

Postprint Version	1.0
Journal website	http://econtent.hogrefe.com/doi/abs/10.1027/1614-2241/a000074?journalCode=med
Pubmed link	
DOI	10.1027/1614-2241/a000074

This is a NIVEL certified Post Print, more info at <http://www.nivel.eu>

Reflections on the Cognitive Interviewing Reporting Framework: Efficacy, Expectations, and Promise for the Future

GORDON WILLIS¹ AND HENNIE BOEIJE²

¹National Cancer Institute, National Institutes of Health, Bethesda, MD, USA,

²Department of Methodology and Statistics, Faculty of Social and Behavioural Sciences, Utrecht University, The Netherlands

ABSTRACT.

Based on the experiences of three research groups using and evaluating the Cognitive Interviewing Reporting Framework (CIRF), we draw conclusions about the utility of the CIRF as a guide to creating cognitive testing reports. Authors generally found the CIRF checklist to be usable, and that it led to a more complete description of key steps involved. However, despite the explicit direction by the CIRF to include a full explanation of major steps and features (e.g., research objectives and research design), the three cognitive testing reports tended to simply state what was done, without further justification. Authors varied in their judgments concerning whether the CIRF requires the appropriate level of detail. Overall, we believe that current cognitive interviewing practice will benefit from including, within cognitive testing reports, the 10 categories of information specified by the CIRF. Future use of the CIRF may serve to direct the overall research project from the start, and to further the goal of evaluation of specific cognitive interviewing procedures.

Boeije and Willis (2013) have argued that the effective conduct of cognitive interviewing requires the use of a systematic, complete, and harmonized system of reporting standards, and have introduced the Cognitive Interviewing Reporting Framework (CIRF) as a means for accomplishing this goal. As an attempt to apply and to initially evaluate the CIRF, the preceding three papers within the current issue each attempted to attend to the CIRF in a variety of ways (Bode & Jansen, 2013; Padilla, Benitez, & Castillo, 2013; Vis-Visschers & Meertens, 2013). Each of these authors adopted the CIRF categories to fashion a testing report, based on information from a cognitive testing project that had already been conducted. Based on their experiences, and on a review of their reports, we make conclusions concerning the

utility of each of the 10 major CIRF checklist categories, and then turn to overall reflections on the CIRF.

ASSESSMENT OF THE 10 CIRF CATEGORIES

CIRF Category 1: Research Objectives

It does appear that the authors of the cognitive interviewing studies within this volume – even when induced to utilize the CIRF – tended to neglect to include a complete description of the research objectives of the investigation. The major assertion made was that a questionnaire required evaluation, and it was assumed that cognitive testing was appropriate for achieving this end. Neither Bode and Jansen, nor Vis-Visschers and Meertens, explain why they felt cognitive testing was an appropriate method, but rather assumed this to be the case – perhaps because cognitive testing has become commonplace and well accepted as a means for evaluating survey questions. On the other hand, Padilla et al. began by describing the decision to include cognitive testing, perhaps because their aim was to contrast this with another method (psychometric analysis).

The CIRF Research Objectives category also suggests providing a description of theoretical perspectives, and in particular to indicate how features of the evaluated questionnaire, and anticipated problems, drive the selection of the particular cognitive approach used. For example, if particular comprehension challenges are expected, the report will presumably mention that probe questions were fashioned to target these areas. However, the three reports featured in this volume tended to simply summarize their theoretical approach in a generic manner – that is, that cognitive testing will be used to examine comprehension, retrieval, judgment, and response processes, according to the standard Tourangeau (1984) model.

There is of course merit to remaining open to a consideration of multiple potential sources of error, rather than overly narrowing one's viewpoint or expectations at the outset. On the other hand, pretesting investigations can sometimes demand a degree of targeting. For example, a questionnaire on lifetime physical activity would seem to demand attention to retrieval processes, given the long reference period involved, and perhaps an approach to cognitive probing that focuses on that issue. We therefore believe that although cognitive interviewers should avoid being narrowly focused, concerning the way theory applies to their investigation, a description of particular hypotheses or expectations that are somewhat more specific may be appropriate, within the opening section of the report. As such, we continue to advocate the CIRF's emphasis on providing information concerning why, and how, cognitive interviewing can be expected to satisfy the demands of the investigation.

