

TEST OF AN APPRAISAL METHOD FOR ESTABLISHMENT SURVEY QUESTIONNAIRES¹

Barbara H. Forsyth, Kerry Levin and Sylvia K. Fisher, Westat
Barbara H. Forsyth, Westat, 1650 Research Blvd Rockville, MD 20850

Key Words: Questionnaire Design, Establishment Survey, Cognitive Laboratory Methods

This paper presents a questionnaire appraisal tool designed to facilitate the review and revision of establishment survey questionnaires. The appraisal method presented here builds on earlier work at Westat and Research Triangle Institute, developing an appraisal for household survey questionnaires. Our goal was to extend the appraisal to establishment and organizational surveys.

The appraisal method was developed to standardize expert questionnaire review. Under the general appraisal method (e.g., Lessler & Forsyth, 1996), questionnaires are reviewed item-by-item. The appraisal system consists of a set of codes that describe question features. Item review involves selecting appraisal codes that identify question design features likely to contribute to response error.

The primary goals of the appraisal are to document problems in question design and identify possible revisions. The review criteria are standard and so, the appraisal offers the promise of increased consistency in questionnaire review and revision. We have found that the household appraisal is a quick and inexpensive source of feedback when more extensive cognitive laboratory testing activities are not feasible. Also, the household survey appraisal has been useful as a first step in developing cognitive research priorities and cognitive testing plans. Successes with the household appraisal method led us to extend the method to organizational survey questionnaires.

It is useful to start with a general model of survey response to assist in developing codes for the organizational appraisal system. We selected a framework by drawing on findings from a variety of researchers, including Edwards and Cantor (1991), Biemer and Fecso (1995), and Tomaskovic-Devey et al. (1995), among others. Following most survey item response models (e.g., Tourangeau, 1984), we assume that question **comprehension** is an early step,

followed by **information retrieval, judgment** and **response selection**.

Our assumptions about **information retrieval** emphasize the role that organization respondents play as managers of information. In establishment and organizational surveys, retrieval involves:

- Identifying appropriate sources of information – these may include the respondents' own memory, organizational record systems, and other people within the organization.
- Selecting strategies for retrieving information from identified sources.
- Assessing the match between information retrieved and information needed to select a response.

Respondents may repeat some or all of these retrieval processes until the match seems sufficient.

In the **judgment** step, the establishment respondent's task is to synthesize or integrate information coming from a variety of sources. Depending on the sources consulted, respondents may need to synthesize information from different individuals within the organization, or they may need to synthesize information from memory with information from administrative records. We expect that in organizational surveys, **response selection** will be influenced by strategic, legal, and economic factors as well as more traditional factors related to desirability and sensitivity.

This framework is useful for identifying potential sources of measurement error that might be eliminated or reduced by question redesign. Table 1 presents the organizational survey appraisal system. The appraisal codes in Table 1 are organized into four general categories: codes identifying problems related to **comprehension**; codes identifying problems related to **information retrieval**; codes identifying problems

¹ Eileen O'Brien and Betsy Bury provided valuable advice on the development and testing research reported here.

related to **synthesis and judgment**, and codes identifying problems related to **response selection**.

Under **comprehension**, there are six subsets of codes. They document problems with *Instruction Content*, with *Navigational Instructions*, with general *Question Content*, with specific *Question Terminology*, with *Question Structure*, and with item *Reference Periods*. For example, under *Instruction Content*, there's a code to identify items with **conflicting instructions** -- instructions that are probably difficult to understand because they are internally inconsistent. Under *Question Terminology*, there is a code to document items that use **ambiguous or vague terms** -- terms that may be interpreted differently by different respondents.

Under the general heading of **information retrieval**, there are four sets of problem codes to document problems related to *Organization Characteristics*, to *Source Identification*, to *Memory Retrieval*, and to *Record Retrieval*. For example, under *Organization Characteristics*, there's a code for **distributed knowledge likely** to document items that ask for information that is likely to require retrieval from several different units or people within an organization. Under *Record Retrieval*, the code for **authority issues** identifies items requesting information that is likely to require review and approval at relatively high levels within the organization.

