

An Innovative Approach to Cognitive Testing: Using Cognitive Probes in Production CATI Interviews

Samantha Collins¹, Georgette Lavetsky², Matt Jans³

¹ICF, 126 College Street, Suite 2. Burlington, VT 05401

²Maryland Department of Health, 201 W. Preston St., Baltimore, MD 21201

³ICF, 530 Gaither Rd, Rockville, MD 20850 USA

Abstract

Cognitive testing is a critical step in developing valid and reliable survey questions. Traditional methods typically include structured interviews with a small number of participants from the target population. Cognitive testing's goal is to elicit respondent feedback regarding their interpretation of the proposed questions and any difficulty answering. This process is generally conducted before production interviewing. This presentation asks, "Can we train standardized survey interviewers to administer cognitive probes and obtain information helpful for question revisions?" The question pretest reported was carried out in the Maryland Behavior Risk Factor Surveillance System (BRFSS) state survey. Each year, states provide between two and 20 new survey questions for field testing to determine how well the questions will perform in production interviewing prior to the upcoming survey year. After pretest fielding, response distributions are reviewed, but there is no method of capturing respondents' difficulty answering in a traditional field test. This innovative pretest method incorporated cognitive probes into the BRFSS CATI interview following new questions. In 2016 the Maryland BRFSS tested nine new questions, after which five follow-up cognitive probes were added (e.g., "What did the word 'neighborhood' mean to you in the preceding set of questions," following questions on neighborhood crime). The cognitive probes were written as standardized survey questions to adapt the method to the training and skills of typical telephone interviewers. ICF collected almost 500 completed interviews that included cognitive probe data. Adding cognitive probes to production CATI interviews provided the ability to review and tailor or re-design questions to meet their goals. By adding the cognitive probe questions directly to the CATI interview we were able to reach a much larger number of respondents than smaller traditional sized cognitive tests. The production cognitive testing process will be discussed in the context of pretesting trade-offs and CATI survey implementation.

Key Words: cognitive testing; questionnaire pretesting; BRFSS

1. Introduction

Cognitive interviewing or testing (Beatty, 2004; Willis, 2005) is a critical step in developing valid and reliable questions and ensuring that respondents' answers align with the researchers' expectations (i.e., they match the constructs we're trying to measure). More specifically, cognitive testing can help identify potential measurement error and item nonresponse risks such as complex or confusing grammar, language complexity and readability/literacy issues, question structure issues (such as mismatches between questions and response options), and, in interviewer-administered surveys, any difficulties the

interviewer may have administering the questions. The overarching goal of any cognitive test is to elicit respondent feedback about their interpretation of the questions being tested, their cognitive process while answering, any challenges when answering, and any misunderstandings or cognitive challenges that lead to inaccurate answers.

Cognitive testing is usually done before fielding and with specially trained interviewing staff. Sample sizes are usually small (sometimes less than 10 participants), and the data collected and analysis methods are usually qualitative. Comparatively, full protocol or “dress rehearsal” pretests usually employ production interviewing staff and larger samples (perhaps 50-100 completed interviews) and test the entire survey protocol. Table 1 juxtaposes cognitive pretesting with production or “dress rehearsal” pretesting or pilot testing in a production environment.

Table 1: Differences Between Standardized Survey and Cognitive Interviews

<i>Standardized Interviewing</i>	<i>Cognitive Interviewing</i>
Goal is to record what the respondent says within the provided response options	Goal is to understand what the respondent is thinking and the true source of their answer
Interviewers trained and expected to read exactly what’s on the screen and only probe with approved definitions or “whatever it means to you”	Interviewers trained and expected to probe and go off-script uncover cognitive problems with the question
Interviewers often have lower education with little or no training in interviewing techniques prior to their employment	Interviewers often have higher education and training in psychology, anthropology, or qualitative interviewing
Interviewers not expected to summarize the information they hear	Interviewers expected to summarize their results of each interview independently and develop recommendations

While pretesting best practice involves cognitive interviewing followed by larger-scale dress rehearsal pilots, ongoing, continuous, large-volume surveys like the Behavioral Risk Surveillance System (BRFSS) surveys face logistical and staffing challenges that can prevent employing such a pretesting plan, or simply do not have the for all these pretesting steps. For example, because the survey is ongoing, the overall protocol has been used previously, and a full-protocol dress rehearsal pilot is not needed. Only new questions need to be tested. Further, because data collection is ongoing, there may be little, if any, time for a dedicated cognitive testing phase, which can take up to a month or more including recruitment.¹

¹ Some cognitive testing phases take longer, depending on the sample size, number of questions to test, and methods of analyzing and summarizing data and revising questions. However, one of the authors was recently involved in a cognitive pretesting (n = 9) that took only one week from recruitment through revised questionnaire and recommendations.