CIRF Category 2: Research Design

The intent of the CIRF category describing Research Design is to describe the features of the overall design, and to describe why particular variants of cognitive testing parameters were selected. For example, testing of a self-administered paper questionnaire involving complex skip instructions may call for purely retrospective verbal probing, in order to avoid interruption of navigational behavior that may occur if concurrent probing were chosen. We note that all three authors selected retrospective verbal probing. However, only Padilla et al. describe why they chose

that technique; whereas Vis-Visschers and Meertens state that they chose it but without explanation as to why; and Bode and Jansen rely on a technique (the Three-Step Test Interview, or TSTI), that appears to involve retrospective probing, but not by that name. Further, although all three investigations appear to have relied on scripted probes (i.e., cognitive probes fashioned prior to the interviews), it is not clear to what extent they also relied on more flexible forms of probing (those labeled *spontaneous* or *emergent* by Willis, 2005). Yet, the selection of such probes constitutes one of the more interesting, though contentious, issues in the field, and is important to describe in a testing report.

We continue to argue that a full description of detailed design information is vital. First, this will be of use in creating a body of evidence concerning how cognitive testing is commonly conducted, in practice. Second, to the extent that authors provide information reflective of the efficacy of their investigations, it may be possible to make conclusions concerning the association between particular approaches and the type or quality of results obtained (e.g., the usefulness of concurrent probing versus retrospective when assessing navigational behavior on a web-based survey).

CIRF Category 3: Ethics

Although cognitive interviews generally do not present the same degree of ethical challenge as, for example a clinical trial involving disease therapy, ethical issues still pertain. Participants within cognitive interviewing studies almost never receive any direct positive benefit, as the entire endeavor is more geared toward the researcher's goals (or those of private industry, in case of marketing surveys).

Therefore, investigators should pay attention to how they respected the interests of participants. Further, they should provide information on basic issues such as whether the interviews were conducted in a private location, and whether individual responses were systematically protected. At the least, a statement should appear stating that the investigation was approved by an ethical review body such as an Institutional Review Board, or else was conducted in a manner consistent with the ethical requirements established by the testing organization. We observed that the manuscripts in the current volume were led, through use of the CIRF, to at least mention issues of research ethics. Vis-Visschers and Meertens state that attention to this topic was prompted explicitly by the CIRF, and that it would not normally be included in a report that adhered to their organization's guidelines. Especially given the tendency of journal editors to increasingly request such information, it seems that inclusion of this topic in the CIRF is a useful contribution.

CIRF Category 4: Participant Selection

All three studies in this volume describe participant selection methods used in the way intended by the CIRF, although these did not universally capture both (a) how recruitment was done (procedurally), and (b) who was recruited (e.g., by providing a table of participant demographic characteristics). Further, despite the intention of the CIRF, no study commented directly (in any part of the report) on whether the individuals successfully recruited did serve to satisfy the purposes of the research, whether enough subjects were studied (e.g., whether saturation was attained), or whether there were "gaps" with respect to particular types of individuals who were observed.

Somewhat in that vein, however, Padilla et al. note that:

“.. the CIRF obliged us to report on methodological issues that are often missed in cognitive interviewing studies. For example, in the Method section for Study 2 we reported demographics such as “type of house- hold” and the reason why it is relevant for the study aims.”

We believe that the CIRF therefore has utility in leading researchers to report not only what was done, but why – for a number of methodological variables such as participant selection. As a negative feature, Padilla et al. also suggest a drawback of the CIRF, concerning its emphasis on issues such as participant selection that are derived from its focus on cognitive interviewing, and apply less well to other forms of question evaluation. In response, we propose that any question pretesting method that relies on an empirical base (i.e., testing on human respondents) necessarily involves elements of participant selection, by design.

Even if the nature of that selection departs markedly from that of cognitive testing, the benefit of including informa- tion concerning how participants were selected remains clear, and calls for the inclusion of this item of information in pretesting reports.

