

Methods for the Analysis of Cognitive Interviews

Johnny Blair¹ and Pat Dean Brick²

¹Abt Associates Inc., 55 Wheeler Street, Cambridge, MA 02138-1168

²Westat, 1600 Research Blvd., Rockville, MD 20850

Abstract

Cognitive interviewing has become a predominant method of survey question pretesting. Despite the stature that cognitive interviewing enjoys as an established and respected pretesting method, there is little consensus on the ways in which the verbal reports should be handled or how the analysis should be conducted. Nor is there a widely accepted notion of what comprises “evidence” from a cognitive session. In many important ways, cognitive interviewing has evolved from and become different from the original think aloud and verbal protocol methodologies of cognitive psychology. We revisit the original think aloud and verbal protocol methodology and examine how current question pretesting practices align with the original goals and limitations of the think aloud method.

Key words: Cognitive interview; think aloud; verbal reports; verbal protocol; analysis of cognitive interviews; questionnaire testing.

1. Introduction

Cognitive interviewing, a pretest method widely used since the late 1980's, has not developed generally accepted practices for analyzing the data that the interviews produce. The data are verbal reports from interview respondents. These reports may result from respondents thinking aloud while answering draft survey questions; or the verbal report may be a response to an interviewer's probe about a question's meaning, or even a volunteered respondent comment. The analysis of cognitive interviews is essentially the extraction of information from these verbal reports about question performance. There are different procedures in use by practitioners to review the verbal reports. Some research evidence indicates that cognitive interview verbal reports are not reliably interpreted (Conrad and Blair 2004).

Beatty and Willis (2007) in discussing error in cognitive interview analysis note that problem identification by independent groups of researchers may be inconsistent. But it is difficult to know how much of that type of inconsistency is due to variations in interviewing processes or to unreliable analysis procedures. Yet neither type of finding seems to have increased practitioners' concerns about analysis methods. Drennan (2002), in an overview of cognitive interviewing concluded that “...the overall approach of analyzing cognitive interview data remains overtly subjective, and this remains the greatest flaw in an otherwise comprehensive method of questionnaire pretesting.”

In this paper, we review some methods that have been used in survey research and other disciplines that use the analysis of various types of verbal reports as a research tool. Our objective is to determine whether some of these analysis methods may usefully inform current analysis practices.

We first review the origins of cognitive interviewing; then consider why some methods previously used for cognitive interview data analysis that have not been adopted. And we briefly examine experiences analyzing verbal reports in a couple of other fields. Based on these observations, we suggest one possible direction for improving analysis practice.

This small inquiry is part of a larger plan to develop a descriptive classification of analysis methods, noting the conditions under which particular methods have been effective— and to provide these findings to practitioners to help inform their choice of analysis techniques for specific applications.

2. Origins of Cognitive Interviewing

Cognitive interviewing to pretest questionnaires was motivated by the work of Ericsson and Simon (1984, 1993) who used “thinking aloud” to elicit data about cognitive processes and analyze that data using protocol analysis. “The central assumption of protocol analysis is that it is possible to instruct subjects to verbalize their thoughts in a manner that does not alter the sequence and content of thoughts mediating the completion of a task and therefore should reflect immediately available information during thinking” Ericsson (2006). It is important to note that Ericsson and Simon devoted nearly as much attention to describing the characteristics of methods that do not elicit valid data as to developing methods that produce veridical verbal reports of cognitive processes.

