Pretesting Survey Questions Via Web Probing – Does it Produce Similar Results to Face-to-Face Cognitive Interviewing?

Timo Lenzner[*], Cornelia E Neuert[†]

Survey Practice

Asking probing questions via web probing has recently been advocated as a promising method for evaluating survey questions. In comparison to standard face-to-face (f2f) cognitive interviewing, the increasing availability of internet non-probability panels allows for recruiting respondents in a quicker and more cost-effective way and a realization of larger sample sizes. In the present study, we examine whether web probing is a potential alternative to standard cognitive interviewing, in particular: Does web probing produce similar results as f2f cognitive interviewing with regard to the problems detected and the item revisions suggested? The study compares the findings of 508 respondents drawn from a non-probability online panel who completed an online survey including four items from the International Social Survey Programme 2013 and 2014 with the results obtained via f2f cognitive interviewing with 20 participants. Findings indicate that web probing and cognitive interviewing detect very similar problems and lead to the same suggestions for item revisions. However, web probing itself has some limitations. Practical implementations and directions for future research are discussed.

INTRODUCTION AND RESEARCH QUESTIONS

Cognitive interviewing is a qualitative method that aims to reveal information from respondents about the cognitive processes they use when answering survey questions and to identify problems with questions (Willis 2005). Conventionally, cognitive interviewing involves conducting face-to-face (f2f) interviews with small sample sizes of five to 30 respondents (Willis 2005). The semi-structured, in-depth interviews are conducted by specially trained cognitive interviewers on the basis of an interview protocol which contains the questions to be tested in the cognitive interview and the techniques to be adopted, in particular think-aloud and follow-up questions (probing). The technique of probing is used to elicit information about how respondents interpret questions or define specific terms and how respondents arrive at their answers. In addition to the scripted probing questions included in the interview protocol, emergent probes can be asked to follow up on respondents' comments during the interview. Probing questions are administered either

* **Institution:** GESIS - Leibniz Institute for the Social Sciences
† **Institution:** GESIS - Leibniz Institute for the Social Sciences

immediately after the subject has answered the survey question (concurrent) or at the end of the cognitive interview (retrospective; Willis 2005).

An alternative to conducting f2f cognitive interviews in the lab is to transfer the probing procedure into an online questionnaire, a method called online or web probing. Here, for the questions to be tested, open and closed probing questions are developed and then implemented into an online questionnaire. In the concurrent probing format, respondents first answer a survey question and after clicking on the next button receive one or more probes on the next survey page. As web probing does not involve a cognitive interviewer, respondents have to answer the probing questions in a self-administered form. The method of web probing has recently been recognized as a promising tool for evaluating survey questions, both during the post-survey assessment of item validity (Behr et al. 2012, 2013) and as a pretesting method to collect data about response strategies (Edgar 2012).

In comparison to cognitive interviewing, web probing has several benefits: First, it allows for recruiting respondents in a quicker and more cost-effective way and thereby a realization of larger sample sizes. This, in turn, allows researchers to quantify their pretest findings (Behr et al. 2012). Second, recruiting participants via the Internet enhances the radius of the regional accessibility. Furthermore, the self-administered mode rules out any interviewer effects and thus increases the reliability and comparability of the results (Conrad and Blair 2009). However, due to the absence of the interviewer, no one can probe for more information, follow up on incomplete answers or provide clarification of the tasks. Probing is restricted to the scripted questions previously programmed and implemented into the Web survey. Moreover, no one can motivate the respondents during completion of the Web survey to answer the (open) probing questions thoughtfully and elaborately. This can result in more satisficing response behavior of the respondents (K rosnick 1991) who then do not provide the same depth of information as participants in a f2f cognitive interview (Meitinger and Behr 2016). Nevertheless, Behr and colleagues have shown that web respondents give meaningful answers to open-ended probing questions (Behr et al. 2012), and Meitinger and Behr (2016) found that there is an extensive overlap between the results of both methods with respect to identified error types and uncovered themes although cognitive interview respondents provided, on average, more indications of errors than web probing respondents.