CIRF Category 5: Data Collection

As developers of the CIRF, we believe that it is important to include information about details such as who conducted the interviews, how many interviewers were involved, and how they were trained. The manuscript authors varied in satisfying these requirements. Bode and Jansen state that in the absence of the CIRF requirement, they would not have focused on interviewer training. However, there are also seemingly mundane issues listed within the CIRF that authors in this volume tended to ignore, such as: Were ses- sions recorded, and if so, was this by audio or video-record- ing?; Were notes taken during the interview, and how were these in turn utilized to create the data that need to be pro- cessed in the analysis and relied upon to make conclusions and recommendations?; Were the responses to the evalu- ated items themselves recorded and made use of in the anal- ysis, or did data consist solely of responses to probe questions, or interviewer comments on subjects’ behaviors and verbalizations?

It may be because of the seemingly uninteresting nature of these details that even investigators who rely on the CIRF left these details out. However, some of the more interesting discussions related to cognitive interviewing procedure hinge on such detail, as the complex journey between the conversation we refer to as the “cognitive interview,” and the set of final statements we present as “recommendations,” is fraught with potential pitfalls and sources of uncontrolled variation (Conrad, Blair, & Tracy, 2000; Willis, 2005). It is only to the extent that researchers can be led to appreciate the importance of such details, and to take steps to report them, that the field will develop the documentation necessary to meaningfully assess how cog- nitive testing is accomplished, and ultimately, whether cer- tain approaches reveal superior outcomes.

To generalize the point, we conclude that it is vital to include appropriately detailed information on the mechanics of data collection, with appropriate information on key staff, for any type of qualitative, empirically-based questionnaire evaluation method. Hence, the basic logic implied by the CIRF can easily be adapted to whatever pre-testing procedure is selected.

CIRF Category 6: Data Analysis

It does seem that data analysis presents a conundrum to cognitive interviewers just as it does for other areas of qualitative research, and the current set of papers reflects this dilemma (for more discussion of this issue see Boeije, 2010). It appeared to be difficult for authors to provide a cogent description of analytic procedures in a way that

the study could be reproduced by making use of the report. As is the case for many cognitive testing reports, the authors tend to present conclusions concerning what was found, in the absence of much information concerning how the data collected led to those conclusions – reports tend to make statements such as “cognitive interviewing revealed that item X was difficult to comprehend,” without indicating how that determination was made. In particular, a section describing the analysis will optimally indicate the criterion used for determination that the data in fact reflect the presence of a problem (e.g., whether at least one participant was recorded as producing a response that indicated a comprehension problem; whether a threshold was used in which “most” participants had problems; or something else). Hence, the CIRF attempts to steer researchers toward an increased specification of such rules. Although qualitative research techniques have often relied on coding of text-based data, and the CIRF explicitly calls for a description of coding procedures, it appears that coding of responses was generally not done within the cognitive interviewing studies reported within this volume (with the exception of Vis-Visschers and Meertens, within the part of their study that involved behavior coding rather than cognitive interviewing). It is common practice within the survey pretesting field to conduct analysis in a manner that foregoes formal coding (Willis, 2005). However, the rationale for the process of moving directly from raw data (participant quotes or interviewer comments) to statements about the efficacy of a particular question wording is seldom explained. As a step toward clarifying the analysis process, Vis-Visschers and Meertens did provide an example of an analysis spreadsheet used to guide the analysis, and that in part addresses the question of how analysis was conducted. Further efforts in this direction would certainly facilitate the objective of making the analysis process more transparent.

The CIRF also calls for a discussion of whether the investigators made an attempt to establish the reliability of the judgments made by the interviewers or analysts. This was not done by the authors in this volume, again as is the case for cognitive testing reports generally (Willis, 2005). Although interobserver reliability of results is a critical issue in the evaluation of cognitive interviewing as a science, it has seldom been directly addressed. It would therefore be valuable for investigators to report the degree of correspondence there was in interpretations concerning item function made by independent reviewers, analysts, or coders. If multiple interviewers each compiled results of their own interviews independently, did their observations converge? If a coding or categorization system was used to organize the results, was

any test of intercoder reliability (such as kappa statistic or intra-class correlation coefficient) used?