1.1 Implementing Cognitive Interviewing in a Production CATI Survey

Alternative methods to traditional small-sample, in-person cognitive interviewing have been proposed (e.g., Fowler and Willis, 2019). However, to our knowledge, cognitive interviewing has never previously been used within production CATI interviewing. That is likely due to assumptions and habits among cognitive pretesters and questionnaire developers. Historically, cognitive testing has been seen as an expertise (rather than a skill) and as the purview of questionnaire experts or psychologists who can, through their magic, discern cognitive challenges that respondents will have. Despite that history, the most recent, formalized cognitive interviewing approaches (e.g., Willis, 2005) incorporate scripted probing so that survey staff with less experience can conduct cognitive interviews along side questionnaire experts. Following this evolution in cognitive interviewing practice, we wondered “Can we fully script very simple cognitive probes, and train production CATI interviewers to field them?” Obviously, there would be some trade-offs. While we would lose the benefit of experienced, trained, and educated survey design experts and the ability to use emergent probing to follow-up on unexpected problems, we thought there would be several benefits to this approach. The first one is the large sample size. With a continuous survey we would be able to pilot test these questions within production data collection and obtain hundreds or thousands of respondents compared to the tens or scores of participants used in typical cognitive interviewing. Second, it would allow us to test the questions with a wider range of people than would ever be possible with standard, qualitative cognitive testing (see Fowler and Willis, 2019 as an exception). In fact, the questions would be tested on the same population and demographic groups as the survey itself. Third, we would be able to assess whether production CATI interviewers would have any challenges reading the questions, and how the questions would fit within the interview. This can only be approximated with traditional cognitive interviewing. Finally, as discussed above, we also thought this would fit well in our project timeline.

The drawbacks to this approach are probably obvious. Using standardized interviewers, there is no flexibility to go off-script or follow-up with probes when a new problem is identified (i.e., no “emergent probing”). Thus, the approach may miss some cognitive problems. Indeed, unless the interviewers happen to notice other problems, and are provided with a mechanism for recording them, the approach explicitly misses cognitive problems that are not asked about in the cognitive probe. We discuss one way to handle this in the next section.

1.2 Research and Implementation Questions

In summary, we had two very simple, proof-of-concept research questions and goals for this study.

- 1) Can standardized interviewers conduct cognitive interviews in the CATI context?
- 2) How useful are quantitative summaries of standardized probes?

2. Method

2.1 Cognitive Testing via CATI for the MD-BRFSS

This cognitive test was conducted during November 2016 within the 2016 Maryland BRFSS (MD-BRFSS) survey in preparation for the 2017 MD-BRFSS. Data were collected as part of the production interviewing sample and schedule, with one small caveat. We targeted 1,150 landline and 500 cell phone completes for the total survey; however, the cognitive probing was conducted on one third of the total November sample (i.e., 496 (30%) of total survey respondents answered the cognitive probe questions, including both landline and cell phone sample).

In total, nine questions were identified as needing cognitive testing during this MD-BRFSS cycle and were included in the cognitive pretest. All questions tested were new questions for the MD-BRFSS. In addition to these questions, five standardized cognitive probes (i.e., follow-up questions) were used. Questions were placed at the very end of the production MD-BRFSS interview so as not to interrupt the flow of the 2016 interview.

Both the test questions and standardized cognitive probes were programmed in production CATI. Interviewers were given minimal “special training” on how to administer the questions, but they were briefed on the presence of these new questions and the goals of the cognitive testing. Training included pre-shift meetings that described the cognitive test and probe questions and explained the purpose of the cognitive probes for each question. Interviewers role-played and used a “practice mode” in the CATI system. In addition to recording answers to the follow-up probes, interviewers were trained to record open-ended notes on challenges with question administration via a feedback form that they could fill out as needed, for example after conducting several interviews that exhibited the same problem.

