SURVEYS · QUESTIONNAIRE DESIGN

Cognitive Interviewing as a Tool for Pretesting Questionnaires

Cognitive interviewing asks a small number of people to explain how they understand and answer draft survey questions. This article explains where the method came from, how think-aloud and probing work, how to plan, conduct and analyse cognitive interviews, and what the method can and cannot tell you.

ModQu research article · 2,829 words · 25 references (APA 7th edition)

A survey question can look perfectly clear to the person who wrote it and still be misunderstood by the people who answer it. Researchers who know a topic well are poor judges of how others will read their questions, because they cannot easily set aside what they already know. The only reliable way to find out how a question works is to watch real people answer it and ask them what they were thinking.

Cognitive interviewing is a method for doing exactly that. In a cognitive interview, a member of the survey's target population answers draft questions while the interviewer explores how they understood each question, what they thought about when answering and how they chose their answer. A small number of such interviews, often fewer than twenty, can reveal problems that would otherwise go unnoticed until the survey had been fielded and the data were already compromised.

Since the 1980s, cognitive interviewing has become the most widely used method for pretesting questionnaires in government statistical agencies, academic research and, increasingly, market research. This article explains where the method came from, describes its main techniques, sets out how to plan, conduct and analyse cognitive interviews and discusses what the method can and cannot achieve.

Origins

The method grew out of a movement in the early 1980s to bring the insights of cognitive psychology to survey research. A seminar held in the United States in 1983 brought together psychologists and survey methodologists to consider how people understand and answer survey questions. One of its products was the four-stage model of the response process: comprehension of the question, retrieval of relevant information, judgement and response (Tourangeau, 1984). The model gave researchers a framework for thinking about where questions go wrong, and it suggested that the methods psychologists used to study thinking could be used to study survey answers (Tourangeau et al., 2000).

The most important of those methods was the collection of **verbal reports**, in which people describe their thinking as they solve a problem. Psychologists Anders Ericsson and Herbert Simon argued that such reports could provide valid evidence about mental processes, provided they were collected carefully: concurrent reports of what is currently in a person's attention are more trustworthy than later explanations of why they acted as they did (Ericsson & Simon, 1980).

Other psychologists had shown that people often cannot accurately report the causes of their own judgements, and will offer plausible but mistaken explanations when asked (Nisbett & Wilson, 1977). Both findings shaped how cognitive interviews are conducted.

Statistical agencies in the United States, including the National Center for Health Statistics, the Census Bureau and the Bureau of Labor Statistics, set up "cognitive laboratories" in the late 1980s to test questions in this way. Early taxonomies described the range of laboratory methods being used (Forsyth & Lessler, 1991), and the approach spread to agencies and research organisations in Europe and elsewhere (DeMaio & Rothgeb, 1996). By the 2000s, cognitive interviewing was established as a standard part of questionnaire development (Presser, Couper, et al., 2004).

Two main techniques

Think-aloud

In the **think-aloud** technique, respondents are asked to say everything that goes through their mind as they answer each question. The interviewer's role is mainly to remind them to keep talking. The appeal of think-aloud is that it captures thinking as it happens, with little influence from the interviewer, in line with Ericsson and Simon's recommendations.

Think-aloud has practical drawbacks. Many people find it unnatural and difficult, and the amount and quality of what they say varies greatly. Some respondents fall silent at the moments of greatest difficulty. And the reports can be long and unfocused, making analysis time-consuming.

Verbal probing

In **verbal probing**, the interviewer asks specific follow-up questions about each survey question, either immediately after it is answered or at the end of a section. Probes can target particular aspects of the response process (Willis, 2005).

