Cognitive Interviewing

A "How To" Guide

Developed by:
Gordon B. Willis
Research Triangle Institute

Reducing Survey Error through Research on the Cognitive and Decision Processes in Surveys
Short course presented at the
1999 Meeting of the American Statistical Association
Rachel A Caspar, Judith T. Lessler, and Gordon B. Willis--Research Triangle Institute

Contents

Preface ii

1. Background: Cognitive Interviewing Techniques 1

2. Cognitive Theory 2

3. Cognitive Interviewing Methods: Think-Aloud and Verbal Probing 3

4. Examples from Cognitive Interviewing Studies 9

5. Detection of Structural Problems in Survey Questions 20

6. The Sequence of Cognitive Interviewing Activities 21

7. Practical Aspects of Cognitive Interviewing 22

8. Interviewing Logistics 25

9. Evaluation of Cognitive Interviewing Techniques 33

References 36

PREFACE

This guide is based on the document "Cognitive Interviewing and Questionnaire Design: A Training Manual," by Gordon Willis (Working Paper #7, National Center for Health Statistics, March 1994). In revised form, this document describes the cognitive interviewing techniques appropriate for questionnaire development and testing, and which are used by the staff of Research Triangle Institute (Laboratory for Survey Methods and Measurement, Research Triangle Park, NC; Cognitive Research Laboratory, Rockville, MD). Although there are several cognitive laboratories currently in operation that may utilize various procedures, the specific methods described were adopted from those used at the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention, where the author previously worked as a senior staff member in the Questionnaire Design Research Laboratory.

1. BACKGROUND: COGNITIVE INTERVIEWING TECHNIQUES

The cognitive interviewing approach to evaluating sources of response error in survey questionnaires was developed during the 1980's through an interdisciplinary effort by survey methodologists and psychologists. Some general features of this approach are as follows:

a) In the form presented here, it focuses mainly on the questionnaire, rather than on the entire survey administration process (the focus is on survey *questions*, as opposed to administration procedures such as Computer Administered Personal Interviewing).

b) It explicitly focuses on the cognitive processes that respondents use to answer survey questions; therefore, covert processes that are normally hidden, as well as overt, observable ones, are studied.

c) For the conduct of the cognitive interview, volunteer subjects are recruited, and are interviewed either in a laboratory environment, or in some other private location (in this application, the term "subject"' refers to an individual who is tested through a cognitive interviewing procedure, and "respondent" defines someone who is interviewed in a fielded survey).

d) The recruitment of subjects targets persons with specific characteristics of interest (for example, the elderly, those who have used illicit drugs in the past 12 months, teenagers who have used chewing tobacco, etc.).

The cognitive approach to the design of questionnaires has generated a body of methodological research (see Campanelli, 1997; Campanelli, Martin, and Rothgeb, 1991; DeMaio and Rothgeb, 1996; Dippo, 1989; Esposito, and Hess, 1992; Jabine, Straf, Tanur, and Tourangeau, 1984; Jobe and Mingay, 1991, Jobe, Tourangeau, and Smith, 1993; Lessler and Sirken, 1985; Royston, Bercini, Sirken, and Mingay, 1986; Sirken, Herrmann, Schechter, Schwarz, Tanur, and Tourangeau, 1999; Willis, DeMaio, and Harris-Kojetin, 1999; Willis and Schechter, 1997). Several Federal statistical agencies, as well as some private survey research organizations, including Research Triangle Institute (RTI), now routinely carry out cognitive interviewing activities on a wide variety of survey questionnaires.

2. COGNITIVE THEORY

The background theory underlying cognitive interviewing has been represented by various models (see Jobe and Herrmann, 1996). The most general model is attributable to Tourangeau (1984), and in brief, consists of the following processes.

1) COMPREHENSION OF THE QUESTION:

a) *Question intent*: What does the respondent believe the question to be asking?

b) *Meaning of terms*: What do specific words and phrases in the question mean to the respondent?

2) RETRIEVAL FROM MEMORY OF RELEVANT INFORMATION:

a) *Recallability of information*: What types of information does the respondent need to recall in order to answer the question?

b) *Recall strategy*: What type of strategies are used to retrieve information? For example, does the respondent tend to count events by recalling each one individually, or does he/she use an estimation strategy?

