

Volume 9 / Number 3 / 2013

ISSN-L 1614-1881 · ISSN-Print 1614-1881 · ISSN-Online 1614-2241

Official Organ of the European Association of Methodology

Methodology

European Journal of Research Methods
for the Behavioral and Social Sciences

3/13

www.hogrefe.com/journals/methodology

Editors

Joop Hox and Nekane Balluerka

Special Issue

Cognitive Interviewing
Reporting Framework

Guest Editors

Gordon Willis
Hennie Boeije

HOGREFE



Contents

Special Issue	Cognitive Interviewing Reporting Framework <i>(Guest editors: Gordon Willis and Hennie Boeije)</i>	
Editorial	The Survey Field Needs a Framework for the Systematic Reporting of Questionnaire Development and Pretesting <i>Gordon Willis and Hennie Boeije</i>	85
Original Articles	The Cognitive Interviewing Reporting Framework (CIRF): Towards the Harmonization of Cognitive Testing Reports <i>Hennie Boeije and Gordon Willis</i>	87
	Examining the Personal Experience of Aging Scale With the Three Step Test Interview <i>Christina Bode and Harrie Jansen</i>	96
	Evaluating the Cognitive Interviewing Reporting Framework (CIRF) by Rewriting a Dutch Pretesting Report of a European Health Survey Questionnaire <i>Rachel Vis-Visschers and Vivian Meertens</i>	104
	Obtaining Validity Evidence by Cognitive Interviewing to Interpret Psychometric Results <i>José-Luis Padilla, Isabel Benítez, and Miguel Castillo</i>	113
	Reflections on the Cognitive Interviewing Reporting Framework: Efficacy, Expectations, and Promise for the Future <i>Gordon Willis and Hennie Boeije</i>	123

Official Organ of the European Association of Methodology

Methodology

European Journal of Research Methods
for the Behavioral and Social Sciences

Your article has appeared in a journal published by Hogrefe Publishing.

This e-offprint is provided exclusively for the personal use of the authors. It may not be posted on a personal or institutional website or to an institutional or disciplinary repository.

If you wish to post the article to your personal or institutional website or to archive it in an institutional or disciplinary repository, please use either a pre-print or a post-print of your manuscript in accordance with the publication release for your article and our “Online Rights for Journal Articles” (www.hogrefe.com/journals).

The Cognitive Interviewing Reporting Framework (CIRF)

Towards the Harmonization of Cognitive Testing Reports

Hennie Boeijs¹ and Gordon Willis²

¹Department of Methodology and Statistics, Faculty of Social and Behavioural Sciences, Utrecht University, The Netherlands, ²Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

Abstract. Cognitive interviewing is an important qualitative tool for the testing, development, and evaluation of survey questionnaires. Despite the widespread adoption of cognitive testing, there remain large variations in the manner in which specific procedures are implemented, and it is not clear from reports and publications that have utilized cognitive interviewing exactly what procedures have been used, as critical details are often missing. Especially for establishing the effectiveness of procedural variants, it is essential that cognitive interviewing reports contain a comprehensive description of the methods used. One approach to working toward more complete reporting would be to develop and adhere to a common framework for reporting these results. In this article we introduce the Cognitive Interviewing Reporting Framework (CIRF), which applies a checklist approach, and which is based on several existing checklists for reviewing and reporting qualitative research. We propose that researchers apply the CIRF in order to test its usability and to suggest potential adjustments. Over the longer term, the CIRF can be evaluated with respect to its utility in improving the quality of cognitive interviewing reports.

Keywords: cognitive interviewing, comparative effectiveness, questionnaire pretesting, standard report format, survey research, checklist

Cognitive Interviewing and Questionnaire Design

A commonly used tool for the development of self-report survey items is the cognitive interview. Most generally, cognitive interviewing (or cognitive testing) is engaged as a means for applying qualitative research methods to the understanding of the functioning of survey questions – as well as of other materials, such as advance letters (Beatty, 2004; DeMaio & Rothgeb, 1996; Willis, 1999, 2005). The premise of this approach is that intensive interviewing of a targeted individual provides rich information that is useful for providing the questionnaire designer with information concerning how questionnaires, and individual survey questions, provide (or fail to provide) desired information.

Although cognitive interviewing is increasingly viewed as a qualitative method within the sociological tradition, its origins are explicitly psychological in nature – as the practice evolved out of the movement known as CASM (Cognitive Aspects of Survey Methodology). During the 1980s, a heavy emphasis was placed on the intersection of survey methods and cognitive psychology, in a manner that produced cognitive interviewing as an interdisciplinary approach to questionnaire development and evaluation (Jabine, Straf, Tanur, & Tourangeau, 1984).