CIRF Category 7: Findings

The CIRF directs that the findings be presented in a clear and systematic way. This requirement demands that the results are logically organized, for instance by first presenting findings related to individual tested item, and then including an overall summary. It should be clear what the data (the observations) consist of, and how they are interpreted. In the current case, all three research reports chose to provide examples of particular items that were found to be problematic, with detailed descriptions of the authors' interpretations, diagnoses, and potential remedies. However, only Padilla et al. provide information concerning the frequency with which various interpretations were observed. Although that information is often not supplied in qualitative research projects, it may sometimes be useful in conveying information concerning whether the examples provided were representative, or rather represent a biased selected by the researchers.

CIRF Category 8: Conclusion, Implications, and Discussion

It is generally assumed in cognitive testing reports that the findings identified are representative of those expected to occur within a field administration of the instrument – even though we generally make no attempt to utilize a representative sample in a statistical sense. At best, the argument is made that because a range of individuals were tested, it was at least possible to identify a range of problems with the item – or to at least gain a comprehensive understanding of the manner in which it can be expected to function. A fair analysis, however, would demand that the investigators reflect upon the role that sample bias may have had in biasing or limiting the results obtained. As such, it would have been useful for the report authors to have considered the element of representativeness in their reports and reflect upon generalization to the populations; however, this element was missing from all three reports (although we note that Bode and Jansen did reflect on the extent to which the environment represented by the cognitive interviews represented that likely to exist in a field survey, and relevant implications).

CIRF Category 9: Strengths and Limitations

Strengths and caveats are a normal part of a usual journal manuscript, but are not heavily featured within the reports in this volume applying the CIRF. Vis-Visschers and Meertens note that a discussion of these issues would not normally be provided in an internal organizational report, as they are not relevant to the consideration of how to improve the questionnaire. We suggest that in some cases, however, an explicit consideration of the relative positive and negative features of the research design – or its implementation – could be very relevant to issues of question improvement. In particular, to the extent that the investigation was seen as methodologically strong (e.g., by recruiting a sufficient number of participants who

vary in ways that satisfied recruitment objectives, and that are deemed to provide a good simulation of field respondents within the final survey), then the results of testing should be relied upon heavily in making re-design decisions. If, on the other hand, the investigation was relatively weak (only a few sub-jects were tested, the results were not fully analyzed before decisions needed to be made, etc.), then this would guide the investigators to perhaps weigh the cognitive testing results more lightly, when considering the wisdom of enact- ing major changes to item content, wording, format, or organization. We therefore continue to believe that the requirement to consider self-identified strengths versus weaknesses is a valid one. To some extent, this practice could benefit future researchers – who can be made more aware not only of the potential of various varieties of cog- nitive testing, but also of its pitfalls.

Further, we feel that such an emphasis will also help researchers to more actively consider the manner in which decision-making should be made, and how the integration of various forms of information (both quantitative and qual- itative) should be carried out. At the least, an honest assess- ment of the strength of the cognitive investigation as implemented – as opposed to simply the argument that cognitive testing is generically effective – may help the field to move away from blanket statements such as “We know the questions work because they were tested.” Beyond specifically reporting the results of a pretest of a survey instrument, researchers are also requested by the CIRF to make suggestions related to methodological devel- opment and future practice of cognitive interviewing. To this end, Padilla et al. concluded that the mixing of qualita- tive and quantitative pretesting methods (in this case cogni- tive interviewing and Item Response Theory) has benefits for instrument evaluators, and should be further developed.

CIRF Category 10: Report Format

Padilla et al. commented that they felt overly constrained by the CIRF format, and used a format that departed in some way from that strictly imposed by adherence to the CIRF. Our intention is that the CIRF be applied in a flexible way in which the ordering of major categories can be somewhat varied, or combined (e.g., Participant Selection with Research Design; or Data Collection with Ethics), as appropriate for a particular study or organization. Authors varied in their assessment of the degree to which the CIRF is either too sparse, versus overly demanding of detail. Vis- Visschers and Meertens describe the CIRF as a “minimal standard” that contains only the necessary infor- mation. On the other hand, Bode and Jansen comment that reporting of the CIRF categories may be excessive, requiring more space than is available in some journal articles.