3) DECISION PROCESSES:

a) *Motivation*: Does the respondent devote sufficient mental effort to answer the question accurately and thoughtfully?

b) *Sensitivity/Social Desirability*: Does the respondent want to tell the truth? Does he/she say something that makes him/her look "better"?

4) RESPONSE PROCESSES:

Mapping the response: Can the respondent match his or her internally generated answer to the response categories given by the survey question?

For survey questions that are non-trivial, the question-answering process may be complex, and involve a number of cognitive steps. Some of these processes may be "conscious", but some are automatic, so that the respondent is not aware of their operation. The cognitive processes used to answer survey questions may also vary, depending on the type of question asked.

Autobiographical questions may place a heavy burden on retrieval processes; asking questions that are sensitive (for example; "Have you ever smoked marijuana?"), may place more demands on the respondent's decision processes.

Survey researchers who apply cognitive interviewing techniques recognize that they cannot know in an absolute sense what transpires in a respondent's mind as he or she answers a survey question. Rather, the cognitive interviewer's goal is to prompt the individual to reveal information that provides clues as to the types of processes mentioned above. The manner in which one may go about this is discussed next.

3. COGNITIVE INTERVIEWING METHODS: THINK-ALOUD AND VERBAL PROBING

There are two major sub-types of cognitive interviewing methods, referred to as think-aloud interviewing, and verbal probing techniques[1]. These are described in turn.

A) "Think-aloud" interviewing

The think-aloud interview derives from psychological procedures described by Ericsson and Simon (1980). Consistent with recent practice (see Willis, et al., 1999), the term think-aloud is used here to describe a very specific type of activity, in which subjects are explicitly instructed to "think aloud" as they answer the survey questions. The interviewer reads each question to the subject, and then records and/or otherwise notes the processes that subject uses in arriving at an answer to the question. The interviewer interjects little else, except to say "tell me what you're thinking" when the subject pauses. For example, a portion of a think-aloud interview might consist of the following:

> INTERVIEWER (reading survey question to be tested): How many times have you talked to a doctor in the last 12 months?
>
> SUBJECT: I guess that depends on what you mean when you say "talked." I talk to my neighbor, who is a doctor, but you probably don't mean that. I go to my doctor about once a year, for a general check-up, so I would count that one. I've also probably been to some type of specialist a couple of more times in the past year - once to get a bad knee diagnosed, and I also saw an ENT about a chronic coughing thing, which I'm pretty sure was in the past year, although I wouldn't swear to it. I've also talked to doctors several times when I brought my kids in to the pediatrician - I might assume that you don't want that included, although I really can't be sure. Also, I saw a chiropractor, but I don't know if you'd consider that to be a doctor in the sense you mean. So, what I'm saying, overall,

[1]This document covers the major techniques used, rather than the full range. For a comprehensive taxonomy of procedures, see Forsyth and Lessler (1991).

> is that I guess I'm not sure what number to give you, mostly because I don't know what you want.

From this "think-aloud protocol," the interviewer may observe that the individual attempts to answer this question by attempting to recall each visit individually, rather than by estimating. It might be concluded that the individual has trouble determining whether a visit was really in the last 12 months. If, after interviewing several subjects, it becomes clear that none could really "think through" with confidence the number of times they had been to a doctor, one might decide that the reference period is simply too long to provide adequate answers. More significantly, the larger problem here seems to be that the subject is clearly unsure about what is to be included and excluded from the question, as far as both a) whether this refers only to doctor contacts that pertain to his/her health, and b) the type of physician or other provider that is to be counted.

Training the subject to perform a 'think-aloud' interview:

The interviewer must teach the subject how to perform the think-aloud procedure. This training generally involves careful practice at the start of an interview. One training approach that has may work is the following:

> "Try to visualize the place where you live, and think about how many windows there are in that place. As you count up the windows, tell me what you are seeing and thinking about."

Depending on how well the subject responds to this exercise, further training may be necessary, prior to beginning the core part of the interview.