In the present study, we replicate the earlier research of Meitinger and Behr (2016) by examining whether web probing produces similar results to f2f cognitive interviewing with regard to the problems detected. In addition, we extend Meitinger and Behr's (2016) research by examining whether both methods produce similar results concerning the item revisions suggested.

METHODS

To examine these research questions, we embedded four items from the International Social Survey Programme (ISSP) 2013 and 2014 into a larger online questionnaire fielded in May 2014. The questionnaire included several methodological studies (of which only the present study applied web probing) and respondents required approximately 25 minutes to complete it. The four items examined in the present study had been tested previously via f2f cognitive interviewing in the GESIS pretest lab in August/September 2013 so that results on the performance of these four items were already available.[1]

The Web survey respondents were drawn from a respondent pool that was assembled during the set-up of the GESIS Online Panel Pilot, however, which is not representative of the German population. Of the 897 respondents who were invited, 534 participated in the survey and 508 completed it, resulting in a response rate of 59.3 percent (American Association for Public Opinion Research RR1). The 20 respondents participating in the f2f cognitive interviews were recruited from a respondent pool maintained by the GESIS pretest lab using quotas for age, education, and gender. The participants received a compensation of 5€ for completing the 25-minute web questionnaire and a compensation of 30€ for taking part in the 60 min f2f cognitive interviews, respectively. Table 1 shows some demographic characteristics of both respondent groups. While the composition of both respondent groups was quite similar with regard to sex and age, they differed somewhat with regard to educational attainment: on average, participants in the web survey had received a higher education than participants in the f2f cognitive interview.

---

1 The items in the online questionnaire were asked as part of an experiment that varied the number of probing questions asked (ranging from 4 to 7 probing questions), the number of nonresponse probes asked (also ranging from 4 to 7), and the number of probing questions presented per page (ranging from 1 to 2 questions per page). The results of this experiment will be presented elsewhere. In this paper, we restrict ourselves to the qualitative analysis of the respondents' answers to the probing questions and the comparison of these results to the findings of the f2f cognitive interviews. Initial analyses comparing the results from the three experimental groups revealed no differences relevant to our research questions, so we combined data from the three sources in the analyses.

Table 1 Demographic characteristics of participants.

	Web survey	F2F cognitive interview
Sex		
Female	227 (45%)	11 (55%)
Male	281 (55%)	9 (45%)
Age		
18–40	187 (37%)	9 (45%)
41 and older	321 (63%)	11 (55%)
Education		
Less than college	178 (35%)	11 (55%)
College and higher	330 (65%)	9 (45%)
N	508	20

The four items to be tested were taken from the modules National Identity and Citizenship of the German questionnaires of the ISSP 2013 and 2014. (See Table 2 in the Results section for the English wording of these items and the Appendix for the original German version.) The items were evaluated by the same probing techniques in both methods, that is by comprehension probes ("What does the term X mean to you?"), elaborative probes ("Could you please explain your answer a little further?"), and specific probes ("What kinds of elections did you think of when answering this question?"). However, in the f2f setting, the interviewers were also encouraged to apply additional probing questions if they deemed it necessary and respondents often commented spontaneously on the items prior to the administration of any probe. Hence, the verbal data obtained by the f2f interviews are based on more information sources than the data obtained by the Web survey. In both groups, the probing questions were administered immediately after respondents answered the target questions (concurrent probing).

Table 2   Identified problems, prevalence of problems, and suggested revisions to items.