DISCUSSION

Within the current issue of *Methodology*, we have attempted (a) to identify a problem – deficiency in the reporting of cognitive interviewing results; (b) to propose a solution consisting of the use of the CIRF; and (c) to pro- vide several examples of its application. On this basis we can venture conclusions concerning the utility of the CIRF approach. As a caveat, we recognize the significant limita- tion to our evaluation that none of the three featured research projects made use of the CIRF

at the outset of project planning, or during its conduct. Hence, authors had no opportunity to utilize the CIRF as a framework for development and reporting, but rather determined whether the information they had already collected could be adapted to develop a CIRF-oriented report. We therefore cannot make conclusions, based on this exercise, concerning the usefulness of the CIRF in driving the overall direction of a cognitive interviewing research project, or its reporting. Despite this limitation, we present several overall conclusions.

Usefulness of the Checklist Approach

As intended, the checklist format of the CIRF effectively leads researchers to include relevant, reportable details they would not otherwise include in their reports. Further, the CIRF may have utility prior to, during, and subsequent to the development of a research protocol for questionnaire evaluation. Researchers might review the CIRF to determine that cognitive testing is especially appropriate, given the resources required and the types of information gained. Even where the investigator already plans to conduct cognitive interviewing, the CIRF checklist may still be an effective guide in leading them to carefully record each critical piece of methodologically-relevant information in a systematic way, which will ease later reporting.

Comprehensiveness of the CIRF Categories

Concerning the adequacy of the CIRF categories for portraying the results of cognitive interviews once these have been conducted, this again seemed to have more positive than negative features. Again, the special issue authors believed that the CIRF helped them to include information they otherwise may not have thought to incorporate. On the other hand, authors cited two detrimental features of the CIRF: (a) The suggested organization and ordering of key categories may in some cases be suboptimal; and (b) Inclusion of all the required information produced reports that may be too long, especially when published as journal articles.

Implications, and Potential Next Steps for CIRF Development

Inclusion of Auxiliary Information Within Websites

We believe that the fundamental approach represented by the CIRF – that of a systematic checklist approach – has merit mainly because it leads to the inclusion of important material within the reports that otherwise may be (and often is) missing. Although additional information does require more space, we suggest that the length of a CIRF-based report can be tempered through a strict attempt to be very precise and efficient in use of language. Further, the provision of more complete information can be satisfied through the practice of making use of auxiliary, supplementary venues, and in particular website repositories. Increasingly, as online journals become more prominent, it may even be relatively straightforward to incorporate active links within the electronic version of a manuscript, that can lead the reader directly to access copies of tested (or final version) questionnaires,

cognitive interviewing protocols, longer reports, or other materials that can be useful in providing a complete picture and that incorporates all of the information referred to in the CIRF. Of course, there are challenges – not unique to cognitive testing reports – to the notion of maintaining supplemental libraries of information: website addresses tend to have limited lifespans, and over time lead to “dead links” that fail to provide the promised information. Therefore, to avoid degradation of available information, organizations must make an active attempt to produce archives that have a reasonably long “shelf life.”

Incorporation of the CIRF Into Other Systems

A further means by which cognitive interviewing reports can be “kept alive” for an extended period – in a way that includes full information as advocated by the CIRF approach – is to develop and populate a self-contained, special-purpose repository for such information that is actively maintained over an indefinite period. The Q-Bank system, currently housed at and maintained by the Centers for Disease Control, National Center for Health Statistics, is designed specifically for this purpose (<http://wwwn.cdc.gov/qbank/Home.aspx>). Currently, cognitive interviewing reports, in their original forms (i.e., absent further restructuring), are included as read-only documents within the system. Further, the Q-Bank system is searchable by topic, by organization, and even by question phrasing, which greatly facilitates access to particular information within these reports. For example, a user interested in how the ubiquitous question “Would you say your health in general is excellent, very good, good, fair, or poor?” has fared under the lens of cognitive interviewing can search on that question, and then read through all reports within Q-Bank in which that item has been evaluated. Further, readers can access methodological details contained within these reports in a less formalized way, by reading the reports themselves, even where these cannot be searched electronically.