Advantages of the think-aloud technique:

a) *Freedom from interviewer-imposed bias*: Because the interviewer contributes little other than the reading of the survey question, except to occasionally ask what the subject is thinking, he or she interjects little that may serve to bias the subject's responses.

b) *Minimal interviewer training requirements*: Again, because the interviewer mainly reads survey questions, and then listens to the respondent talk, little training or special expertise is usually necessary.

c) *Open-ended format*: Because the subject's verbalization is guided only minimally, he or she may provide information that is unanticipated by the interviewer. Therefore, think-aloud interviewing is especially valuable when the subject is outgoing, articulate, and has had significant experience with the topics covered by the survey questions.

Disadvantages of the think-aloud technique:

a) *Need for subject training*: Because thinking-aloud is somewhat unusual for most people, the technique typically requires a non-trivial amount of preliminary training of lab subjects, in order to elicit a sufficient amount of think-aloud behavior. Such training may eat into the amount of productive time that can be devoted to the interview.

b) *Subject resistance*: Even given training in the activity, many individuals are not proficient at the think-aloud activity. In particular, they tend to simply answer the questions that are asked, without further elaboration.

c) *Burden on subject*: Related to the point above, the think-aloud activity places the main burden on the subject. The alternative, as described next, is to place more of the relative burden on the cognitive interviewer.

d) *Tendency for the subject to stray from the task*: Under think-aloud, the subject controls the nature of much of the elaborative discussion. Therefore, it is very easy for a "free associating" subject to wander completely off-track, and to spend a significant amount of time on one question, often delving into irrelevant areas, so that the interviewer must struggle to "bring the subject back." In general, the think-aloud technique results in relatively few survey questions being tested within a particular amount of time, relative to alternative approaches (again, see the discussion that follows).

e) *Bias in subject information processing*: By its nature, thinking-aloud forces subjects to think. As such, subjects may invest a considerable amount of mental effort into processing the survey questions, relative to what they do when simply answering the questions. Thinking-aloud typically entails more intensive effort, and more justification of each answer, than when one simply provides an answer such as "yes," "no," or "I agree." Therefore, it is very possible that the activities associated with think-aloud speech may serve to burden or contaminate the cognitive processes used in answering the question. This issue is clearly still open to debate, as there are no direct physiological measures, from either the cognitive interview or the usual survey interview, of the quantitative amount of information processing that is typically involved in answering survey questions.

B. The use of Verbal Probing techniques

As an alternative to the think-aloud, the use of verbal probing is the basic technique that has increasingly come into favor by cognitive researchers (see Willis, et al., 1999). After the interviewer asks the survey question, and the subject answers, the interviewer then asks for other, specific information relevant to the question, or to the specific answer given. In general, the interviewer "probes" further into the basis for the response. The following table contains basic categories of cognitive probes, and an example of each:

Comprehension/ Interpretation probe:	What does the term "outpatient" mean to you?
Paraphrasing[2]:	Can you repeat the question I just asked in your own words?
Confidence judgment:	How sure are you that your health insurance covers drug and alcohol treatment?
Recall probe:	How do you remember that you went to the doctor five times in the past 12 months?
Specific probe:	Why do you think that cancer is the most serious health problem?
General probes:	How did you arrive at that answer? Was that easy or hard to answer? I noticed that you hesitated - tell me what you were thinking[3]

Advantages of the Verbal Probing technique:

a) *Control of the interview.* The use of targeted probing to guide the subject tailors the interchange in a way that is controlled mainly by the interviewer. This practice avoids a good deal of discussion that may be irrelevant and non-productive. Further, the interviewer can focus on particular areas that appear to be relevant as potential sources of response error.

b) *Ease of training of the subject.* It is fairly easy to induce subjects to answer probe questions, as these probes often do not differ fundamentally from the survey question they are otherwise answering. In fact, subjects will sometimes begin to expect probes, and to offer their own spontaneous thoughts and critiques, so that the interview comes to

[2]Paraphrasing has been classified by other authors as a specific type of cognitive method, apart from cognitive interviewing (see Forsyth and Lessler, 1991), whereas this guide categorizes paraphrasing as a sub-type of verbal probing. Note that in practice, to the degree that one chooses to simply make use of each method as appropriate, such nomenclature differences have few serious implications, as far as how interviews are conducted.

[3]Note that the probe "tell me what you were thinking" is virtually identical to the general practice sometimes used in think-aloud interviewing to elicit responding. From this perspective, to the extent that the interviewer uses this type of probe when conducting a think-aloud, the think-aloud procedure can be conceptualized as a specialized form of verbal probing.

resemble a think-aloud.