Tested item	F2f problems (prevalence across interviews in %)	F2f revision	Web problems (prevalence across interviews in %)	Web revision	Are revisions the same for the two methods?
I1. How important is it that citizens may engage in acts of civil disobedience when they strictly oppose government actions? Response scale ranging from 1 to 7, where 1 is not at all important and 7 is very important	– The term 'civil disobedience' is unfamiliar/undefined (30%) – 'Civil disobedience' is associated with violent behavior (15%) – The response scale is interpreted as reaching from nonviolent to violent behavior (5%)	How important is it that citizens may engage in acts of nonviolent protest when they strictly oppose government actions?	– The term 'civil disobedience' is unfamiliar/undefined (5%) – 'Civil disobedience' is associated with violent behavior (12%) – The response scale is interpreted as reaching from non-violent to violent behavior (2%)	How important is it that citizens may engage in acts of nonviolent protest when they strictly oppose government actions?	Yes
I2. How important is it that long-term residents of a country, who are not citizens, have the right to vote in that country's national elections? Response scale ranging from 1 to 7, where 1 is not at all important and 7 is very important	– Respondents think of all sorts of elections when answering the question and not only about elections on the national level (55%) – Respondents do NOT think about national elections when answering the question and would answer differently if they did so (15%) – It is unclear what time period the term 'long-term' refers to (5%) – One respondent says he would rather answer whether he is	How important is it that long-term residents of a country, who are not citizens, have the right to vote in that country's nationwide elections?	– Respondents think of all sorts of elections when answering the question and not only about elections on the national level (53%) – Respondents do NOT think about national elections when answering the question and would answer differently if they did so (23%) – It is unclear what time period the term 'long-term' refers to (1%)	How important is it that long-term residents of a country, who are not citizens, have the right to vote in that country's nationwide elections?	Yes

Tested item	F2f problems (prevalence across interviews in %)	F2f revision	Web problems (prevalence across interviews in %)	Web revision	Are revisions the same for the two methods?
	in favor of this issue or not than rating its importance*				
I3. How much do you agree or disagree with the following statement: The world would be a better place if Germans acknowledged Germany's shortcomings Answer categories: Agree Strongly, Agree, Neither Agree nor Disagree, Disagree, Disagree Strongly, Don't Know	– Respondents do not understand what the question is actually about (20%) – Respondents do not see a connection between the acknowledgment of Germany's shortcomings and the state of the world (45%) – Respondents ignore the causal relation stated in the question and answer it only with regard to Germany's shortcomings (10%) – Respondents disagree with the presupposition(s) that (a) Germans do not acknowledge this or (b) there are shortcomings in Germany (20%)	How much do you agree or disagree with the following statement: The world would be a better one if Germany admitted to other countries that over here there are shortcomings too	– Respondents do not understand what the question is actually about (10%) – Respondents do not see a connection between the acknowledgment of Germany's shortcomings and the state of the world (27%) – Respondents ignore the causal relation stated in the question and answer it only with regard to Germany's shortcomings (21%) – Respondents disagree with the presupposition(s) that (a) Germans do not acknowledge this or (b) there are shortcomings in Germany (8%)	How much do you agree or disagree with the following statement: The world would be a better one if Germany admitted to other countries that over here there are shortcomings too	Yes
I4. How much do you agree or disagree with the following statement: I feel more like a citizen of the world, and thus connected to the world as a whole, and less as a citizen of a particular country	The term 'citizen of the world' is understood incorrectly: 1. Due to globalization (import/ export of goods), we are connected to the whole world (10%) 2. Citizen of the	How much do you agree or disagree with the following statement: I feel more connected to the world as a whole than to a particular country	– The term 'citizen of the world' is understood incorrectly: 1. Due to globalization (import/export of goods), we are	How much do you agree or disagree with the following statement: I feel more connected to the world as a whole than to a particular country	Yes

Tested item	F2f problems (prevalence across interviews in %)	F2f revision	Web problems (prevalence across interviews in %)	Web revision	Are revisions the same for the two methods?
Answer categories: Agree Strongly, Agree, Neither Agree nor Disagree, Disagree, Disagree Strongly, Don't Know	world=inhabitant of the world (15%) 3. Due to modern media (Internet) we are connected to the whole world (10%) 4. Due to migration we live in a multi-cultural society (5%)*		connected to the whole world (8%) 2. Citizen of the world=inhabitant of the world (13%) 3. Due to modern media (Internet) we are connected to the whole world (6%) – The term 'citizen of the world' is unfamiliar (2%)*		

*Asterisk indicates that the problem was detected by one but not the other pretest method. Revisions are translations from the original German revisions. The original German item wordings and revisions are listed in the Appendix.