Unlike traditional cognitive testing, interviewers were *not* instructed to use probes other than those presented on screen, or to go off-script and probe on any cognitive issues that arose. Similarly think-aloud techniques were not used. This approach was chosen to test whether cognitive interviews could be conducted in a truly standardized setting. Similarly, we did not select high-performing or specially trained interviewers. Any MD-BRFSS interviewer working that month had a chance of administering test questions and cognitive probes.

The cognitive probes in this test were standardized follow-up questions designed to draw out various types of knowledge, comprehension, recall and other potential cognitive problems, similar to scripted probes used in traditional cognitive testing. Unlike traditional, qualitative cognitive interviewing, both the responses to the test questions and answers to the cognitive probes were recorded in a closed-ended format. Table 2 contains a list of those questions and their cognitive probes in the order they occurred in the instrument. This paper focuses on interpretation of the two questions in bold.

Table 2: Test Questions and Cognitive Probes

<i>Test Question</i>	<i>Cognitive Probe</i>
<i>Topic 1: Community Support (6 test questions; 2 probes)</i>	
Question 1: Think about your neighborhood when answering the following questions. For this interview, neighborhood is defined as the area within one-half mile or a ten-minute walk from your home. (MD18_1)	- NONE -
Does your neighborhood have any sidewalks?	
1) Yes 2) No ²	
Question 2: For walking at night, would you describe the street lighting in your neighborhood as (MD18_2):	- NONE -
1) Very Good 2) Good 3) Neutral 4) Poor 5) Very Poor	
Question 3: Bike Lanes: How many of the roads and streets in your neighborhood have shoulders or lanes that are marked for bicycling? (MD18_3)	Probe 1: In the preceding question when we said streets “marked for bicycles”, what did that phrase mean to you? (MD18_3c)
Would you say...	SELECT ALL THAT APPLY. (INTERVIEWER READ OUT LOUD)
1) none are marked, 2) some are marked, 3) most are marked, 4) all are marked?	1) Streets that contain shoulders available for bicycling 2) Streets that are marked with defined bicycle lanes 3) Presence of street signage indicating “share the road” with bicycles

² Following survey best practices, “Don’t Know/ Not Sure” and “Refused” were accepted as responses but were not read out loud.

Table 2 (continued): Test Questions and Cognitive Probes

<i>Test Question</i>	<i>Cognitive Probe</i>
<i>Topic 1 (continued): Community Support (6 test questions; 2 probes)</i>	
Question 4: Neighborhood: “On a scale from 1 to 5 where 1 is very unsafe and 5 is very safe, how safe from crime do you consider your neighborhood to be? (MD18_4)	Probe 2: What did the word neighborhood mean to you in the preceding set of questions? (MD18_4C)
<ul style="list-style-type: none"> 1) Very unsafe 2) Unsafe 3) Neutral 4) Safe 5) Very safe 	<p>SELECT ONE (INTERVIEWER READ OUT LOUD)</p> <ul style="list-style-type: none"> 1) The residential street on which you live 2) The residential community in which you live within ½ mile or a 10 minute walk of your home 3) The entirety of the surroundings within ½ mile or a 10 minute walk of your home, including any shopping areas or parks
<p>Question 5: During the past 30 days, for about how many days did you walk in your neighborhood for leisure or as a way to get to your destination? (MD18_5)</p> <p>__ Number of days [range = 01-30] 88 None</p>	- NONE -
<p>Question 6: What is the number one reason that you did not walk more frequently in your neighborhood? (MD18_6)</p> <ul style="list-style-type: none"> 1) Lack of time 2) No sidewalks 3) Too much traffic/traffic noise 4) Medical condition 5) Lack of energy/motivation 6) Exercise elsewhere 7) Concerns for safety/crime 	- NONE -