Cognitive Interviewing relies on several key premises. First, its practitioners assert that it is effective to interview small numbers of individuals who serve as “stand ins” for survey respondents, and to conduct these interviews in such a way that we ask *about* the tested survey questions – rather than simply collecting answers to those questions. Second, such interviews are viewed as providing a “window into the mind” that in turn can be used to unearth insights about the evaluated questions. Early cognitive interviews, such as by Loftus (1984), examined issues such as how survey respondents retrieved information from memory, when asked a specific question (e.g., on recall of health care visits). Specifically, when counting such visits over a specific time period, do respondents engage in a forward, backward, or idiosyncratic temporal recall order?

Despite the intrinsic variability in respondent approaches to answering such survey questions, in order to supply a common theoretical framework, Tourangeau (1984) developed a simple yet elegant model of the survey response process that has stood for over 30 years, and explicates four major cognitive processes that respondents are generally presumed to engage in when attempting to answer survey questions: Comprehension of the question; Retrieval of relevant information needed to answer it; a range of Judgment or Estimation processes that are used to integrate and edit this information; and finally, a

Response process in which the individuals convert their internally constructed representation of the answer to one that constitutes their answer to the question, either in spoken or written form (e.g., saying “yes” rather than providing a meandering, conversational response).

Since its inception, notions surrounding the classic cognitive model have been embellished in at least three ways. First is the recognition that motivational, as well as purely mechanistic cognitive processes, are vital to the task of responding to survey questions. Krosnick (1991) in particular has established that respondents make an executive decision concerning the amount of effort they will apply to the question-answering task, and may engage in “satisficing,” in order to produce an answer that matches the requirements, but which may not be the result of complete processing. Second, it has been recognized that the cognitive stages may not be carried out in invariant fashion, but are more akin to subroutines which are selectively engaged by the respondent, and that may be missed, repeated, or reordered (Herman, Johnson, McEvoy, Herzog, & Hertel, 1996). Third, and perhaps most importantly, more recent conceptualizations of the survey response process have viewed the entire activity in a wider manner, by appropriating concepts from disciplines other than psychology. In particular, linguistic and sociological-anthropological viewpoints have broadened the scope of answering survey questions to include a range of sociocultural elements, in which the question is answered in a situational or life context, rather than within a world that consists only of the respondent interacting with the survey question (Collins, 2007; Gerber & Wellens, 1997; Miller, 2003, 2011). For example, questions concerning physical activity that specify leisure-time activities of high-income Non-Hispanic respondents in the United States (e.g., tennis, golf, weightlifting, running) may fail to capture the types of activities that are carried out by low-income women within Hispanic cultures.

Conduct of the Cognitive Interview

The conceptualization of the cognitive interview, as discussed above, impacts the manner in which the interview is carried out. Classically – based on a purely cognitive point of view – the investigator carries out the activity by focusing on the tested subject’s cognitive processes, to view what is presumably happening inside the “black box” of the mind (again, Comprehension, Retrieval, Decision/Judgment, and Response). This is accomplished through two operational means. The first is the “think-aloud,” in which the participants are asked to report everything that comes to mind as they are mentally processing a presented survey item, and then answering it (Ericsson & Simon, 1980). The second fundamental approach is for the interviewer to administer verbal probe questions that are specifically designed to target the key underlying cognitive processes; that is, to metaphorically “probe the mind” for specific information. Hence, the notion is that one can test the survey question – for example, “Overall, how happy are you

these days?” – by following up the response (“Somewhat happy”) with the probe “And what, to you, is ‘happiness’ as it is used in this question?” To the extent that subjects have access to useful information concerning their own conceptualization of what it means to be “happy,” and are able to articulate that in response to a probe question, cognitive testing may inform the questionnaire design and evaluation process.

The specific activities involved in collecting information through either think-aloud or verbal probing are specified in detail in a separate book (Willis, 2005). To summarize, critical points are as follows:

- (1) Think-aloud and verbal probing can be used in conjunction, and are in fact normally combined within the same study, although practitioners have increasingly come to rely on targeted probing, in part because many individuals are not adept at think-aloud. Further, probing is under the control of the investigator and puts less demand on the subject;
- (2) Verbal probing may be either concurrent or retrospective: The former involves probing immediately after the subject has answered a tested survey question, prior to administering the next evaluated question. Retrospective probing is also known as debriefing: The interviewer defers probing until all evaluated survey questions are administered. Both concurrent and retrospective probing forms persist, as there are tradeoffs in the utility of their usage (Willis, 2005).
- (3) Probing is done with a specific purpose – normally the aim is viewed as locating problems in survey questions (e.g., a term, such as “abdomen,” is not well understood; or a question like “how many times have you ridden in a passenger airplane?” poses a difficult recall task). However, an alternative to a viewpoint emphasizing the identification of design defects instead advocates the objective of understanding how a question works, and “what it captures” – without necessarily seeking to remediate sources of error. For example, the investigator may endeavor to capture the full range concerning what individuals think when asked about “health in general,” and to determine how this conceptualization varies across cultural groups, without necessarily seeking a solution for any particular deficiency with respect to item construction.