Before analyzing the respondents' answers to the probing questions, the f2f interview data were transcribed from the video recordings of the interviews. Afterward, the data in both groups were analyzed by two researchers, working independently and each one reviewing both data sets, as follows: first, they openly coded respondents' answers to the probes with regard to the kinds of information they provided. Second, they organized these codes into larger categories and specified the core themes and types of problems that emerged from the analysis. Finally, they developed draft revisions for the questions. The researchers then met to discuss the findings, to resolve minor discrepancies in the codings, and to make a final decision about the recommendations for revision.

RESULTS

The results of our analyses are displayed in Table 2. All in all, the f2f cognitive interviews and the web probing method identified very similar question problems and led to identical suggestions for revising the items. Differences in the types of problems detected were only found for item 2 and item 4. In item 2, one respondent in the f2f setting said that he would rather answer whether he is in favor or not of the issue in question (i.e., whether long-term residents of a country, who are not citizens, have the right to vote in national elections) than rating how important he finds the issue. This problem was not found in the web probing data. In item 4, f2f cognitive interviewing revealed that some respondents misinterpreted the term "citizen of the world" as referring to people living in a multicultural society (e.g., ID 12: "Nowadays, people from all over the world are living here, and we have got so used to it that you could indeed say one feels rather connected to the whole world."). Again, this interpretation was not found in the web probing data. By contrast, web probing revealed that the term "citizen of the world" was unfamiliar to some respondents who, as a consequence, were not able to answer the question meaningfully (e.g., ID 178: "What is a 'citizen of the world' supposed to be?"). Despite these minor differences, both methods resulted in the same recommendations for revising item 2 and item 4, namely in replacing the term "national elections" with "nationwide elections" (item 2) and deleting the unclear term "citizen of the world" (item 4).

With regard to the prevalence of the problems detected, we found some substantial differences between the two methods. For example, while 30 percent of the f2f cognitive interview respondents said that the term "civil disobedience" in item 1 was unfamiliar to them, only 5 percent of the web respondents did so. This might be due to the fact that participants in the f2f setting often spontaneously commented on an item before answering one of the probing questions. Hence, some of these participants first said that they were unsure about the meaning of the term and afterward (in response to the probing question) explained what they thought the term most likely referred to. In the web probing setting, respondents had no means to comment on an item spontaneously and were thus more focused on answering the probing

questions. Again, however, the differences in the prevalence of problems had no effects on the suggested item revisions. Irrespective of their prevalence, the same problems were either deemed significant or insignificant for rendering item revisions necessary in both methods.

Finally, we examined whether the problems detected had any effects on measurement quality, in particular, whether respondents misinterpreting an item or having any other difficulty answering it systematically erred in one direction when responding to the item. This response behavior was found in three of the four items (I1, I2, I4). In item 1, respondents who associated the term "civil disobedience" with violent behavior were more likely to rate the item as not important than respondents who (correctly) interpreted the term as referring to nonviolent behavior. In item 2, we found that respondents who were primarily thinking of local elections when answering the item valued long-term residents' right to vote more important than if they thought of national elections. And finally, in item 4, respondents misinterpreting the term "citizen of the world" were more likely to agree that they "feel like a citizen of the world" than to disagree with this statement. Hence, the proportion of respondents who really hold cosmopolitan views might be overestimated when using this item. In sum, the problems detected by both methods were indeed severe enough for rendering revisions necessary.