Table 2 (continued): Test Questions and Cognitive Probes

<i>Test Question</i>	<i>Cognitive Probe</i>
<i>Topic 2: Substance Abuse (3 test questions; 3 probes)</i>	
Question 1: Now I'm going to ask you about <u>non-medical use</u> of drugs. <u>'Non-medical use' means using drugs not prescribed by a doctor, or are used to get high, or for curiosity. Please do not include alcohol or tobacco.</u> (MD19T)	[IF YES TO Q1] Probe 1: Please tell me which types of drugs the preceding question was asking about. For each, please tell me Yes or No. [EACH DRUG PRESENTED ON A NEW SCREEN]
Remember, all answers are kept confidential and no identifying information is recorded. You may skip any question you do not wish to answer.	a) Crack or cocaine? (MD19_2a)
	b) Marijuana? (MD19_2b)
	c) Heroin? (MD19_2c)
	d) Alcohol? (MD19_2d)
In the past 12 months, did you use any drug, on one or more occasions? (MD19_1)	e) Prescription medication taken exactly as prescribed by a doctor? (MD19_2e)
1) Yes	f) Prescription medication NOT taken as prescribed, for example taking double the dose of the medicine? (MD19_2f)
2) No [GO TO NEXT SECTION]	g) Taking prescription medication NOT prescribed directly to you? (MD19_2g)
	h) If someone said they took drugs only for experience, would you describe that as non-medical use of a drug? (MD19_2h)
	Probe 2: Now I am going to read you a list of ways someone might describe drug usage. Please answer Yes or No to the following. (MD19_3)
	a) If someone said they took drugs recreationally, would you describe that as the non-medical use of a drug? (MD19_3a)
	b) If someone said they took drugs only for experience, would you describe that as non-medical use of a drug? (MD19_3b)
	c) If someone said they took drugs for the feelings the drug caused, would you describe that as non-medical use of a drug? (MD19_3c)
	d) If someone takes a medication other than prescribed, would you describe that as non-medical use of a drug? (MD19_3d)

Table 2 (continued): Test Questions and Cognitive Probes

<i>Test Question</i>	<i>Cognitive Probe</i>
<i>Topic 2 (continued): Substance Abuse (3 test questions; 3 probes)</i>	
Question 2: In the past 12 months, did you use heroin or any type of opioid that you did not have a prescription for or that you took more frequently than prescribed, on one or more occasions?	Probe 3: Did you understand the previous question to mean that we were interested in asking about use of heroin, OR all opioids including heroin, OR something else? (MD19_4c)
Opioids include certain painkillers, such as morphine, hydrocodone, and oxycodone; and prescription drugs such as OxyContin, Percocet, and Vicodin. (MD19_4)	1) Heroin 2) All opiates 3) Something else, SPECIFY: (FREE TEXT FIELD)
1) Yes 2) No	
Question 3: In the past 12 months, did you inject or shoot any drug, on one or more occasions? (MD19_5)	- NONE -
1) Yes 2) No	

3. Results

Table 3 shows the quantitative cognitive results for the bike lanes and neighborhood probes within the Community Support topic. Interpreting the follow-up probe frequencies was relatively simple. For example, when asking about bike lanes and markings, it was clear that the vast majority of respondents thought that the term meant that streets had dedicated bike lanes, and that about a quarter thought it meant shoulders big enough for biking. However, 15% of respondents thought that “marked for bicycles” only included signs like “share the road.” Five percent of respondents indicated that all of those definitions apply.

For defining the word “neighborhood,” the largest percentage of respondents (43%) understood the term to mean about a half mile (10-minute walk) around where they live. A substantial percentage (35%) also included shopping areas and parks in that area. Fewer respondents (13%) interpreted the term to only mean the street on which they live.

Just as with substantive survey questions, the cognitive probes included response options for the interviewer to record a “don’t know” response or when the respondent “refused” any specific question. These were not explicitly offered to respondents, but the interviewer could record them. “Don’t know” and “refused” have been cited as indicators of cognitive difficulty (e.g., Beatty and Herrmann; 2002; Jans, 2010), but they may take on a slightly more specific meaning on cognitive probes than substantive questions. On substantive questions we often wonder whether respondents are telling us that they actually don’t know the answer to the question, that they don’t know what the question is asking (or key concepts of the question), or that the “don’t know” response is just a polite way of refusing. In the cognitive probe context, it seems clearer that a “don’t know” tells us that the

respondent cannot decide between the definitions provided. Nine percent of respondents to the bike lanes and markings question, and the same percentage of respondents to the neighborhood question responded with “don’t know” or refused to respond. For both questions about twice as many respondents said “don’t know” as refused. Taken at face value, this suggests question/concept understanding, fatigue, or grammatical issues (i.e., cognitive issues) were more prominent on these two questions than question sensitivity. Despite being provided with candidate definitions, a nontrivial percentage of respondents were not able to map their implicit understanding of the terms to those definitions. In addition to the frequencies of closed-ended probes below, responses to the test questions themselves, and interviewer feedback forms were reviewed for additional information that might help with revising the questions