Regardless of the approach taken to gathering data in the cognitive interview, or the underlying objectives, this information is overwhelmingly qualitative, in the common and well-accepted sense of the term (Collins, 2007; Conrad & Blair, 2004; Miller, 2011). Decisions concerning item functioning mainly derive from written, descriptive information that is gathered by the trained cognitive interviewer (e.g., “None of the subjects I tested was able to recall how many times they had consulted the internet for health information within the past 12 months”). As such, it is necessary to gain a sufficient level of expertise, in interviewing (gathering

meaningful and clear qualitative data), performing qualitative data-reduction activities, and then in interpreting the information collected (to avoid idiosyncratic or biased interpretation). It is the analysis of cognitive interviewing in particular that may be the least-developed aspect of the entire process. Again, the various approaches to analysis that have been used over the past 30 years are detailed in Willis (2005), and additional approaches that are especially appropriate for cross-cultural or cross-national research are suggested by Miller (2011), Miller, Mont, Maitland, Altman, and Madans (2010), and by Fitzgerald, Widdop, Gray, and Collins (2011).

Towards the Harmonization of Cognitive Interviewing Reports

Variation in Cognitive Interviewing Approaches

As mentioned previously, there are significant differences in the nature of data collection (think-aloud vs. probing), probing (concurrent vs. retrospective), and analysis. Beyond this, there are also key issues concerning selection of appropriate sample sizes, and the way in which cognitive interviews are divided into “rounds” of testing. A common approach has been to conduct a small number of interviews (8–12), and then to stop and assess results, making modifications to tested questions before retesting through a subsequent round. Such testing is iterative in nature, in that the sequence of testing-and-modification may be carried out through three or even more such rounds. Alternatively, however, a researcher could decide to conduct all the interviews without stopping to make changes – or, the complete opposite: through modifying questions after each interview, such that the size of the “testing round” is effectively one interview. The relative benefits of these potential alternative approaches have not been well studied, and little empirical evidence exists to suggest that common practices represent optimal (or even minimally acceptable) solutions.

Further sources of variation in cognitive interviewing practice exist with respect to a multitude of variables: for example, the number of cognitive interviewers who conduct the interviews (e.g., one interviewer conducting all the interviews, vs. a larger number who conduct several interviews each); the nature of selection and training of interviewers; or the details of the probing approach used (with respect to what are labeled by Willis (2005) as Proactive vs. Reactive probes).

In itself, variation in procedures is not necessarily a negative feature. It does, however, produce two vexing problems for practitioners: (a) Because procedures tend not to be well described, it can be difficult to determine the methodological steps that a particular investigation has taken; and (b) It therefore follows that a determination of the efficacy of any particular variant of cognitive interviewing will be exceedingly difficult to ascertain. Given the nature of current practices in the production of cognitive interviewing

reports, resolving these issues, and developing a set of best practices, relies on a body of evidence that simply does not exist at this time. However, in order to reach the point at which comparative effectiveness of varied approaches can be established (*outcome evaluation*), it is first necessary to take the initial step of creating a means for systematically describing the procedures that have been used in cognitive interviewing studies, so that the requisite body of evidence in fact exists (i.e., to support the initiation of the *process evaluation* of cognitive interviewing approaches).

Benefits of a Standard Reporting Format

We argue that the lack of comprehensiveness of the information contained in cognitive testing reports is largely because there is currently no well-defined, standard report format, or even a standard for specifying the minimal level of information that should be contained within a report. As a result, reports tend to be idiosyncratic in both the types and ordering of information contained, certainly between organizations, and sometimes even within. Hence, reports may be missing crucial pieces of information (e.g., the number of cognitive interviews, or cognitive interviewers, involved in the testing project), or may present these in ways that makes searching for the information difficult.

Developing a Reporting Framework

The problems mentioned above call for the development of a unified approach to the types of information to be contained in a cognitive testing report. The current article, therefore, is intended to present a solution to this challenge, through the introduction of a conceptual reporting framework. As previously mentioned, cognitive interviewing represents a qualitative research approach. Hence, cognitive interviewing practices can benefit from what has been learned in qualitative research in fields other than survey question evaluation, with regard to the reporting and evaluation of the conducted research project.