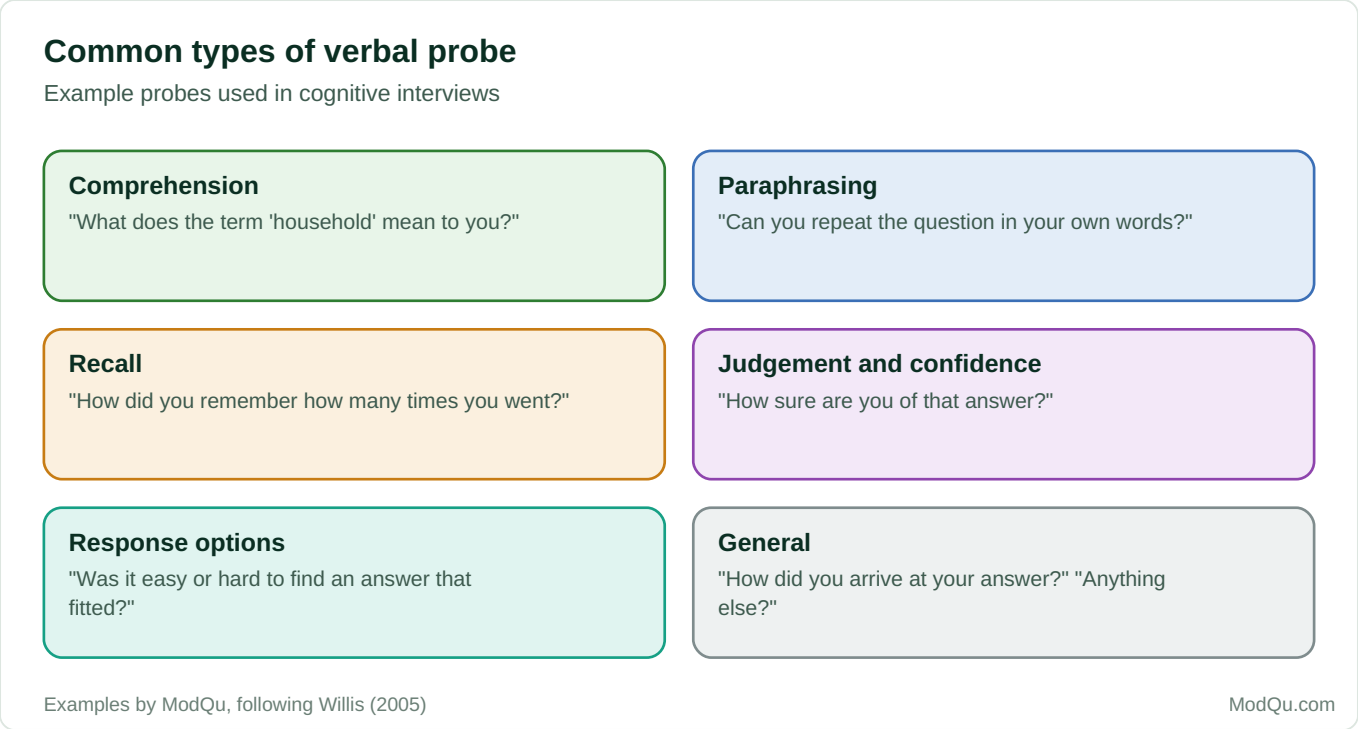


Figure 1. Common types of verbal probe used in cognitive interviews. Examples by ModQu, following Willis (2005).

Probes may be **scripted** in advance, so that every respondent is asked the same follow-ups, or **spontaneous**, arising from something the respondent said or did, such as hesitating or asking for clarification. Most cognitive interviewing combines the two.

Verbal probing is easier for respondents and gives the interviewer more control, which makes it the dominant technique in practice. A review of how cognitive interviewing is actually carried out found that most practitioners rely mainly on probing, often combined with some think-aloud (Beatty & Willis, 2007). Its risk is that probes can create problems that would not arise in the real survey, for example by making respondents reconsider an answer they gave without difficulty, or by suggesting interpretations they would not otherwise have considered.

Comparisons of the two techniques suggest that they find overlapping but not identical sets of problems, and that probing tends to produce more information about specific problems while think-aloud reveals more about respondents' own reasoning (Priede & Farrall, 2011).

