collection. For example, these include information on the device type used by respondents, response times, keystrokes, and mouse movements (see GESIS Survey Guideline “[Web Paradata in Survey Research](#)”, Kunz & Hadler, 2020). All of these variables can provide important information on how respondents process and answer survey questions (e.g., whether they click on icons for additional information) and on difficulties they may have in answering questions (e.g., whether the questions are displayed correctly on devices with different screen sizes). As mentioned above, in experimental designs, response times can also be indicative of the cognitive burden associated with different questions versions (Lenzner et al., 2010).

Finally, web probing can easily be integrated into the pilot testing of online questionnaires and function as a respondent debriefing (see section 2.1 above). In this scenario, probing questions are implemented at the end of the questionnaire and respondents are probed retrospectively about a selection of the questions they previously answered. By placing the probes at the end of the questionnaire, the natural situation of completing the survey is not interrupted by the probes (Hadler, 2023), and both information on the overall length of the questionnaire as well as information on the understanding or comprehensibility of selected questions can be collected.

## 5. References

- Beatty, P. C., Collins, D., Kaye, L., Padilla, J.-L., Willis, G. B., & Wilmot, A. (Eds.) (2019). *Advances in questionnaire design, development, evaluation and testing*. Hoboken, NJ: John Wiley & Sons.
- Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, 71(2), 287-311. <https://doi.org/10.1093/poq/nfm006>
- Behr, D., Braun, M., Kaczmirek, L., & Bandilla, W. (2014). Item comparability in cross-national surveys: Results from asking probing questions in cross-national web surveys about attitudes towards civil disobedience. *Quality & Quantity*, 48(1), 127-148. <https://doi.org/10.1007/s11135-012-9754-8>
- Behr, D., Kaczmirek, L., Bandilla, W., & Braun, M. (2012). Asking probing questions in web surveys: Which factors have an impact on the quality of responses? *Social Science Computer Review*, 30(4), 487-498. <https://doi.org/10.1177/0894439311435305>
- Behr, D., Meitinger, K., Braun, M., & Kaczmirek, L. (2017). *Web probing - implementing probing techniques from cognitive interviewing in web surveys with the goal to assess the validity of survey questions*. Mannheim, GESIS - Leibniz Institute for the Social Sciences (GESIS - Survey Guidelines). [https://doi.org/10.15465/gesis-sg\\_en\\_023](https://doi.org/10.15465/gesis-sg_en_023)
- Behr, D., Meitinger, K., Braun, M., & Kaczmirek, L. (2020). Cross-national web probing: An overview of its methodology and its use in cross-national studies. In P. C. Beatty, D. Collins, L. Kaye, J.-L. Padilla, G. B. Willis, & A. Wilmot (Eds.), *Advances in questionnaire design, development, evaluation and testing* (pp. 521-543). Hoboken, NJ: John Wiley & Sons. <https://doi.org/10.1002/9781119263685.ch21>
- Blair, J., & Conrad, F. G. (2011). Sample size for cognitive interview pretesting. *Public opinion quarterly*, 75(4), 636-658. <https://doi.org/10.1093/poq/nfr035>
- Campanelli, P. (2008). Testing survey questions. In E. de Leeuw, J. Hox, & D. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 176-200). New York: Erlbaum.
- Collins, D. (2015). *Cognitive interviewing practice*. Thousand Oaks, CA: Sage.
- Conrad, F. G., & Blair, J. (2009). Sources of error in cognitive interviews. *Public Opinion Quarterly*, 73(1), 32-55. <https://doi.org/10.1093/poq/nfp013>
- Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. Beverly Hills, CA: Sage.
- Fowler, S., & Willis, G. B. (2020). The practice of cognitive interviewing through web probing. In P. Beatty, D. Collins, L. Kaye, J. L. Padilla, G. Willis, & A. Wilmot (Eds.), *Advances in questionnaire design, development, evaluation and testing* (pp. 451-469). Hoboken, NJ: John Wiley & Sons. <https://doi.org/10.1002/9781119263685.ch18>
- Hadler, P. (2023). *Context effects in question evaluation via web probing: Exploring the interaction of open-ended and closed survey questions*. Mannheim: MADOC. <https://madoc.bib.uni-mannheim.de/66207/>
- Hadler, P., Lenzner, T., Finzer, M., Scholl, J., & Neuert, C. (2024). FReDA-W4 – Fragen zu den Themen Haushaltsgröße, Einkommen, idealer Erwerbsumfang von Eltern, Arbeitszeit und regionale Daseinsvorsorge. Kognitiver Online-Pretest. *GESIS Projektbericht*. Version: 1.0. <http://doi.org/10.17173/pretest139>
- Hadler, P., Lenzner, T., Schick, L., Steins, P., & Neuert, C. (2022). European Working Conditions Survey 2024. Cognitive Pretest. *GESIS Project Reports*. Version: 1.0. <http://doi.org/10.17173/pretest116>