Procedure

We began by examining instruments that have been developed for assessing and reviewing the quality of qualitative studies. Quality is a much-debated topic in qualitative research, and the lack of consensus on what constitutes quality is reflected in the large number of different instruments to review qualitative investigations (Boeije, van Wesel, & Alisic, 2011; Cohen & Crabtree, 2008). We first selected four quality-oriented checklists as a source for generating items for our own purposes (see Table 1). We preferred checklists with explanation attached to the different items as to be sure about their meaning. We sought checklists that originated in different disciplines or branches, that had been generated over different time

Table 1. Checklists used to generate the Cognitive Interviewing Reporting Framework (CIRF)

	Number of items
<i>Quality checklists used as a start</i>	
1 Qualitative research checklist British Medical Journal http://www.bmj.com/about-bmj/resources-authors/article-types/research/editors-checklists	12
2 Critical Review Form – Qualitative studies (version 2.0) Letts, L., Wilkins, S., Law, M., Stewart, D., Bosch, J. & Westmorland, M. (2007). McMaster University. http://www.srs-mcmaster.ca/Portals/20/pdf/ebp/qualreview_version2.0.pdf	20
3 Quality in qualitative evaluation: a framework for assessing research evidence Spencer, L., Ritchie, J., Lewis, J. & Dillon, L. (2003). National Centre for Social Research. http://collections-r.europarchive.org/tna/20070705130742/ http://www.policyhub.gov.uk/docs/qqe_rep.pdf	18
4 Criteria for the evaluation of qualitative research papers British Sociological Association Medical Sociology Group Seale, C. (1999). <i>Quality in qualitative research</i> . London: Sage.	20
<i>Quality checklists used to check coverage</i>	
5 Critical Appraisal Skills Programme (CASP) (2011) http://www.casp-uk.net	10
6 Step-by-step guide to critiquing research. Part 2: qualitative research Ryan, F., Coughlan, M. & Cronin, P. (2007). <i>British Journal Nursing</i> ; 16(12): 738–744.	16
7 Critical appraisal checklist for qualitative research studies Treloar, C., Champness, S., Simpson, P.L. & Higginbotham, N. (2000). <i>Indian Journal of Pediatrics</i> , 67(5): 347–351.	10
8 Critical appraisal of focus group research articles Vermeire, E., Royen, P. van, Griffiths, F., Coenen, S., Peremans, L. & Hendrickx, K. (2002). <i>European Journal of General Practice</i> , 8(3): 104–108.	13
9 Qualitative research review guidelines – RATS (Relevance, Appropriateness, Transparency, Soundness) Clark, J.P. (2003). How to peer review a qualitative manuscript. In: F. Godlee & T. Jefferson (Eds.). <i>Peer review in Health Sciences</i> , pp. 219–235. London: BMJ Books. http://www.biomedcentral.com/info/fora/rats	10
10 Evolving guidelines for publication of qualitative research studies in psychology and related fields. Elliot, R., Fischer, C. & Rennie, D.L. (1999). <i>British Journal of Clinical Psychology</i> , 38: 215–229.	14
<i>Reporting guidelines for qualitative research</i>	
11 Consolidated criteria for reporting qualitative research (COREQ) Tong, A., Sainsbury, P. & Craig, J. (2007). <i>International Journal for Quality in Health Care</i> . 19(6): 349–357.	32
12 Reading qualitative studies Sandelowski, M. & Barroso, J. (2002). <i>International Journal of qualitative methods</i> . 1(1): 74–108. Creativecommons.org/licenses/by/2.0	15

periods, and that varied in length, in order to cover a broad spectrum.

Second, we extracted items from the four quality checklists and grouped them into eight relevant clusters that became the skeleton of the framework (analysis is available from authors on request): (1) research objectives, (2) research design, (3) ethics, (4) sampling, (5) data collection, (6) data analysis, (7) results, and (8) documenting the study (auditability). As a consequence of combining the different checklists, some items were redundant and eliminated in each category. The first draft of the resulting framework was developed by the first author and checked by the second.

Third, the framework that resulted was assessed by consulting six other checklists assessing quality of qualitative research (see Table 1). On the basis of these new materials we decided to divide the cluster “results” into (a) “findings” and (b) “conclusions, implications, and discussion.” The cluster “documenting the study (auditability)” was split into (a) “quality and auditability” and (b) “report format.” This operation resulted in 10 total clusters.

Fourth, in addition to consulting checklists devoted to aspects of quality, we consulted a checklist intended for the reporting of qualitative research (see Table 1), to explicitly formulate criteria for providing information concerning studies, using the appropriate style, and degree of

specificity. Finally, we adjusted the new checklist, through adoption of terminology that adheres to cognitive interviewing methods and results. For instance, the word “sampling” was replaced with “participant selection,” and “quality and auditability” was replaced with “strengths and limitations of the study.” The resultant framework was labeled the Cognitive Interviewing Reporting Framework, or CIRF (see Table 2).

The CIRF: Cognitive Interviewing Reporting Framework

In this section we describe the categories in the CIRF conceptual framework, to facilitate its use by other researchers.