As a by-product of the analyses presented above, we additionally found some substantial differences between both methods regarding item nonresponse and the proportion of meaningful and interpretable answers respondents provided to the probes. While nearly all f2f respondents provided interpretable answers to the probing questions asked, many web respondents did not answer the probing questions meaningfully, but simply skipped these questions, provided unintelligible or very short answers or copied definitions from the Web. On average, this behavior occurred in 14 percent of the cases.

DISCUSSION

In this study, we examined whether traditional f2f cognitive interviewing and web probing yield similar results in pretesting survey questions. Our findings indicate that both methods detect very similar problems and lead to the same suggestions for item revisions. Hence, web probing appears to be a promising method for pretesting questionnaires, and our findings suggest that it may be used as an alternative to standard cognitive interviewing.

On the positive side, web probing additionally allows researchers to quantify their pretest findings and to estimate the measurement error associated with the problems detected if large sample sizes are used. In addition, almost no staff resources are needed for recruiting participants and conducting interviews, and incentives are generally lower in online surveys than in f2f interviews. On the negative side, we found that a considerable amount of the web respondents did not provide meaningful answers to the probing questions, and thus, it seems

important that practitioners recruit larger sample sizes than necessary when conducting a web probing pretest to obtain a suitable amount of interpretable responses.

There are several limitations to this study calling for future research. First, it is important to note that we applied only one of several existing cognitive interviewing techniques (i.e., verbal probing) in both pretesting methods in this study. Thus, our findings are restricted to this particular technique and do not generalize to other techniques commonly used in f2f cognitive interviews, such as thinking aloud, for example. Given that it is technically possible to do an audio and screen recording of the web respondents' answering process, future studies should look into whether web respondents can be motivated to perform think-aloud tasks while answering the online questionnaire, and if so, whether the web and f2f settings again yield similar pretesting results. Second, it seems worthwhile to examine whether additional behavioral data, such as keystrokes, response times, and mouse movements, which can be collected easily in Web surveys, could provide further insights on response difficulties. Finally, our study focused exclusively on attitudinal questions and did not examine the performance of both methods in testing factual and behavioral questions. Hence, future research should ideally include a broader set of question types.

Given that the use of web probing as a pretesting method is still in its infancy, there are several other issues worth to be addressed in future studies. For example, future research should investigate the potential merit of implementing nonresponse probes into the online questionnaires, that is, motivating probes (e.g., "Please answer this question. It is of great importance to this study.") automatically triggered by undesired respondent behavior (e.g., providing very short or no answers to probing questions). Moreover, it should be examined whether web respondents can be motivated to answer as many probing questions as f2f cognitive interview respondents, that is, to fill in a questionnaire for up to 60 minutes. And finally, future research should study the minimum sample size necessary to ensure a sufficiently high likelihood that a problem is being detected in a web probing pretest.

ACKNOWLEDGMENT

APPENDIX

ORIGINAL GERMAN VERSION AND ENGLISH TRANSLATIONS OF ITEMS AND SUGGESTED REVISIONS

I1: Wie wichtig ist es für Sie, dass Bürger die Möglichkeit des zivilen Ungehorsams haben, um ihre deutliche Ablehnung gegenüber Regierungsentscheidungen zum Ausdruck zu bringen?

[How important is it that citizens may engage in acts of civil disobedience when they strictly oppose government actions?]

1 – Überhaupt nicht wichtig, 2, 3, 4, 5, 6, 7 – Sehr wichtig, Kann ich nicht sagen.

[1 – Not at all important, 2, 3, 4, 5, 6, 7 – Very important, Don't know.]

Revision I1: Wie wichtig ist es für Sie, dass Bürger die Möglichkeit des gewaltlosen Protests haben, um ihre deutliche Ablehnung gegenüber Regierungsentscheidungen zum Ausdruck zu bringen?

[How important is it that citizens may engage in acts of nonviolent protest when they strictly oppose government actions?]

I2: Wie wichtig ist es für Sie, dass Menschen, die schon lange in einem Land leben, aber dort nicht eingebürgert sind, das Recht haben, bei nationalen Wahlen abzustimmen?