- Kaczmirek, L., Meitinger, K., & Behr, D. (2017). *Higher data quality in web probing with EvalAnswer: a tool for identifying and reducing nonresponse in openended questions*. (GESIS Papers, 2017/01). Köln: GESIS – Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.21241/soar.51100>
- Kunz, T., & Hadler, P. (2020). *Web Paradata in Survey Research*. Mannheim, GESIS - Leibniz Institute for the Social Sciences (GESIS - Survey Guidelines). [https://doi.org/10.15465/gesis-sg\\_037](https://doi.org/10.15465/gesis-sg_037)
- Lenzner, T., Hadler, P., & Neuert, C. (2021). *How to Conduct Cognitive Interviews in Times of COVID-19*. Mannheim, GESIS - Leibniz Institute for the Social Sciences (GESIS Blog). <https://doi.org/10.34879/gesisblog.2021.34>
- Lenzner, T., Hadler, P., & Neuert, C. (2023). An experimental test of the effectiveness of cognitive interviewing in pretesting questionnaires. *Quality & Quantity*, 57(4), 3199-3217. <https://doi.org/10.1007/s11135-022-01489-4>
- Lenzner, T., Kaczmirek, L., & Lenzner, A. (2010). Cognitive burden of survey questions and response times: a psycholinguistic experiment. *Applied Cognitive Psychology*, 24(7), 1003-1020. <https://doi.org/10.1002/acp.160>
- Lenzner, T., & Neuert, C. E. (2017). Pretesting survey questions via web probing – Does it produce similar results to face-to-face cognitive interviewing? *Survey Practice*, 10(4) <https://doi.org/10.29115/SP-2017-0020>
- Lenzner, T., Schick, L., Hadler, P., Behnert, J., Steins, P. & Neuert, C. (2022). FGZ Cohesion Panel: Wave 2 – Questions on climate change, antisemitism, and gender equality (English Version). Cognitive Online Pretest. *GESIS Project Reports*. <http://doi.org/10.17173/pretest129>
- Meitinger, K., & Behr, D. (2016). Comparing cognitive interviewing and online probing: Do they find similar results? *Field Methods*, 28(4), 363–380. <https://doi.org/10.1177/1525822X15625866>
- Meitinger, K., Braun, M., & Behr, D. (2018). Sequence matters in web probing: The impact of the order of probes on response quality, motivation of respondents, and answer content. *Survey Research Methods*, 12(2), 103–120. <https://doi.org/10.18148/srm/2018.v12i2.7219>
- Meitinger, K., & Kunz, T. (2024). Visual design and cognition in list-style open-ended questions in web probing. *Sociological Methods & Research*, 53(2). <https://doi.org/10.1177/00491241221077241>
- Neuert, C. E., & Lenzner, T. (2016). A Comparison of Two Cognitive Pretesting Techniques Supported by Eye Tracking. *Social Science Computer Review*, 34(5), 582-596. <https://doi.org/10.1177/0894439315596157>
- Neuert, C. E., & Lenzner, T. (2019). *Use of Eye Tracking in Cognitive Pretests*. GESIS - Leibniz Institute for the Social Sciences (GESIS - Survey Guidelines). [https://doi.org/10.15465/gesis-sg\\_en\\_025](https://doi.org/10.15465/gesis-sg_en_025)
- Neuert, C., & Lenzner, T. (2023). Design of multiple open-ended probes in cognitive online pretests using web probing. *Survey Methods: Insights from the Field (SMIF)*. <https://doi.org/10.13094/SMIF-2023-00005>
- Neuert, C. E., Meitinger, K., & Behr, D. (2023). Open-ended versus closed probes: Assessing different formats of web probing. *Sociological Methods & Research*, 52(4), 1981-2015. <https://doi.org/10.1177/00491241211031271>
- Porst, R. (2014). *Fragebogen: Ein Arbeitsbuch*. Wiesbaden: Springer.
- Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly*, 68(1), 109–130. <https://doi.org/10.1093/poq/nfh008>
- Prüfer, P., & Rexroth, M. (1996). Verfahren zur Evaluation von Survey-Fragen: Ein Überblick. *ZUMA-Nachrichten*, 39, 95-116.