Research Objectives

As in all research projects, researchers need to formulate and describe the research objectives of the investigation. Cognitive testing research normally involves the pretesting of new questions prior to fielding, or else conducting quality assessment of already developed questions. Efforts can also be aimed at finding problem areas, and discovering resolutions, with respect to the entire survey. They can also be aimed at examining possible problems in only one target group or finding problem areas in only the items that were recently adjusted. Alternately, the objective might be to determine if an item, or full questionnaire, is interpreted similarly across cultural groups or languages. This initial section is also an appropriate place (but not the only one) in which the to-be-evaluated items can be listed or referred to.

A justification for cognitive pretesting or evaluation could be the identification of clearly anticipated problems, within the initial stages of developing questionnaire. In particular, a previous expert review can indicate some anticipated problem areas that are subsequently dealt with in a cognitive interviewing project. Alternatively, when an instrument is closer to fielding, the research objectives can be more confirmatory in nature, and intended to determine that no serious problems remain.

Concerning the context of cognitive pretesting, the report can address questions such as: Who has decided to evaluate the instrument (e.g., a client or sponsor, researchers internal to the organization, students, and so on)? What parties are involved in the pretest? How much modification to items will be allowed, and what other constraints exist? In particular, are there any different agendas involved that pose conflicting demands (e.g., pressure for maintaining data trends which may restrict capacity for item modification)?

A review of background literature may be relevant, for users to know whether any previous research has been conducted into the use of the current items or the questionnaire as a whole; for instance, where the questions have been used previously, and whether other pretesting studies were

done. Or, for newly developed items, a literature review might reveal existing instruments that measure the phenomenon of interest, and identify gaps or limitations to be redressed.

We also make reference to theoretical perspective, because cognitive interviewing is rooted in diverse disciplines and it is important to report on the specific theoretical perspective that underlies the pretesting of the survey. Again, some researchers make use of the well-known cognitive model developed by Tourangeau (1984); whereas those having a sociological perspective may make explicit use of a Grounded Theory approach (e.g., Boeije, 2010). Cross-cultural studies may rely on relevant classification schemes such as those by Fitzgerald, Widdop, Grey, and Collins (2009).

Research Design

Research design is concerned with the study's methodology. The study might be designed as an experiment comparing different cognitive interviewing methods; or may emphasize subgroup comparison. For instance, Willis and Zahnd (2007) included groups based on both cultural group membership (Koreans vs. non-Koreans) and level of acculturation to the US society (high vs. low). Also relevant to research design is the process used for cognitive probing (e.g., based on interviewer-participant verbal interaction with concurrent probing, versus unprobed self-administration followed by a debriefing session). Concerning the structural design of the pretest, the project might also involve a clear sequencing of identifiable testing stage – for example, the Three-Step-Test-Interview (TSTI) introduced by Hak, van der Veer, and Jansen (2004). Finally, researchers can indicate the degree to which the procedures were fixed, as opposed to flexible and modifiable, through the course of pretesting or evaluation. It is important that cognitive interview researchers provide adequate description, and optimally justification, for their choice of each component of the research design.

Ethics

This category describes the relevant ethical issues, and how possible benefits and costs to participants were considered, for example with respect to appropriate level of monetary compensation for participation. We acknowledge that issues of ethics and of human subjects protection may not be pronounced in many cognitive interviewing investigations. There are, however, situations in which this topic can be significant, where sensitive or emotional content is involved, or where the establishment of a relationship with participants might influence their expectations concerning services or advice provided by the interviewer or agency conducting the interview (e.g., for a cognitive interview involving risk behavior or medical care, is the interviewer precluded from providing advice or information?). In general, any potential harm or threats posed by the interview

Table 2. Cognitive Interviewing Reporting Framework

1. Research objectives

Define the research objectives

- What are the aims of the study?
- What is the context that gave rise to pretesting the instrument?

Provide a review of relevant background literature

- What is the theoretical perspective for the cognitive interviewing study?

2. Research design

Describe the features of the overall research design

- What was the basis for each feature of the design?

3. Ethics

Present evidence of thoughtfulness about research contexts and participants

- Was the study approved by an ethics committee or IRB? (consent procedures)
- How was the research project introduced to settings and participants?
- How were people motivated to participate?
- How was confidentiality and anonymity of participants/sources protected?

4. Participant selection

Describe the participant selection methods used

- What are participants details with respect to demographics and other project-specific items of information
- Did the selection of participants satisfy the study objectives?

5. Data collection

Provide information about the data collection methods

- Who conducted the interviews and how many interviewers were involved?
- How were the interviewers trained?
- Were sessions recorded and if so, was audio or video used?
- Were notes taken and how was this employed?
- What type of verbal reporting method was employed, that is, think-aloud, probing, or combinations?
- Was the interview protocol adjusted during the research process and if so, how?
- Was saturation achieved?

6. Data analysis

Describe methods of data analysis in this research project

- How were raw data transformed into categories representing problem areas and solutions?
- What software programs were used?
- Has reliability been considered, including the repetition of (parts of) the analysis by more than one researcher?
- How did researchers work together and how were systematic analysis procedures encouraged, especially between laboratories or testing locations?
- Were there any efforts for seeking diverse observations, that is, triangulation?
- Was quantitative evidence used to supplement qualitative evidence?