Under the general heading of **synthesis and judgment**, there are three sets of codes for problems due to the *Match between Record and Item Content*, to *Judgment Processes*, and to *Judgment Task Characteristics*. Under the general heading of **response selection** there are three sets of codes for problems with *Response Terminology*, with the *Response Units*, and with *Response Structure*. For example, under *Response Structure*, there are codes to identify response sets that include **overlapping categories** and response sets with **missing categories**. Both of these features make response selection more difficult and more error-prone.

The appraisal codes in Table 1 provide a detailed item-level analysis. The establishment appraisal is a relatively new development, and the specific codes included are likely to change as we gain experience using the system. We will use an example to illustrate how the system is used. In this example, we used the new appraisal system to evaluate draft items from a business survey questionnaire that Westat is

redesigning for a Federal Agency. The agency funds a program that provides consulting assistance to some businesses, and the survey is designed to assist the agency with program evaluation. The survey is conducted by telephone. Paper copies of the questionnaire are mailed to identified respondents to help them prepare for the telephone interview.

This paper focuses on one draft item. It asks, "Did the assistance you received lead to any changes at your company?" Table 2 shows the appraisal codes assigned to this item. Consider the **comprehension** codes. This item is the first item on the questionnaire. The old form provided no instructions for respondents. We selected the *instruction code* labeled **transition needed** to indicate that orienting instructions printed on the form would make it easier to understand the form and the items on it.

In practice, the paper form is mailed with a cover letter explaining the purposes of the survey. The cover letter could provide the missing orienting instructions, but there's room for comprehension error if the letter is separated from the form and unavailable to the respondent. We used the *instruction code* labeled **provide information on finding details** because we'd expect fewer comprehension errors if the form included an instruction to refer to the cover letter. We selected the code for **under specified topic** because the item does not identify the particular services covered by the item. For example, businesses that use the program more than once may have trouble determining which services to include in their responses. It is also possible for a consultant to work with several different people at a client business. In this case, there is a risk that respondents will report on too narrow a set of services because they are not familiar with all of the services received by their organization.

We assigned two additional codes related to **Comprehension**. "Any changes" is vague wording, and so we selected the code for **ambiguous or vague terms**. The kinds of changes that respondents think about may differ across respondents, possibly depending on the unit the respondent works in or the respondent's role in the business. We also selected the code for **undefined reference period** because the item does not specify a time frame, and respondents are likely to choose different time periods to report on.

Under **Information Retrieval**, we selected the code labeled **distributed knowledge likely** because the information needed to answer this item may be

distributed across units within the client business. We used the code labeled *provide instructions to identify source(s)* to indicate that error might be reduced by providing assistance to help respondents identify useful sources of information. Tailored cover letters or well-trained interviewers might be able to use information provided by the program consultants to remind respondents about other sources in the business that may have information to contribute to the evaluation.

We selected the code labeled *shortage of memory cues* to indicate that the item provides few cues to assist recall. For example, reminders about the kind of assistance provided might enhance recall for the changes a business experienced because of the consulting assistance.

We have already noted that this question about “any changes” is very general. We used the code for **records unavailable or don’t support estimation** to indicate that it’s unlikely businesses will keep the kinds of records needed to answer a question at this level of generality. To the extent that records are relevant, it seems likely that respondents will have to synthesize or summarize record information in order to answer this item. We selected none of the codes under **Synthesis and Judgment** and none under **Response Selection**. While the item’s generality makes it prone to comprehension errors, the generality seems to ease the tasks involved in judgment and response selection.

From the detailed description we’ve given for this item, it would be easy to infer that there are lots of problems with it. It’s important to note that several of the problems identified could be fixed with a well-crafted set of instructions and a well-defined reference period.