Verbal protocol analysis has been an important research method in a number of fields. It has been effective partly due to following a well-defined set of procedures for both the collection and analysis of verbal reports. An illustration of the method is given by Ericsson (2006) who explains how studies of the structure of expert performance, using chess expertise as an example, have been done. The “structure of expert performance” in this case is not about measuring expert players’ knowledge of the game or testing their memory for games they have studied. Instead, the objective summarized by Ericsson (2006) is for “researchers [to determine] how players win tournament games.” He uses a study by Charness (1981) who elicited 136 think aloud protocols¹ from club-level and expert-level chess players. Each verbal report was coded into a “problem behavior graph” which is a detailed structural representation of the search process for deciding chess moves. Thus coded, the data were analyzed both qualitatively and statistically to identify differences in the search processes of players of different skill levels. The key features of this research being strict think aloud procedures, an a priori coding scheme based on a model of chess move decision making, and analyses that were specified before data collection. This type of data collection and analysis has been used successfully in survey methods research, and attempted in cognitive interview practice. It would be useful to understand why it succeeded in the one area, but not the other.

Some early uses of verbal reports in survey research were investigations of the response process. The focus on process is a natural extension of classic protocol analysis. One example is provided in Blair and Burton (1987) who used a “retrospective protocol” that asked respondents “How did you come up with that answer?” to study the cognitive

¹The terminology in most fields that collect verbal reports labels the verbatim transcription the “protocol,” unlike survey cognitive interviewing where the term protocol is used to describe the set of procedures for conducting the cognitive interview.

strategies used to answer questions about frequency of occurrence of particular events. Their research tested hypotheses about the use of “episodic enumeration” (counting individual event occurrences) versus other strategies. They developed a coding frame of different cognitive processes, such as “simple episodic enumeration” or “simple rate-based estimation,” into which the verbal reports were coded for analysis. The coded data were used to test the posited hypotheses.

In the early development of cognitive interviewing, some survey researchers believed that by understanding the response process, one could identify question comprehension problems, recall difficulties and other factors that impair the accuracy of question answers (Sudman et al. 1996). This, again, was a natural application of the original method. It followed the use of the method in most other disciplines, where thinking aloud was used to generate data to understand a particular cognitive process or validate an information processing model. While early survey applications were concerned with the response process, the objective soon shifted to the goal of identifying when the response process (as posited in a response model) failed.

3. Cognitive Interview Pretesting

As the main use of cognitive interviewing shifted to pretesting questions, the interviewing techniques expanded far beyond eliciting think aloud reports. Paralleling the changes in interview procedures were changes in how interview results were interpreted, to the point that if we now take “methods for the analysis of cognitive interviews” to mean procedures used to determine when respondents’ verbal reports constitute evidence of a survey question problem, those procedures have become so varied that it is not surprising no generally-applicable analysis paradigm has emerged.

A cognitive interview analysis method can be nothing more than noting respondent comments during the interview, e.g. “I don’t know what ‘internist’ means.” Or analysis can employ trained coders review verbatim interview transcripts to identify word strings that, according to a coding frame, indicate a particular type of question problem. In the first example, the “analysis” occurs during the interview and requires nothing more than listening to the respondent; the second example is far more elaborate and costly to implement, but may be no more valid or reliable than the first. Moreover, the two approaches could coexist in the same pretest.

4. Application of Protocol Analysis in Survey Pretesting and Other Fields

The literature reviewed to date confirms that verbal protocol analysis, as formulated by Ericsson and Simon, has rarely been applied in cognitive interview pretests, though other somewhat similar methods have been— but such approaches have not been generally adopted by survey researchers. The few uses of the method in pretesting are informative.

We also discovered that a huge verbal protocol analysis (VPA) literature exists in several other disciplines— in some of which the question of the proper analysis of verbal reports has periodically surfaced. And following the original intent of VPA, the applications in other fields have most often been concerned with the study of some type of cognitive process, rather than matters akin to problem identification. Only usability testing seems to have used verbal reports (generated by respondents thinking aloud or otherwise verbalizing their experiences) mainly as a tool for problem identification.

We next:

1. Examine the application of formal, and quasi-VPA methods to pretesting;
2. Describe some issues of verbal reports analysis raised in other fields, and note possible lessons for pretesting; and
3. Consider a model-based method for evaluating and testing survey questions.