7. Findings

Present findings in a systematic and clear way, either per-item, per meaningful part of the questionnaire, or per entire questionnaire

- What was observed concerning subject behavior with respect to *each* evaluated item?
- To what extent did results differ as a function of subject characteristics, behaviors, or status?

8. Conclusions, implications, and discussion

Address the realization of the objectives

- If possible, include a copy of the modified questions if one was produced as a product of testing.
- How do findings and solutions relate to previous evidence?

9. Strengths and limitations of the study

Discuss strengths and limitations of the design and employment of the study and how these could have affected the findings

- What were relevant a priori expectations or previous experiences?
- What are the implications of findings for generalization to the wider population from which the participants were drawn, or applicability to other settings?
- What is the study's contribution to methodological development and future practice?

10. Report format

Use a structured and accepted format for organizing the report

- Include main study documents that are relevant for independent inspection by others as appendix or online materials.

should be discussed. Further, if the research was reviewed by an Institutional Review Board (IRB) or other body, that should be stated. Note that information on Ethics need not constitute a separate section, but should be included where it is appropriate and readily identifiable.

Participant Selection

Reports should identify the target population for the questionnaire and in particular the (sub) populations that were involved in pretesting the questionnaire. Describe how setting (e.g., organization providing facilities or space; physical interviewing location) was chosen and how participants were selected. Describe recruitment strategies used (advertisement, word-of-mouth, etc.). In particular, if critical subgroups were included, these should be indicated (e.g., for a project that tests tobacco questions, indication of how many participants were current smokers, former smokers, and never smokers).

Data Collection

Details of data collection are sometimes seen as mundane and unworthy of mention, yet several key issues are vital for purposes of full disclosure of methods, and enabling replication of the processes used. Reports should make clear details such as whether a single cognitive interviewer, as opposed to a team, was used, and their relevant training and backgrounds. Data collection varies significantly, depending on whether this involved simple note-taking during the interview, as opposed to audio/video recordings which are reviewed later. Procedural variables include the nature of cognitive probes used, and whether they were standardized or administered flexibly. Because best practices have yet to be developed with respect to all of these areas, it behooves investigators to indicate which practices they have made use of, as the field works toward identifying those that prove most effective.

Critically, indicate the degree to which *saturation* was achieved. Saturation is in theory obtained when no new information is gathered once additional participants are recruited and interviewed. It is possible that researchers decide to stop data collection when the most serious problems seem to have been detected and/or resolved. There might be other reasons to stop data collection as well, such as time constraints and lack of funds or staff. In particular, determinants of the sizes of the participant groups, and number of testing rounds, should be well explicated.

Data Analysis

Analysis may be the single most serious “black box” in cognitive interviewing reports, as investigators rarely describe how they moved from data collection to the production of results and recommendations. As such, reports

must be careful to describe how the analysis took place within the current project. Overall, investigators should indicate the method of data management and data reduction they used to transform a series of individual comments, pertaining to separate survey items across multiple interviews, into a coherent set of summary findings that transcend the individual interview level. This may have involved the use of a coding scheme, or else the use of thematic reduction intended to capture major themes. For investigations that involve multiple investigators, or groups of investigators, data review and summary steps are sometimes done independently by each researcher subgroup, before being compared and further combined. In other cases, original comments consisting of raw data from interviews are reviewed by all investigators – that is, at the lowest possible level – before being further processed. Again, current guides to cognitive interviewing provide little guidance in these areas, so a series of reports that chronicle the approaches commonly taken would at least provide a rich description of the strategies that are currently in use.

Reports should also describe the ways in which discussions with those not directly involved in the testing, such as clients, stakeholders, or other researchers, were used to inform analysis and interpretation. Finally, given the increased application of mixed-method approaches for pretesting and question evaluation, if quantitative methods were used in association with qualitative ones (e.g., psychometric evaluation of responses), these should also be described.

Findings

Findings need to be presented in a systematic and clear way, either per-item, per meaningful part of the questionnaire, or per the entire questionnaire. Preferably, findings not only include descriptions of problem areas, but also of insights into what caused these problem areas, and indications of potential solutions. If the intent of the investigation is not to identify problems, but simply to capture the range of interpretation concerning each evaluated item, then this full range should be described, for every target item (that is, reporting should not be selective or arbitrary). Reports of findings may involve examples, but these should not be “cherry picked” to support the researcher’s a priori hypothesis – rather, the totality of the relevant findings should be presented.