Although Q-Bank represents a major step forward in enhancing the accessibility of cognitive testing reports, it cannot in itself ensure that the elements recommended by the CIRF, let alone the full set of methods-related details that allow one to fully understand the study, are contained in the reports. In fact, a perusal of the reports currently contained in Q-Bank reveals that these vary in organization, style, and content. As such, it is very difficult to quickly read through reports in a way that one can quickly locate critical information, or even to expect that it is contained within. Clearly, a more standardized, parallel structure for report-writing would ease the uptake of information, and ease the user’s cognitive burden. We propose that the CIRF could have a positive role in this regard, by serving as a template for writers of reports that are to be submitted to

Q-Bank – or at the least, through its promotion of the concept of report-writing standards.

Extension to Other Pretesting Methods

Beyond its application to cognitive interviewing, we also view the CIRF as a potentially expandable means for producing more standardized report writing across a range of questionnaire pretesting and evaluation methods. The current volume anticipates this trend, by including a paper (Padilla et al.) that considered whether the CIRF can be extended to a study that includes multiple question evaluation procedures (cognitive testing and Item Response Theory analysis). Based on the Padilla et al. report, it does appear that the fundamental elements of the investigation are somewhat similar, across methods, and that relatively modest modifications to the CIRF categories can be made to generalize the approach. A range of pretesting and evaluation procedures depart from cognitive interviewing mainly in the types of details that are appropriate to discuss under each major heading. It would not seem that difficult to tailor the more detailed information under the major CIRF categories to a range of evaluation techniques.

CONCLUSION

We view the CIRF as a dynamic system that will evolve over time, as it either becomes clear that it can be made more efficient, more encompassing, or otherwise more effective; or as the question development field evolves and places more or different demands on a report-writing system. Development of the CIRF probably will not require full-scale revision, as its core elements are likely to be acceptable to researchers who conduct cognitive interviews. Rather, the major challenge we face involves disseminating information about the CIRF, and enticing writers of cognitive testing reports to make use of this tool, both as a heuristic device and as a practical means for communication of their results. Presumably, the exposition of the method within a Special Issue of a journal devoted to the topic of research methodology is a major initial step. In order for a new procedure to gain traction within the research community, however, it is necessary that it become adopted by a critical mass of individuals and organizations that constitute the target audience.

Facilitating such a trend

objective of producing report-writing standards is a potentially beneficial direction; (b) that this is in fact feasible; and (c) that the CIRF is the optimal (or at least reasonable) approach. We are hopeful that practitioners of cognitive interviewing will agree that a common reporting system would in fact be both beneficial and feasible, and that adoption of the CIRF by the scientific community will turn it into a useful and well-regarded tool.

REFERENCES

- Bode, C., & Jansen, H. (2013). Examining the Personal Experience of Aging Scale with the Three Step Test Interview. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 9, 96–103. doi: 10.1027/1614-2241/a000071
- Boeije, H. R. (2010). *Analysis in qualitative research*. London, UK: Sage.
- Boeije, H., & Willis, G. (2013). The Cognitive Interviewing Reporting Framework (CIRF): Towards the harmonization of cognitive testing reports. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 9, 87–95. doi: 10.1027/1614-2241/a000075

- Conrad, F., Blair, J., & Tracy, E. (2000). Verbal reports are data! A theoretical approach to cognitive interviews. In *Proceedings of the 1999 Federal Committee on Statistical Methodology Research Conference* (pp. 317–326). Washington, DC: Office of Management and Budget. Retrieved from <http://www.bls.gov/osmr/pdf/st990240.pdf>
- Padilla, J. L., Benitez, I., & Castillo, M. (2013). Obtaining validity evidence by cognitive interviewing to interpret psychometric results. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 9, 113–122. doi: 10.1027/1614-2241/a000073
- 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 CIRF by re-writing a Dutch pretesting report of a European health survey questionnaire. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 9, 104–112. doi: 10.1027/1614-2241/a000072
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.