5. Methods of Analyzing Verbal Reports to Test Survey Questions

Bolton (1991) proposed a pretest methodology in the spirit of Ericsson and Simon in which she sought evidence from think aloud verbal reports of problems in one of the four stages of the common question response model (Tourangeau et al. 2000). The audio-taped interviews were transcribed and segmented (using syntactical and other types of markers). Coding categories were developed based on words and word strings, and content analysis was used to identify problems in any of the response model stages to compare original and revised versions of questions. This methodology produced informative data both to identify problems and provide direction for further revisions. However, Bolton notes that this approach, in addition to being time consuming and labor intensive, depends “heavily on the specification and measurement of theoretically-justified content characteristics,” and, in part for this very reason, “...cannot identify all types of defective questions.”

In a later paper, Bolton (1993) extended the methods and compared the content analysis approach to “observational monitoring,” a variant of the Cannell et al. (1977) method of using respondent-interviewer interactions to identify problematic questions. The main goal of Bolton’s project was to assess improvement from original to revised questionnaires. She concluded that the content analysis (which used automated coding) was the more effective method for identifying a range of types of question problems and also provided more detailed diagnostic information. Of course, the limitations of interaction behavior coding were known at that time and have been confirmed in the years since this early study. The content analysis approach permitted effective, if limited, quantitative analyses, but again at substantial cost. The huge advances in programming and computational power since that time certainly make content analysis more efficient, but there may still be a prohibitive difference in cost between this type of approach and methods that more directly observe question defects.

Knafl et al. (2007) provides a good example of recent efforts to systematically analyze cognitive interviews. The cognitive interviews in their study used a “verbal probing approach” in which respondents were asked to state their interpretation of each item rather than actually answer the questions. In contrast to Bolton, Knafl’s analysis approach was qualitative, but was similarly motivated to employ less subjective methods than are often used. After entering all the data from each respondent into an Access database, a matrix-display approach was used to construct item summaries, which were also linked to a set of respondent characteristics. A coding scheme was developed that reflected problem types specific to the Family Management Measurement items being tested. The item summaries, examined across respondents, were successfully used to make decisions about retention, deletion or revision of individual items. These examples illustrate from the first uses of cognitive interviewing to the present, researchers have sought more reliable analysis methods.

The main reasons for not adopting these types of methods seem to be partly practical limitations of time and resources, but also that users of cognitive interviewing to pretest questionnaires are satisfied with the widely-used, informal methods. Willis (2005), after providing a thorough review of these different approaches, seems to find little justification to pursue more formal analysis methods.

6. Issues of Verbal Report Analysis in Other Fields

Boren and Ramey(2000) have observed uses of verbal reports in usability testing that are very similar in some ways to what had occurred in survey pretesting. Their paper acknowledges that “usability professionals may contribute positively to a system’s development,” but then charge that “usability practices are not systematic or rigorous enough to merit the distinction of being called a method (a defined practice with clear rules for correct performance) [their bold type].” The particular usability practice they target is “thinking aloud.” They describe in detail many ways in which usability applications of the think aloud procedure differ from Ericsson and Simon’s work, even though that is the single most cited source of justification for the use of thinking aloud. As for actual applications, Boren and Ramey note that in their study of analysis methods note that “Only rarely did practitioners say they analyzed verbalizations closely, and then only for particularly problematic segments of action; in other words, there was no protocol analysis.” Boren and Ramey then propose an alternative theoretical basis more in tune with the way verbal reports are actually used in usability testing. It may be useful to explore similar solutions for survey cognitive interview practices.