Conclusion, Implications, and Discussion

Conclusions can be narrow – that is, restricted to the use of the current set of questions within the context of the current survey; or they can be wide, to the extent that they address the use of the evaluated items more generally, whether in terms of other surveys, or of measurement objectives other than those within the current investigation. Often the overall conclusions can be made prominent through their inclusion in an executive summary; whereas more detailed,

question-by-question conclusions can be listed under each evaluated survey item.

Strengths and Limitations of the Study

In all cases, researchers should indicate their level of confidence in the results they have obtained, based partly on the clarity of the results, and on the extensiveness of the investigation. If sample sizes were small, testing until saturation not done, and the reports by different interviewers in conflict, it would be imperative to state these limitations.

Report Format

The CIRF does not imply a strict sequence of elements: our suggested section ordering may be appropriate for some reports but not others. It may be especially useful to include an executive summary at the beginning of the report for readers uninterested in methodological details.

Discussion

Our overall assertion is that cognitive interviews can be considered to be a type of qualitative research, and that cognitive interviewing practice can benefit from what has been learned and realized in qualitative research in other fields. On the basis of 12 existing quality and reporting checklists that have been developed for qualitative research reports, we generated the CIRF as a framework for reporting cognitive face-to-face interviewing studies. The ten-category CIRF checklist is meant to encourage researchers to report on their cognitive interview projects, and to do so in a clear and comprehensive way. In our experience, many researchers refrain from disseminating cognitive interview reports in a wide manner.

However, within other areas of research, there is evidence that the quality of reporting is improved when an organizing scheme is used, and results are shared. As a precedent, the use of a statement of reporting randomized trials (RCTs), the CONSORT statement, has been evaluated for its effects on the quality of reports of RCTs (Moher, Jones, & Lepage, 2001). These authors found a decrease of unclear reporting, and an improvement in the quality of reports. This finding supports the use of a reporting framework to encourage the development of best practices within the field of cognitive interviewing. As a consequence, surveys in turn will benefit from these systematic reports, as our cognitive pretesting processes become more reliable and effective.

We have refrained from being prescriptive and rigid with respect to best practices for the *conduct* of cognitive interviewing. Rather, we strive for clear *reporting* of cognitive interviewing studies, in order for the users to be able to assess the value of these studies. However, the framework offers the users flexibility. This is deemed necessary, as not

all studies contain all elements that the checklist covers (or, they may include additional elements).

A logical first step is the use of the CIRF framework by authors to report their work, and then to reflect upon the framework, in terms of its clarity and usefulness. Within this special issue, several authors take that step. First is “*Examining the personal experience of Aging Scale with the Three Step Test Interview*,” by Bode and Jansen (2013). The investigators apply a particular pretesting variant that includes cognitive interviewing – the TSTI – and attempt to fashion their write-up in a manner that utilizes the CIRF framework. Following this, they consider potential positive and negative effects of following the CIRF.

Next, Vis-Visschers and Meertens (2013) present “*Evaluating the CIRF by rewriting a Dutch pretesting report of a European Health Survey Questionnaire*.” The investigators reformatted an existing cognitive testing report so that it matches the CIRF format, and upon reflection, make conclusions concerning the comprehensiveness of the CIRF in providing relevant information.

The subsequent case study, “*Obtaining validity evidence by cognitive interviewing to interpret psychometric results*,” by Padilla, Benítez, and Castillo (2013), broadens the potential application of the CIRF, by including both cognitive interviewing and psychometric approaches according to a mixed-method pretesting and evaluation model. The authors consider the extent to which a CIRF-type checklist can be expanded to include pretesting procedures other than cognitive testing.

To round out the Special Issue, Willis and Boeije (2013) summarize the research reported, in “*Reflections on the Cognitive Interviewing Reporting Framework: Efficacy, expectations, and promise for the future*.” The authors present common themes from the three case studies, in terms of how the CIRF can be applied across a range of survey questionnaire pretesting environments. Then, they suggest further directions that can be taken for purposes of (a) facilitating the future use of the CIRF by a range of researchers; (b) evaluating and modifying the CIRF further; and (c) proposing specific points of entry within existing systems (e.g., the extant Q-Bank database system) into which the CIRF may be incorporated.

Following this set of enhancements, we expect to further the objective of producing clear reports of cognitive interviewing studies, through further elaboration of the CIRF checklist. Eventually, we hope to produce a clear, logical, and easy-to-use standard that is also effective in enhancing the quality of cognitive testing research.

References

- Beatty, P. (2004). The dynamics of cognitive interviewing. In S. Presser, et al. (Ed.), *Questionnaire development evaluation and testing methods* (pp. 45–66). Hoboken, NJ: Wiley.
- Bode, C., & Jansen, H. (2013). Examining the personal experience of Aging Scale with the Three Step Test Interview. *Methodology*, 9, 96–103.
- Boeije, H. R. (2010). *Analysis in qualitative research*. London, UK: Sage.