Planning a cognitive interviewing study

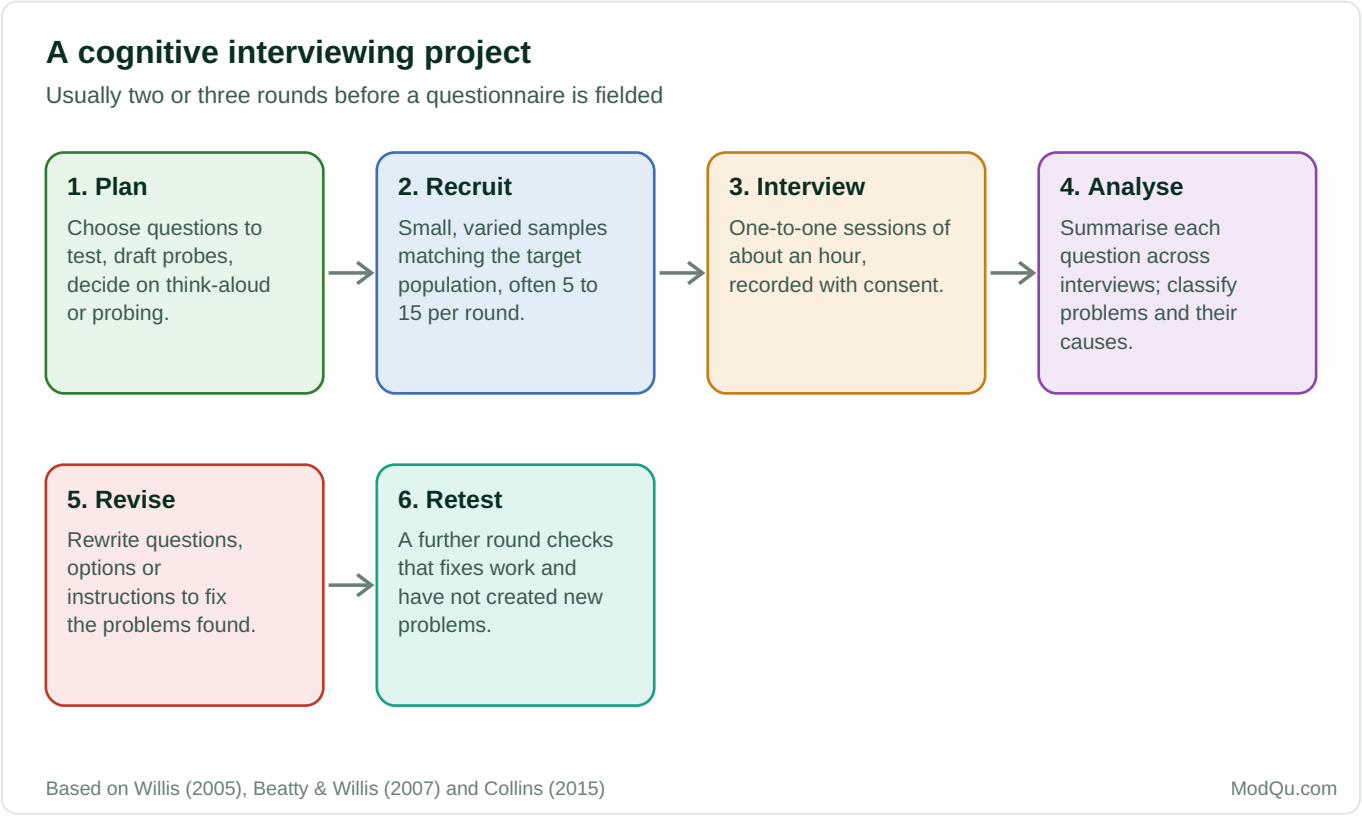


Figure 2. The stages of a cognitive interviewing project. Based on Willis (2005), Beatty and Willis (2007) and Collins (2015).

Decide what to test

Few studies can test every question in a long questionnaire in depth. Priority should go to new questions, questions on difficult or sensitive topics, questions with complex definitions or reference periods, questions adapted from other contexts or translated from another language, and the key measures on which the main conclusions will rest.