Beyond implications for the validity of verbal reports collected by methods at variance with the Ericsson and Simon model, dispensing with protocol analysis has contributed to the use of an array of diverse practices. “Lacking the guidance of unifying principles, such changes [that they describe practitioners employing] vary greatly in degree and kind, not only from Ericsson and Simon’s model, but from each other.” This aspect of the usability experience, as described by Boren and Ramey, also bears a pronounced resemblance to survey cognitive interviewing: a menu of variations and deviations from Ericsson and Simon that, while having strong face validity, have become untethered from any underlying theory, and produced a set of practices differing from one researcher or project to another. Does this matter? If less formal procedures produce useful results at lower costs, there may be no need for concern. However, if there is evidence that those results are sometimes unreliable, it may be that some deviations from a theory-based methodology— in this instance a methodology for data analysis— have a hidden cost.

Chi (1997), motivated by the lack of a guide for the “analysis of verbal data more generally,” proposed her own guide to the analyses of verbal data, when the general goal is to understand representation of knowledge used in cognitive performances. While she notes that one way to develop such a guide would be, as we are planning, “...to survey the literature, identify all those studies that have used some kind of qualitative analysis of verbal data, then describe, analyze, and synthesize all the various methods,” she takes a different path. Her “verbal analysis” method is intended to quantify the qualitative coding of the contents of verbal reports. Chi’s guide is based on methods from her verbal analysis research at that time. Her approach is different from VPA in that it is less concerned with the processes of problem solving and more on representation of the knowledge that the problem solver has. This different perspective complements the VPA method in a way compatible with some cognitive interview objectives. (In more recent later work (Chi 2006), she also employs laboratory methods that are similar to some

cognitive laboratory techniques.) It is interesting that even a general purpose guide that employs formal coding and analysis methods has not, to our knowledge, found wide use.

7. Analysis of Cognitive Interviews Utilizing Subject-Specific Models and Coding Frames

In this section, we consider a method that uses general response models and tailored coding frames for the analysis of cognitive interviews. The relationship between models of the survey question response process (see Tourangeau et al. 2000 for a discussion and Jobe 2003 a review) and coding frames designed to enumerate response problems have been mainly, but not invariably, used for methodological research.

In one example, of many available, Conrad and Blair (1996) suggested a minor variation of the four-stage, general model: comprehension, recall, judgment and reporting. Their variation on the model allows for respondents proceeding from one stage to another with flawed or insufficient information (such as some amount of miscomprehension or incomplete recall), which may possibly affect response accuracy. In the subsequent stage, the respondent may become aware of the inadequate information, which can lead to:

- A. Proceeding even knowing about the information problem;
- B. Returning to the previous stage (e.g. trying to better understand the meaning of the question or trying to recall additional material) or
- C. Giving up and not producing an answer.

As shown in Blair and Burton's work, coding frames can be finely tuned to particular research topic. Blair et al. (1991) similarly developed a detailed coding frame for a study of self vs. proxy reporting. There are several other instances of such specialized coding frames for research.

Response models also have been adapted to particular populations. An excellent illustration is provided by Willimack and Nichols (2010) who report on a hybrid response model designed for business surveys. This model has guided the development of cognitive pretesting of establishment survey questionnaires at the U.S. Census Bureau. Frames that rely on response models (Bolton 1991) and those that do not (Knafl et al. 2007) have been developed to address types of problems expected in a particular instrument.

These examples suggest that analysis plans may be designed for particular populations, or otherwise tailored for specific types of surveys. We think that this is a direction that may improve analysis reliability, and address some of the issues that have mitigated against more formal analysis methods, even when those methods appear to be sound.

Presser and Blair (1994) elaborated on the four-stage response model to develop a detailed coding frame (Figure 1). To provide guidance to coders in that study, examples of problems were developed to illustrate every category in the coding frame.