Table 3: Frequencies of Cognitive Probe Responses for Probes 1 and 2 in the Community Support Topic

<i>Probe Question</i>	<i>Response Option</i>	<i>Percentage</i>
<i>Bike Lanes and Markings (Probe 1 in Community Support topic)</i>		
In the preceding question when we said streets “marked for bicycles”, what did that phrase mean to you?	Streets that contain shoulders available for bicycling (only)	25%
	Streets that are marked with defined bicycle lanes (only)	52%
	Presence of street signage indicating “share the road” with bicycles (only)	15%
	<i>Total substantive (single selections)</i>	92%
SELECT ALL THAT APPLY	Don’t Know / Not sure	7%
	Refused	2%
	<i>Total nonsubstantive</i>	9%
<i>Total (incl. substantive & nonsubstantive responses)</i>		101%*
All 3 response options selected		
<i>Neighborhood (Probe 2 in Community Support topic)</i>		
What did the word neighborhood mean to you in the preceding set of questions?	Residential street on which you live	13%
	Residential community in which you live within ½ mile or a 10-minute walk of your home	43%
	Entirety of the surroundings within ½ mile or a 10-minute walk of your home, including any shopping areas or parks	35%
SELECT ONE	<i>Total substantive</i>	91%
	Don’t Know / Not sure	6%
	Ref	3%
	<i>Total nonsubstantive</i>	9%
<i>Total (incl. substantive & nonsubstantive responses)</i>		100%

*Percentages do not add to 100% due to rounding.

Table 4 shows the original and revised questions for the two focus questions of this paper. For bike lanes and markings, cognitive testing made us comfortable with how respondents understood the core concept, and no changes were needed. However, for feeling safe in one's neighborhood, cognitive testing led to several changes described in the table below. It is important to notice that the changes were not about the neighborhood construct, which was the focus of the cognitive probe. Thus, by using a feedback form and responses to the test questions we were able to uncover problems that were not specifically probed.

Table 4: Original and Revised Questions

<i>Original Question</i>	<i>Revised Question</i>	<i>Summary of Changes</i>
<p>How many of the roads and streets in your neighborhood have shoulders or lanes that are marked for bicycling?</p> <p>Would you say none are marked, some are marked, most are marked, all are marked?" (MD18_3)</p>	<p>How many of the roads and streets in your neighborhood have shoulders or lanes that are marked for bicycling? (MD6_3)</p> <p>Would you say...</p> <p>1 None are marked 2 Some are marked 3 Most are marked 4 All are marked</p>	<p>None. Fielded as tested.</p>
<p>On a scale from 1 to 5 where 1 is very unsafe and 5 is very safe, how safe from crime do you consider your neighborhood to be? (MD18_4)</p> <p>1) Very unsafe, 2) unsafe, 3) neutral, 4) safe, 5) very safe?</p>	<p>How safe from crime do you consider your neighborhood to be? (MOD24_3)</p> <p>Would you say...</p> <p>1) Extremely safe 2) Safe 3) Unsafe 4) Extremely unsafe</p>	<p>1) Removed explanation of scale points from question because it was unnecessary</p> <p>2) Added "Would you say..." intro to response options to be consistent with other similar BRFSS questions</p> <p>3) Removed the "neutral" response option because to force respondents to choose one of the scale or the other (17% of respondents chose the option in the cognitive test)</p> <p>4) Changed scale poles to "extremely" instead of "very"</p> <p>5) Ordered response options from "safe" to "unsafe", reversing the order tested</p>

4. Discussion

This paper reports on a successful quantitative cognitive interviewing pretest conducted in production CATI interviewing. The approach shows promise for future development of cognitive interviewing and testing methods. With the caveats and extensions discussed below, we think that this approach provides a reasonable method for cognitive pretesting when staff, time, or expertise to conduct standard cognitive interviewing are unavailable.

Our first research question was, “Can standardized interviewers conduct cognitive interviews in the CATI context?” Based on this implementation we think the answer is clearly “Yes!”