Disadvantages of probing techniques:

a) *Artificiality*. Occasionally, the criticism is made that the validity of verbal probing techniques is suspect, because the interjection of probes by interviewers may produce a situation that is not a meaningful analog to the usual survey interview, in which the interviewer simply administers questions, and the respondent answers them. However, note that the verbal probing technique is certainly no more unrealistic than the alternative of thinking-aloud. Further, this criticism may also not be particularly relevant; the basic purpose of the pretest cognitive interview is very different than that of the fielded interview (the former analyzes questions, the latter collects data). Alternatively, one might consider making use of retrospective probing (see below).

b) *Potential for Bias*. A related criticism is that the use of probes may lead the respondent to particular types of responses. This is of course possible, but can be minimized through the careful selection of "non-leading" probing techniques that minimize bias. For example, in conducting probing, rather than suggesting to the subject one possibility ("Did you think the question was asking just about physicians?"), it is preferable to list all reasonable possibilities ("Did you think the question was asking only about physicians, or about any type of health professional?"). In other words, probes should be characterized by unbiased phrasing, in the same manner that survey questions are intended to.

Concurrent versus retrospective probing:

The two general approaches to probing are: a) *concurrent probing*, and b) *retrospective probing*. With concurrent probing, the interchange is characterized by: a) the interviewer asking the survey question, b) the subject answering the question, c) the interviewer asking a probe question, d) the subject answering the probe question, and e) possibly, further cycles of (c-d). In retrospective probing, on the other hand, the subject is asked the probe questions after the entire interview has been administered (sometimes in a separate part of the interview known as a "debriefing session"). Overall, it appears that concurrent probing is more frequently used at present, mainly because the information to be asked about is still fresh in the subject's mind at the time of the probing. It may seem more realistic to wait and to debrief the subject by probing after the questions have been administered (in order to avoid the potential for bias mentioned above). However, there is then a significant danger that subjects may no longer remember what they were thinking as they answered a question, and will instead fabricate an explanation.

Retrospective probing can be very useful, however, under certain circumstances:

a) *When testing self-administered questionnaires*. Retrospective probing is useful when the purpose of testing is mainly to determine the subject's ability to complete the

instrument unaided, and especially to follow sequencing instructions.

b) *In later stages of questionnaire development*. When a questionnaire is in latter stages of development, and one wants to simulate a more "realistic" type of presentation, it makes sense to administer the questionnaire "straight," and to then conduct probing afterward.

How are the specific probes developed?

Whether probing is done concurrently or retrospectively, there are two basic categories of probe questions:

a) *Scripted probes*: For use by all interviewers-- these are developed prior to the interview.

b) *Spontaneous probes*: Used by a particular interviewer-- these are usually “thought up” during the interview.

Scripted probes are meant for use by all interviewers who will be conducting interviews, and are developed before interviewing commences by either a questionnaire development group or by a lead individual. For example, if it is anticipated that a particular term may not be universally understood, all interviewers can be instructed to apply the probe: "What does (TERM) mean to you?" These probes are often typed directly into the questionnaire draft.

Scripted probes are practical and useful when:

a) There is sufficient time to prepare for interviews.

b) Resources exist to plan and execute a fairly standardized testing approach.

c) Some interviewers are relatively inexperienced and would benefit from the guidance provided by a structured protocol.

<u>Choice of scripted versus spontaneous probes</u>. Admittedly, the "spontaneous" approach to probing appears to be somewhat unscientific or haphazard, especially because there is no coordination of probing across interviewers. However, there are particular advantages to this approach. In particular, the most interesting and productive forms of probing often develop through the course of the interview, as a product of the particular relationship between the interviewer, subject, and survey questionnaire. One of the key underlying assumptions of the cognitive interviewing approach is that these developments often cannot be anticipated in advance of the interview. Over time, interviewers become very proficient in using spontaneous probing. Further, the subject’s answer to a particular probe may well lead the interviewer to use other probes, and to follow-up on the issues that emerge as the most interesting and important.