| | |
|---|--|
| <p>I. Semantic:</p> | <p>Problems affecting how readily the question is understood or remembered due to:</p> <ul style="list-style-type: none"> A. Amount of information A. Structure/organization B. Flow or relation between questions |
| <p>II Semantic:</p> | <p>Problems affecting the way the question (or some part of it) is understood due to:</p> <ul style="list-style-type: none"> A. Boundary lines B. Insufficient knowledge <ul style="list-style-type: none"> 1. Technical term is not understood 2. Common term is not understood A. Multiple subjects |
| <p>III Respondent Task: Problems for the respondent retrieving information or formulating or reporting a response</p> <ul style="list-style-type: none"> A. Recall/Response Formation is: <ul style="list-style-type: none"> 1. Difficult 2. Impossible 3. Redundant 4. Resisted by respondent A. Report affected by <ul style="list-style-type: none"> 1. Overlapping response categories 2. Response categories being insufficient 3. Response categories making too fine a distinction 4. Response categories not appropriate to question 5. Sensitivity <p>For each of the code categories, they constructed an example of a flawed question and an example of how that flaw might be expressed in a verbal report from a cognitive interview. For example,</p> | |
| <p>II Semantic</p> <ul style="list-style-type: none"> A. Boundary lines <p>Question example: <u>How long have you lived in College Park?</u></p> <p>Report examples:</p> <p><i>"Well I have moved back and forth. Do you mean this most recent time?"</i></p> <p><i>"Is living in a dorm counted as living here, or only living in a house or apartment?"</i></p> | |

Figure 1: Problem Coding Frame

This procedure was developed for a questionnaire that included a wide range of topics, but the same approach could be used to adapt the frame to a particular topic. A frame with this kind of paired Question–Verbal Report examples for each code category could just as well be subject matter specific. For a specific survey, the examples of question flaws and reports could all be on the survey topic, e.g. health, education, travel etc.

Subject can be broadly specified, as in health surveys, or more narrowly, as in health surveys of the elderly or dietary behaviors. Ongoing tailoring of such frames could be along whatever dimensions (e.g. mode of administration) and detail that is useful. The first key for this approach to be effective is repeated use.

There are several possible advantages to taking this direction:

First, a tailored frame is likely to be more effective in helping coders to identify problems of particular interest.

Second, if repeated surveys on a topic are conducted, the frame can be revised and improved over time and its development costs spread across multiple projects. There would also be savings in start-up time and training.

Third, an investment in software to automate part of the problem identification process may be justified.

The second key to effectiveness would be shared or pooled information. If researchers working in a particular area or with a particular population contributed real examples of questions, question problems and verbal reports that revealed the problems, the development of such frames as general purpose analysis tools would be enhanced.

One possibility is that a resource such as Q-Bank — that has already compiled and classified a large database of survey questions and linked each question to its pretest findings— could be used to develop some subject-matter specific frames to be used in pilot studies.

The frame does not restrict the type of analysis to be done. Informal or qualitative methods, content analysis or verbal protocol analysis all benefit from reducing the cost of coding and improving coding accuracy. The frame does not automatically increase coding reliability, but is an essential part of that effort.

8. Summary

This approach, which we do not claim to be entirely new, might begin— and only begin— to address some of the problems that have prevented some effective methods from being more widely used:

- The need to reduce the large investment in time and labor to create a coding frame and coding rules for each survey.
- The need to accommodate a range of types of cognitive interview techniques and the different kinds of verbal reports they elicit

It may be that for many researchers the cost of more formal methods will still be too high. Some practitioners have pointed out that their, and their clients', goals do not require such thorough methods or even identifying most of the problems in an instrument. This is certainly true. And this proposal will not address their needs.

Reducing the cost of coding would permit trying different analysis methods that start with coding. This is one part of determining which methods produce more reliable (and replicable) results. It is yet to be seen whether or not the reconsideration of some of the more formal methods, and their theoretical bases, would strengthen cognitive interviewing practice.