- Boeije, H. R., van Wesel, F., & Alisic, E. (2011). Making a difference: Towards a method for weighing the evidence in a qualitative synthesis. *Journal of Evaluation in Clinical Practice*, 17, 657–663.
- Cohen, D. J., & Crabtree, B. F. (2008). Evaluative criteria for qualitative research in health care: Controversies and recommendations. *Annals of Family Medicine*, 6, 331–339.
- Collins, D. (2007). Analysing and interpreting cognitive interview data: A qualitative approach. In *Proceedings of the 6th Questionnaire Evaluation Standard for Testing Conference* (pp. 64–73). Ottawa: Statistics Canada.
- Conrad, F., & Blair, J. (2004). Data quality in cognitive interviews: The case for verbal reports. In S. Presser, et al. (Ed.), *Questionnaire development evaluation and testing methods* (pp. 67–87). Hoboken, NJ: Wiley.
- 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. 175–195). San Francisco, CA: Jossey-Bass.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87, 215–251.
- Fitzgerald, R., Widdop, S., Gray, M., & Collins, D. (2011). Identifying sources of error in cross-national questionnaires: Application of an error source typology to cognitive interview data. *Journal of Official Statistics*, 27, 1–32.
- Fitzgerald, R., Widdop, S., Gray, M., & Collins, D. (2009). Testing for equivalence using cross-national cognitive interviewing. *Center for comparative social surveys*. Working Paper Series, No 01.
- Gerber, E. R., & Wellens, T. R. (1997). Perspectives on pretesting: Cognition in the cognitive interview? *Bulletin de Methodologie Sociologique*, 55, 18–39.
- Hak, T., van der Veer, K., & Jansen, H. (2004). The three-step test-interview (TSTI): An observational instrument for pretesting self-completion questionnaires. In *ERIM Report ERS-2004-029-ORG*. Rotterdam, The Netherlands: Erasmus Research Institute of Management. Retrieved from <http://hdl.handle.net/1765/1265>
- Herman, D., Johnson, M., McEvoy, C., Herzog, C., & Hertel, P. (1996). *Basic and applied memory research: Vol. 2. Practical application*. Hillsdale, NJ: Erlbaum.
- Jabine, T. B., Straf, M. L., Tanur, J. M., & Tourangeau, R. (Eds.). (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, DC: National Academy Press.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213–236.
- Loftus, E. (1984). Protocol analysis of responses to survey recall questions. In T. B. Jabine, M. L. Straf, J. M. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 61–64). Washington, DC: National Academy Press.
- Miller, K. (2003). Conducting cognitive interviews to understand question-response limitations. *American Journal of Health Behavior*, 27(Suppl. 3), S264–S272.
- Miller, K. (2011). Cognitive interviewing. In K. Miller, J. Madans, A. Maitland, & G. Willis (Eds.), *Question evaluation methods: Contributing to the science of data quality*. New York, NY: Wiley and Sons.
- Miller, K., Mont, D., Maitland, A., Altman, B., & Madans, J. (2010). Results of a cross-national structured cognitive interviewing protocol to test measures of disability. *Quality & Quantity*, 4, 801–815.
- Moher, D., Jones, A., & Lepage, L. (2001). Use of the CONSORT statement and quality of reports of randomized trials: A comparative before-and-after evaluation. *JAMA*, 285, 1992–1995.
- Padilla, J.-L., Benítez, I., & Castillo, M. (2013). Obtaining validity evidence by cognitive interviewing to interpret psychometric results. *Methodology*, 9, 113–122.
- Tourangeau, R. (1984). Cognitive science and survey methods: A cognitive perspective. In T. Jabine, M. Straf, J. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey design: Building a bridge between disciplines* (pp. 73–100). Washington, DC: National Academy Press.
- Vis-Visschers, R., & Meertens, V. (2013). Evaluating the Cognitive Interviewing Reporting Framework (CIRF) by rewriting a Dutch pretesting report of a European Health Survey Questionnaire. *Methodology*, 9, 104–112.
- Willis, G. B. (1999). *Cognitive interviewing: A how-to guide*. Retrieved from <http://appliedresearch.cancer.gov/areas/cognitive/interview.pdf>
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.
- Willis, G., & Boeije, H. (2013). Reflections on the Cognitive Interviewing Reporting Framework: Efficacy, expectations, and promise for the future. *Methodology*, 9, 123–128.
- Willis, G., & Zahnd, E. (2007). Questionnaire design from a cross-cultural perspective: An empirical investigation of Koreans and Non-Koreans. *Journal of Health Care for the Poor and Underserved*, 18, 197–217.

Received August 9, 2012

Accepted April 22, 2013

Published online August 2, 2013

Gordon Willis, PhD

National Cancer Institute
6130 Executive Blvd., MSC 7344, EPN 4005
Bethesda, MD 20892
USA
Tel. 001-301-594-6652
E-mail willisg@mail.nih.gov