We will close this paper by highlighting a few benefits of the appraisal methodology and by suggesting some directions for future research. The establishment appraisal has three key benefits. First, the appraisal is relatively easy to implement. Second, it provides quick feedback about draft questionnaire items. Third, the standardized format means that expert reviews address a constant set of design principles -- both across questionnaires and across reviewers. In addition, as the example illustrated, the focus on cognitive response processes makes it easy to use the appraisal codes to identify useful revisions. In large-scale projects, an initial appraisal can be one resource for informing cognitive testing designs. In smaller

projects where extensive testing is not feasible, the questionnaire appraisal may be a useful alternative.

We have several goals for continued research. One of the first is to explore whether we can train relatively junior staff to conduct the appraisal coding. If junior-level staff can be trained to use the appraisal consistently, then it will especially useful as a cost-effective method for providing feedback and identifying possible revisions. Another goal for future research is to explore similarities and differences between appraisal results and results from other cognitive laboratory methods. It will be interesting to see whether our ability to substitute the appraisal for other cognitive methods differs for household and organizational surveys. It may be risky to use appraisal methods as a substitute for cognitive interviewing in organizational surveys because response processes are influenced by factors related to content and organizational structure. In these areas, organizational respondents have special expertise. Interviews with these experts may be a critical component in an effective questionnaire evaluation design.

Table 1. Organizational Survey Appraisal System

Comprehension		Information Retrieval	Synthesis and Evaluation	Response Selection
Instruction Content	Question Terminology	Organization Characteristics	Match: Record and Item	Response Terminology
Conflicting instructions	Critical definition(s) missing	Distributed knowledge likely (or multiple sources)	Incompatible with regulatory requirements	Critical definition(s) missing
Inaccurate instructions	Add or add to examples	Seasonal or periodic trends	Incompatible with organizational or unit objectives	Vague term(s)
Hidden instructions	Ambiguous or vague term(s)	Source Identification	Survey-specific system unlikely (e.g., panel versus one-time survey)	Mismatch to technical language
Complicated content	Multiple definitions		Variability in recorded units	Industry-specific terminology
Complex syntax	Mismatch to technical language	Provide assistance to help identify source(s)	Incompatible time frames	Response Units
Separate from item	Industry-specific terminology	Sources may not be accessible (for survey purposes)		Response unit mismatch to organizational units
Nearby but not embedded in item	Question Structure	Memory Retrieval	Judgment Processes	Responses use wrong units
Instructions provided too late	Hidden question	Non-routine summary or breakdown required	Coordination or collaboration	Response Structure
Unclear examples	Complex syntax	Shortage of (memory) cues	Guessing or estimation likely	
Unclear layout	Implicit assumption	Detail problem/item specificity	Task Characteristics	
Transition needed	Several questions	Unanchored reference period	Non-routine time frame	
Provide info on finding details	Unclear goal	Rolling reference period	Complex estimation	
Navigational Instructions	Q/A mismatch	Record Retrieval	Potentially sensitive	
Inaccurate instructions (move to wrong place)	No question	Records unavailable or don't support estimation	Social desirability	
Confusing convention, flow or typographic	Reference Period	Record access issues	Proprietary information	
Complex information	Carry-over reference period	Authority issues	Strategic factors	
Not salient	Undefined reference period		Timing issues	
Question Content	Embedded reference period			
Complex topic	Abrupt change			
Under specified topic	Problematic length			
Topic carried over				
Assumes consistent behavior				
Provide assistance to identify source(s)				

Table 2. Appraisal codes assigned to example item

COMPREHENSION CODES	
Instruction Content	Question Terminology
Transition needed	Ambiguous or vague term(s)
Provide information on finding detail	Reference Period
Question Content	Undefined reference period
Under specified topic	
INFORMATION RETRIEVAL CODES	
Organization Characteristics	Memory Retrieval
Distributed knowledge likely (multiple sources)	Shortage of cues
Source Identification	Record Retrieval
Provide instructions to help identify source	Records unavailable or don't support estimation
SYNTHESIS AND JUDGMENT	
RESPONSE SELECTION	