This is very similar to a value of information problem in sample design. Sudman (1976) began his book on survey sampling with the question “How good does the sample need to be?” going on to state that “there is no uniform standard of quality that must always be reached by every sample.” Further on, puts the same matter more specifically. Considering the “decreasing marginal utility of additional information, “...there comes a time when additional information is not worth the cost of obtaining it...The value of information to its user depends on how likely the information is to influence an action or decision.” This reasoning seems quite applicable to investments in pretesting generally, and analysis in particular. The redesigning, for example, of a large ongoing federal survey has a greater need for very thorough problem identification than many, probably most, other surveys do not. It is not necessary to develop an analysis method that can benefit the entire range of survey needs and resources. It is an open question how widely applicable and efficient the use of survey-specific frames based on general response models may be.

References

- Beatty, P., and G. Willis. 2007. The Practice of Cognitive Interviewing. In *Public Opinion Quarterly*, Vol. 71 No. 2, 287-311.
- Blair, E., and S. Burton. 1987. Cognitive Processes Used by Survey Respondents to Answer Behavioral Frequency Questions. In *Journal of Consumer Research*, Vol. 14
- Blair, J., G. Menon, and B. Bickart. 1991. Measurement effects in self versus proxy response to survey questions: an information processing perspective. In P. Biemer et al. (eds) *Measurement Error in Surveys*, Wiley, New York
- Bolton. 1991. An Exploratory Investigation of Questionnaire Pretesting with Verbal Protocol Analysis. In *Advances in Consumer Research*, Vol. 18.
- Bolton. 1993. Pretesting Questionnaires: content analyses of respondents' concurrent verbal protocols. In *Marketing Science*, Vol. 12 No. 3.
- Boren and Ramey. 2000. Thinking Aloud: Reconciling Theory and Practice. In *IEEE Transactions on Professional Communication*. Vol. 43 No. 3.
- Cannell, C., L. Oksenberg, and J. Converse. 1977. Experiments in interviewing techniques. National Center for Health Services Research, Hyattsville MD.
- Charness, N. 1981. Sear in Chess: Age and Skill Differences. In *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 7, No. 2, 476-76
- Chi, M. 1997. Quantifying Qualitative Analyses of Verbal Data: A Practical Guide. In *The Journal of the Learning Sciences*, 6(3).
- Chi, M. 2006. Laboratory Methods for Assess Experts' and Novices' Knowledge. {working paper for book chapter}
- Conrad and Blair. 1996. From Impressions to Data: Increasing the Objectivity of Cognitive Interviews. In *JSM Proceedings*, Section on Survey Research Methods. Alexandria, VA: American Statistical Association.
- Conrad and Blair. 2009. Sources of Error in Cognitive Interviews. In *Public Opinion Quarterly* 73(2).
- Ericsson, A. 2006. Protocol Analysis and Expert Thought: Concurrent Verbalizations of Thinking during Experts' Performance of Representative Tasks. in K. A. Ericsson, N. Charness, P. Feltovich, and R. Hoffman, *The Cambridge Handbook of Expertise and Expert Performance*, Cambridge: Cambridge University Press.
- Ericsson, A., and H. Simon. 1993. Protocol Analysis: Verbal Reports as Data, The MIT Press, Cambridge
- Jobe, J. 2003. Cognitive psychology and self-reports: Models and methods. In *Quality of Life Research*, 12.

- Knafl et al. 2007. The Analysis and Interpretation of Cognitive Interviews for Instrument Development. In *Research in Nursing & Health*, 30.
- Presser, S., and J. Blair. 1994. Survey Pretesting: Do Different Methods Produce Different Results? In *Sociological Methodology*, Vol. 24, 73-104.
- Sudman, S. 1976. *Applied Sampling*, Academic Press Inc. New York
- Sudman, S., N.M. Bradburn, and N. Schwarz. 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*, Jossey-Bass San Francisco
- Tourangeau, R.L., and K. Rasinski. 2000. *The Psychology of Survey Response*, Cambridge University Press
- Willimack, D.K., and E. Nichols. 2010. A Hybrid Response Process Model for Business Surveys. In *Journal of Official Statistics*, Vol. 26, No. 1, 3-24
- Willis, G. 2005. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*, Sage Thousand Oaks CA.