Write an interview protocol

The protocol contains the draft questions, the scripted probes for each and instructions for the interviewer. Probes should be chosen to investigate the problems the designer suspects. For a question asking about "household income," for example, a probe might ask who the respondent included in their household and which sources of income they counted.

Recruit participants

Cognitive interviews use small, purposive samples rather than random ones. The aim is not to estimate how common a problem is, but to discover what problems exist and why. Participants should reflect the range of people who will answer the survey, with particular attention to groups likely to have difficulty, such as people with lower literacy, older people,

people answering in a second language or people whose circumstances are complex. If a question about employment will be asked of everyone, the sample should include people with several jobs, the self-employed and people who are not working.

Decide how many interviews

There is no fixed rule for sample size. In practice, many studies conduct between five and fifteen interviews per round, often in two or three rounds with revisions in between (Willis, 2005). Research on this question has shown that the number of problems detected keeps increasing as more interviews are conducted, and that some important problems are found only after many interviews. Small samples are therefore likely to miss some problems, and larger samples are advisable when the stakes are high (Blair & Conrad, 2011).

Conducting the interviews

Cognitive interviews are usually conducted one to one and last about an hour. Longer sessions tire both participants and interviewers. They can take place in a laboratory, in participants' homes, by telephone or by video call. Sessions are normally recorded, with consent, so that the interviewer can concentrate on the conversation.

Interviewers need training and skill. They must put participants at ease, explain that it is the questions being tested and not the participant, administer the draft questions as written, ask probes in a neutral way and follow up on unexpected comments without leading. Studies of cognitive interviews have shown that interviewers' behaviour affects what they find, for example through how often they probe and whether their probes suggest answers (Beatty, 2004; Conrad & Blair, 2009).

Self-administered questionnaires can be tested by asking participants to complete them while the interviewer observes, followed by retrospective probing. For web surveys, observation can be combined with screen recording so that hesitations, changed answers and navigation problems can be reviewed later.

Analysing cognitive interviews

Analysis turns interview notes and recordings into decisions about questions. A common approach is to summarise, for each question, what each participant understood, how they answered and any difficulty they had, and then to look across participants for patterns (Willis, 2015).

Problems can be classified using the stages of the response process. A participant who understood "regularly" as "every day" has a comprehension problem; one who could not remember how many times they had visited a doctor has a retrieval problem; one who did not want to report their alcohol consumption has a response problem. Classification helps identify the right remedy: a comprehension problem calls for clearer wording or a definition, while a retrieval problem may call for a shorter reference period or memory cues.

More recent approaches emphasise interpretation as much as error. Rather than asking only whether a question is "wrong," they ask what the question actually measures for different groups of people, recognising that participants interpret questions through the lens of their own lives. This interpretive approach has been developed particularly for testing questions across cultures and languages (Miller et al., 2014; Willis & Miller, 2011).

Findings are usually written up in a report that describes each question, the problems found, the evidence for them and the recommended changes. Clear documentation allows others to understand and evaluate the decisions made, and it builds a record of how particular question types perform (Collins, 2015).

An example

Suppose a public transport agency wants to ask riders: "In the past month, how often did you use public transport to travel to work?" with options from "never" to "every working day." The question seems simple. Cognitive interviews with eight riders might reveal several problems.

One participant who works from home three days a week is unsure whether "every working day" means every day they work or every day they travel to a workplace. Another, who drives to a station and then takes a train, is unsure whether the trip counts as public transport. A third, a shift worker, points out that their "working days" include weekends. A fourth answers about a typical month rather than the past month, because the past month included a holiday. Probing shows that two participants counted only trips to work, while two others counted round trips as two uses.

None of these problems would be visible in the survey data. Every participant gave an answer, and every answer looked plausible. But the answers mean different things for different people. The revision might define the trip ("any journey where you used a bus, train or tram for at least part of the way"), ask for a count of days rather than a frequency label ("On how many days in the past four weeks did you..."), and state that the count should include only days on which the respondent travelled to a workplace. A second round of interviews would then check that the new version is understood consistently.