[How important is it that long-term residents of a country, who are not citizens, have the right to vote in that country's national elections?]

1 – Überhaupt nicht wichtig, 2, 3, 4, 5, 6, 7 – Sehr wichtig, Kann ich nicht sagen.

[1 – Not at all important, 2, 3, 4, 5, 6, 7 – Very important, Don't know.]

Revision I2: Wie wichtig ist es für Sie, dass Menschen, die schon lange in einem Land leben, aber dort nicht eingebürgert sind, das Recht haben, bei landesweiten Wahlen abzustimmen?

[How important is it that long-term residents of a country, who are not citizens, have the right to vote in that country's nationwide elections?]

I3: Inwieweit stimmen Sie den folgenden Aussagen zu oder nicht zu? Die Welt wäre besser, wenn die Deutschen zugeben würden, dass in Deutschland nicht alles zum Besten steht.

[How much do you agree or disagree with the following statement: The world would be a better place if Germans acknowledged Germany's shortcomings.]

Stimme voll und ganz zu, Stimme zu, Weder noch, Stimme nicht zu, Stimme überhaupt nicht zu, Kann ich nicht sagen.

[Agree strongly, Agree, Neither agree nor disagree, Disagree, Disagree strongly, Don't know.]

Revision I3: Inwieweit stimmen Sie den folgenden Aussagen zu oder nicht zu? Die Welt wäre eine bessere, wenn Deutschland gegenüber anderen Ländern einräumen würde, dass hierzulande auch nicht alles zum Besten steht.

[How much do you agree or disagree with the following statement: The world would be a better one if Germany admitted to other countries that over here

there are shortcomings too.]

I4: Inwieweit stimmen Sie den folgenden Aussagen zu oder nicht zu? Ich fühle mich eher als Weltbürger und somit verbunden mit der Welt insgesamt und weniger als Bürger eines bestimmten Landes.

[How much do you agree or disagree with the following statement: I feel more like a citizen of the world, and thus connected to the world as a whole, and less as a citizen of a particular country.]

Stimme voll und ganz zu, Stimme zu, Weder noch, Stimme nicht zu, Stimme überhaupt nicht zu, Kann ich nicht sagen.

[Agree strongly, Agree somewhat, Neither agree nor disagree, Disagree somewhat, Disagree strongly, Don't know.]

Revision I4: Inwieweit stimmen Sie den folgenden Aussagen zu oder nicht zu? Ich fühle mich eher mit der Welt insgesamt verbunden als mit einem bestimmten Land.

[How much do you agree or disagree with the following statement: I feel more connected to the world as a whole than to a particular country.]

REFERENCES

Behr, D., M. Braun, L. Kaczmirek, and W. Bandilla. 2013. "Testing the Validity of Gender Ideology Items by Implementing Probing Questions in Web Surveys." *Field Methods* 25 (2): 124–41. https://doi.org/10.1177/1525822X12462525.

Behr, D., L. Kaczmirek, W. Bandilla, and M. Braun. 2012. "Asking Probing Questions in Web Surveys: Which Factors Have an Impact on the Quality of Responses?" *Social Science Computer Review* 30 (4): 487–98. https://doi.org/10.1177/0894439311435305.

Conrad, F.G., and J. Blair. 2009. "Sources of Error in Cognitive Interviews." *Public Opinion Quarterly* 73 (1): 32–55. https://doi.org/10.1093/poq/nfp013.

Edgar, J. 2012. "Cognitive Interviews without the Cognitive Interviewer?" Presented at the 67th Annual Conference of the American Association for Public Opinion Research. Orlando, FL.

Krosnick, J.A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5 (3): 213–36. https://doi.org/10.1002/acp.2350050305.

Meitinger, K., and D. Behr. 2016. "Comparing Cognitive Interviewing and Online Probing: Do They Find Similar Results?" *Field Methods* 28 (4): 363–80. https://doi.org/10.1177/1525822X15625866.

Willis, G.B. 2005. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. London: Sage.