Combination of probing types. The most effective interviews may consist of a combination of scripted and spontaneous probes described above, rather than either type by itself. By way of analogy, a cognitive interview is similar to a session with a clinical psychologist; the "therapist" has certain guiding principles, and perhaps specific questions or comments, to apply during a session with the patient. However, much of the interchange emerges spontaneously during the course of therapy. The clinical session may be approached in ways similar to other sessions, and be somewhat "scripted", but every interview is different, entails its own developmental sequence, and makes a unique contribution as far as the “diagnosis” of problems.

For the remainder of this manual, probing rather than the strict think-aloud procedure is emphasized. However, practitioners of cognitive interviewing techniques often mix these techniques into the same interview. In fact, procedural flexibility, as opposed to rigid adherence to one dogmatic approach, is often viewed as one of the most attractive features of the cognitive interviewing approach.

4. EXAMPLES FROM COGNITIVE INTERVIEWING

In order to better illustrate the above discussion of cognitive techniques, and the use of verbal probing in particular, a list of examples of survey questions that have been tested using verbal probing techniques is presented below[4]. Each example consists of:

1) The question in its original form.

2) A list of several probes that would be appropriate to use in testing that question.

3) A short description of the problems found, through cognitive testing of these questions, using probes of the types suggested. Each of the examples is classed generally according to whether the problems found are representative of the cognitive categories defined earlier. However, some questions may have more than one type of problem, and in some cases it is arguable what class of problem is really being reflected. This type of classification ambiguity may not be problematic, to the extent that the nature of the specific problem (and perhaps its resolution) is clear.

4) Finally, a suggested resolution to the problem is presented, based on the testing results.

[4]These questions were developed during the time the author worked in the Questionnaire Design Research Laboratory at the National Center for Health Statistics, CDC, in Hyattsville, MD (see Willis, 1994). The tested questions were mainly intended for use in the National Health Interview Survey (NHIS), a household-interview-based health survey conducted annually by NCHS.

EXAMPLE 1:

1) Original form of survey question:

Has anyone in the household ever received vocational rehabilitation services from-

... The State Vocational Rehabilitation program?

... another vocational rehabilitation program?

---

2) Probes:

a) **Can you repeat the question in your own words?**
(To test how well the subject comprehends the question.)

b) **What, to you, is a "vocational rehabilitation program"?**
(To test comprehension of a particular term.)

c) **How sure are you that (person) got this type of service?**
(To determine the subject's ability to recall information confidently.)

3) Results:

Comprehension problems: Subjects found it difficult to understand the question, because of its length and technical nature. Further, the meaning of "vocational rehabilitation" was not at all clear; some subjects thought this just meant any type of physical therapy. Because of the comprehension problems in the original form, we suggested the following change:

4) Suggested revision:

Has anyone in the household ever received job rehabilitation services?

If YES, ask WHO, and:

Was (person's) rehabilitation from the state, or from another job rehabilitation program?

Note: The question is "decomposed", or divided up, to make it easier to understand. The term "vocational" is also changed to the more understandable form "job".

EXAMPLE 2:

1) Original form of question:

How long has (name) used the (cane, wheelchair, walker...)?

2) Probes:

a) **How did you get the answer of (x) years?**
(To determine the overall cognitive strategy used.)

b) **When did (x) first use the (device)?**
(To test comprehension/interpretation of the question.)

c) **How well do you remember this?**
(To test recall of the relevant information.)

3) Results:

It was found that for target individuals whose use was intermittent over a long period of time, the question was interpreted in two distinctly different ways:

1) "How long has it been since (person) first used the (device)? For example, the subject may say: "since 1960, so about 30 years".

2) "For how long, overall, has (person) actually used the device since first having it? The subject counts up periods of use within a longer time- for example: "For two five-year periods since 1960, so 10 years".

Note that the problem identified can be considered a type of "comprehension" problem, but doesn't involve a failure of comprehension of a key term, as did the last example. Rather, subjects simply have alternate, but reasonable, interpretations of the question intent.

4) Suggested revision:

This required consultation with the client, in order to clarify the objective of the question. It became clear that the desired expression was:

How long ago did (person) first use a (device)?

EXAMPLE 3:

1) Original form:

About how many miles from here is the home (child) lived in before (he/she) moved to this home?