Common mistakes

Several mistakes recur in cognitive interviewing studies, particularly when the method is used for the first time.

- **Testing too late.** Cognitive interviews are most useful while the questionnaire can still change. Testing a questionnaire that has already been programmed and scheduled leaves little room to act on the findings.
- **Recruiting the wrong people.** Colleagues, friends or highly educated volunteers are convenient but rarely represent the people who will answer the survey. They tend to understand questions as the designer intended and so miss the problems that matter.
- **Leading probes.** Probes such as "Did you find that confusing?" invite participants to report problems they did not have. Neutral probes, such as "How did you go about answering that?", are more informative.
- **Treating every comment as a problem.** Not everything a participant says calls for a change. Analysts need to judge whether a difficulty is likely to affect many respondents in the real survey and whether a change would introduce new problems.

- **Not retesting.** Revised questions can fail in new ways. A second round of interviews is the only way to know whether the fixes work.

Experienced practitioners recommend combining cognitive interviewing with other evaluation methods in a planned sequence, and recording what was tested and what was changed so that later surveys can build on the results (Presser, Rothgeb, et al., 2004).

What cognitive interviewing can and cannot do

Strengths

Cognitive interviewing is unusually good at revealing why a question fails. Other pretesting methods, such as pilot surveys, can show that a question has a high rate of missing answers or produces unexpected distributions, but they rarely explain why. Cognitive interviews show directly that participants misread a word, could not recall the information, did not find a suitable answer option or found the question intrusive. This makes it possible to fix the problem rather than guess at it.

The method also reveals problems that do not show up in data at all. A respondent who misunderstands a question may give an answer that looks perfectly normal. Classic research showed how often everyday survey questions were understood in ways their designers had not intended (Belson, 1981), and studies of unclear terms have shown how much answers can change when key words are defined (Fowler, 1992). Only by talking to respondents do such misunderstandings come to light.

Limitations

Cognitive interviewing also has well-known limitations.

- **Small samples.** Findings cannot show how common a problem is in the population, and small samples miss some problems (Blair & Conrad, 2011).
- **Artificial setting.** Participants in a cognitive interview pay more attention than typical survey respondents and are aware that they are being studied. Some problems that arise from inattention or fatigue in real surveys will not appear.
- **Interviewer effects.** Probes can create problems as well as reveal them, and different interviewers may reach different conclusions (Conrad & Blair, 2009).
- **Variable results.** A study in which several organisations tested the same questionnaire found that they identified overlapping but different sets of problems (Rothgeb et al., 2007). This does not mean that the findings are unreliable, but it does mean that no single study finds everything.
- **Limits of self-report.** People cannot always report the reasons for their answers accurately (Nisbett & Wilson, 1977), so participants' explanations must be interpreted with care.

For these reasons, cognitive interviewing is best combined with other pretesting methods, such as expert review, pilot testing and, where possible, experiments comparing alternative versions (Presser, Couper, et al., 2004; Lenzner et al., 2016).

Newer developments

Web probing

Cognitive interviewing is labour-intensive, which limits sample sizes. **Web probing** adapts the method for online surveys by inserting open-ended probes after selected questions, such as "What did you think of when you answered this question?" Because probes can be included in surveys with hundreds of respondents, web probing can show how common different interpretations are and can compare groups, for example across countries. Research has examined how the design of web probes affects the quality of responses (Behr et al., 2012) and found that web probing and face-to-face cognitive interviewing identify many of the same issues, while each also finds problems the other misses (Meitinger & Behr, 2016).

Remote interviewing

Video calls have made it easier to conduct cognitive interviews with participants in different locations, including people who would not travel to a laboratory. Remote sessions also make it possible to observe participants completing a web survey on their own device, which is valuable for testing mobile layouts.

Cross-cultural testing

When questionnaires are translated for use in several countries or languages, cognitive interviews in each language can check whether the translated questions are understood in the same way. Coordinated cross-cultural cognitive interviewing studies use common protocols and joint analysis to compare interpretations across groups (Willis & Miller, 2011).