- Prüfer, P., & Rexroth, M. (2005). Kognitive Interviews. *ZUMA How-to-Reihe, Nr. 15*.
- Repke, L., Birkenmaier, L., & Lechner, C. M. (2024). *Validity in Survey Research – From Research Design to Measurement Instruments*. Mannheim, GESIS - Leibniz Institute for the Social Sciences (GESIS - Survey Guidelines). [https://doi.org/10.15465/gesis-sg\\_en\\_048](https://doi.org/10.15465/gesis-sg_en_048)
- Revilla, M., & Höhne, J. K. (2020). How long do respondents think online surveys should be? New evidence from two online panels in Germany. *International Journal of Market Research*, 62(5), 538-545. <https://doi.org/10.1177/1470785320943049>
- Ridolfo, H., & Schoua-Glusberg, A. (2011). Analyzing cognitive interview data using the constant comparative method of analysis to understand cross-cultural patterns in survey data. *Field Methods*, 23(4), 420-438. <https://doi.org/10.1177/1525822X11414835>
- Schick, L., Lenzner, T., Hadler, P., Behnert, J., & Neuert, C. (2022). DEval-Meinungsmonitor Entwicklungspolitik. Kognitiver Online-Pretest. *GESIS Projektbericht*. Version: 1.0. <http://doi.org/10.17173/pretest122>
- Schick, L., Lenzner, T., Hadler, P., & Neuert, C. (2023). FReDA-W3b – Fragen zu den Themen Partnerschaftsstatus, Ernährungsstile, globale Unsicherheit und Vertrauen in Institutionen. Kognitiver Online-Pretest. *GESIS Projektbericht*. Version: 1.0. <http://doi.org/10.17173/pretest127>
- Schmidt, I., & Lechner, C. M. (2020). *Documenting Measurement Instruments for the Social and Behavioral Sciences*. Mannheim, GESIS - Leibniz Institute for the Social Sciences (GESIS - Survey Guidelines). [https://doi.org/10.15465/gesis-sg\\_en\\_033](https://doi.org/10.15465/gesis-sg_en_033)
- Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design*. San Francisco, CA: Jossey-Bass.
- Walter, J. G. (2018). The adequacy of measures of gender roles attitudes: A review of current measures in omnibus surveys. *Quality & Quantity*, 52(2), 829–848. <https://doi.org/10.1007/s11135-017-0491-x>
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.
- Willis, G. B. (2015). *Analysis of the cognitive interview in questionnaire design*. Oxford University Press.