Our second research question was, “How useful will quantitative summaries of standardized probes be?” Our scripted probe results helped us understand how respondents were understanding key terms, and, in this implementation, made us comfortable with the interpretations we saw. The changes reported for the two items presented in this paper came mostly from the qualitative information that interviewers provided on the feedback form or based on the quantitative distributions of the tested questions themselves, not the probes. Interestingly, qualitative data can play an important role in CATI-based cognitive interviewing.

We see several benefits and drawbacks to this method. The first benefit includes larger sample size than what is available with traditional cognitive interviewing. While we did not take advantage of the large sample sizes to evaluate whether cognitive issues differ by segments of the sample (e.g., across demographic subgroups), that could be done in future implementations. Second, we were able to test questions in the same mode, with the same sample, and with the same type of interviewers that would be employed when they are no longer “test questions.” Third, we found that having to specify one or two scripted cognitive probes for each test question or concept forced us to prioritize our testing needs and develop extremely clear cognitive goals. That is, we could not simply count on cognitive interviewers to suss out the cognitive issues present. We had to think carefully about what, specifically, we wanted to know, where we thought the largest cognitive issue lay, and how to word that in the same way we would word any other standardized survey question.

4.1 Recommendations and Potential Improvements

Despite the relative standardization and automation of the process we describe, we still recommend using a survey design expert to a) review the questions and their goals, and revise questions to remove any obvious problems prior to testing; b) help develop the probes; and c) participate in the analysis and question revisions. Conducting an expert review prior to testing will make the testing more efficient and effective. Including questionnaire design experts in probe development will help ensure that the probes are as useful and targeted as possible.

It is important to remember that, while the overall approach is heavily quantitative, the final analysis is still somewhat qualitative. By that we mean, there are no strict objective cut-offs (i.e., percentages of definition endorsement that would clearly indicate a problem), no statistical hypothesis testing, nor other common quantitative methods. Rather, frequencies of the quantitative cognitive probes were reviewed to assess whether most respondents tended to interpret each question the way we expected. Further, information

provided in the feedback form was not analyzed via a formal process but was reviewed by staff to a) add context to the quantitative results and b) check for any obvious additional problems. When possible, we advise listing which definitions are acceptable (or expected) and which are not. Determining whether definition endorsement rates indicate a problem is significantly easier if the intended or expected meaning is clearly stated.

Finally, our example used fewer cognitive probes than questions tested. However, future implementations should consider using multiple probes for each question like what is done in traditional cognitive interviewing. Similarly, we recommend following up don't know and refusal (i.e., nonsubstantive) responses on cognitive probes. In this study, almost 10% of respondents to each probe gave such a response. This is a fairly large percentage of respondents who could not, or would not, fit their implicit definition of the key term probed into the response options provided. We wonder if those respondents truly did not have access to their implicit definitions, or if their responses are telling us that they had a different understanding of the term that was not offered as a response option. Probing don't know and refusal responses or including a "some other definition" with space for interviewers to record that definition would help provide a clearer picture of respondents' cognitive challenges.

Readers may debate whether our pretest was truly a "cognitive interview," and we welcome that debate. However, at minimum, we have evidence that data and results like those gathered by traditional, small-scale cognitive interviewing can be obtained through large-scale, standardized CATI pretesting. We sincerely encourage readers to try similar techniques in their own pretesting with the goal of expanding the ways in which cognitive pretesting is used in questionnaire design.

Acknowledgements

The authors thank Traci Creller with her assistance preparing this paper.

References

- Beatty, P. (2004). The dynamics of cognitive interviewing. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (Vol. 1–Book, Section, pp. 45–66). John Wiley & Sons, Inc.
- Beatty, P., & Herrmann, D. (2002). To answer or not to answer: Decision processes related to survey item nonresponse. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (Vol. 1–Book, Section, pp. 71–85). Wiley Series in Survey Methodology.
- Fowler, S., & Willis, G. B. (2019). The Practice of Cognitive Interviewing Through Web Probing. In *Advances in Questionnaire Design, Development, Evaluation and Testing* (pp. 451–469). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119263685.ch18>
- Jans, M. (2010). Verbal paradata and survey error: Respondent speech, voice, and question-answering behavior can predict income item nonresponse. <http://deepblue.lib.umich.edu/handle/2027.42/75932>
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Sage Publications.