Cognitive interviewing in business research

For businesses that run customer, employee or market surveys, cognitive interviewing offers a practical and affordable way to improve data quality. A handful of interviews with customers can reveal that internal terms, such as product category names, service tiers or department titles, mean little to people outside the company. Interviews can show whether customers interpret a satisfaction scale in the intended way, whether they can recall the information requested and whether answer options cover their situations.

A modest investment is often enough. Two rounds of six to eight interviews, with revisions in between, can be completed within a couple of weeks and will typically improve a questionnaire substantially. For major surveys whose results will drive large decisions, or for questionnaires that will be used repeatedly over many years, a larger programme of testing, combined with a pilot survey, is worthwhile.

Conclusion

Cognitive interviewing gives questionnaire designers direct evidence of how respondents understand and answer draft questions. Grounded in cognitive psychology and refined through decades of use in statistical agencies, it combines think-aloud and verbal probing to uncover problems at every stage of the response process. It cannot measure how common problems are, and its findings depend on careful sampling, skilled interviewing and systematic analysis. Used alongside other pretesting methods, it is one of the most effective ways to make sure that survey questions measure what they are meant to measure (Schaeffer & Dykema, 2020).

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* (pp. 45–66). Wiley. https://doi.org/10.1002/0471654728.ch3

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., 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

Belson, W. A. (1981). *The design and understanding of survey questions*. Gower.

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

Collins, D. (Ed.). (2015). *Cognitive interviewing practice*. SAGE. https://doi.org/10.4135/9781473910102

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

DeMaio, T. J., & Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 177–195). Jossey-Bass.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215–251. https://doi.org/10.1037/0033-295X.87.3.215

Forsyth, B. H., & Lessler, J. T. (1991). Cognitive laboratory methods: A taxonomy. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 393–418). Wiley. https://doi.org/10.1002/9781118150382.ch20

Fowler, F. J., Jr. (1992). How unclear terms affect survey data. *Public Opinion Quarterly, 56*(2), 218–231. https://doi.org/10.1086/269312

Lenzner, T., Neuert, C., & Otto, W. (2016). *Cognitive pretesting* (GESIS Survey Guidelines). GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_en_010

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

Miller, K., Willson, S., Chepp, V., & Padilla, J.-L. (Eds.). (2014). *Cognitive interviewing methodology*. Wiley. https://doi.org/10.1002/9781118838860

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231

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

Presser, S., Rothgeb, J. M., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., & Singer, E. (Eds.). (2004). *Methods for testing and evaluating survey questionnaires*. Wiley. https://doi.org/10.1002/0471654728

Priede, C., & Farrall, S. (2011). Comparing results from different styles of cognitive interviewing: 'Verbal probing' vs. 'thinking aloud'. *International Journal of Social Research Methodology, 14*(4), 271–287. https://doi.org/10.1080/13645579.2010.523187

Rothgeb, J., Willis, G., & Forsyth, B. (2007). Questionnaire pretesting methods: Do different techniques and different organizations produce similar results? *Bulletin de Méthodologie Sociologique, 96*(1), 5–31. https://doi.org/10.1177/075910630709600103

Schaeffer, N. C., & Dykema, J. (2020). Advances in the science of asking questions. *Annual Review of Sociology, 46*, 37–60. https://doi.org/10.1146/annurev-soc-121919-054544

Tourangeau, R. (1984). Cognitive science and survey methods. In T. B. Jabine, M. L. Straf, J. M. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 73–100). National Academy Press.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press. https://doi.org/10.1017/CBO9780511819322

Willis, G. B., & Miller, K. (2011). Cross-cultural cognitive interviewing: Seeking comparability and enhancing understanding. *Field Methods*, *23*(4), 331–341. https://doi.org/10.1177/1525822X11416092

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. SAGE. https://doi.org/10.4135/9781412983655

Willis, G. B. (2015). *Analysis of the cognitive interview in questionnaire design*. Oxford University Press.