(THE RESPONSE CATEGORIES ARE PRINTED ON THE QUESTIONNAIRE, BUT NOT READ):

[ ] less than 1 mile
[ ] 1-50 miles
[ ] 50+ miles

---

2) Probes:

a) **How sure are you of your answer?**
(to determine overall level of confidence)

b) **How hard was this to answer?**
(to determine level of difficulty, and likelihood of estimation/guessing)

3) Results:

No one had difficulty understanding the question as posed. However, some subjects needed to think for a fairly long time before giving an answer. Further, some subjects struggled needlessly with the level of specificity they thought was required (for example, deciding whether the distance was closer to 20 or to 25 miles, when this information was ultimately irrelevant, as the interviewer would mark "1-50 miles" in either case).

The problem can be described as one involving a difficult recall task, as opposed to comprehension. A rephrasing of the question that incorporated response alternatives was necessary to make clear to subjects the degree of precision that was necessary in their answer.

4) Suggested revision:

About how far from here is the home ____ lived in before (he/she) moved to this home- less than a mile, 1 to 50 miles, or more than 50 miles?

EXAMPLE 4:

1) Original form:

We are interested in your lifetime exercise patterns.
First, when you were 14 to 19 years old:

How many hours a week of brisk walking did you do?

How many hours a week of vigorous exercise such as running, cycling, swimming, or aerobics did you do?

How many hours a week of activities that required you to be on your feet (excluding running or walking) such as dancing, hiking, ... did you do?

---

2) Probes:

a) **Was this hard or easy to answer?**
(to determine comprehension, and overall ability to recall)

b) **How do you remember this?**
(to study recall strategy)

c) **How sure are you of your answer?**
(confidence probe)

d) **What, to you, is "vigorous exercise?"**
(comprehension/interpretation of a specific term)

3) Results:

Subjects found it very difficult to remember back to the time period specified, at the required level of detail. In fact, it seemed that some subjects really could not even answer this with respect to their current behavior, let alone their behavior many years ago. Recall of information (assuming it was ever "learned" in the first place) seemed to be the dominant problem.

As for the previous example, the cognitive interviewing staff needed to confer with the sponsor/client to clarify question objectives. We were able to determine that use of a broad scale of level of activity, comparing past and present behavior, would satisfy the data objectives:

4) Suggested revision:

We are interested in your lifetime exercise patterns. When you were 14 to 19 years old, were you more active than you are now, less active than now, or about as active as now?

EXAMPLE 5:

1) Original version:

During a typical work day at your job as an (occupation) for (employer), how much time do you spend doing strenuous physical activities such as lifting, pushing, or pulling?
[CATEGORIES ARE CONTAINED ON A CARD SHOWN TO RESPONDENT]

___ ***None***
___ ***Less than 1 hour***
___ ***1-4 hours***
___ ***4+ hours***

---

2) Probes:

a) **What type of work do you do? Describe a typical workday.**
b) **How did you arrive at the answer of X hours?**

3) Results:

Careful probing revealed that people who gave reports of 1-4 hours often were office workers who did little or no heavy physical work. This appeared to be due to biasing characteristics of the question; saying "none" makes one appear to be entirely "non-physical", and is therefore somewhat socially undesirable. This problem was seen as related to respondent decision processes, rather than to comprehension or recall. A resolution was needed to make it "easier" for someone to report little work-related physical activity:

4) Suggested revision:

The next questions are about your job as a ____ for _____.

Does your job require you to do repeated strenuous physical activities such as lifting, pushing, or pulling heavy objects?

(IF YES:) During a typical work day, how many minutes or hours altogether do you spend doing strenuous physical activities?

Note that the results of a field-based survey experiment by Willis and Schechter (1997) have supported the contention that the revised question form is very likely a better expression than was the initial version.

EXAMPLE 6:

1) Original:

Do you believe that prolonged exposure to high levels of radon gas can cause:

	YES	*NO*	*Don't Know*
Headaches?	__	__	___
Asthma?	__	__	___
Arthritis?	__	__	___
Lung Cancer?	__	__	___
Other cancers?	__	__	___

2) Probes:

a) **Why do you believe this?**

b) **How sure are you of this?**

c) **Is it difficult to answer these?**

3) Results:

Simple observation of subjects made it clear that this question is difficult to answer. Subjects required a long time to respond to each item, and tended to be unsure about several of the items. Further, probing revealed that the format encouraged a "guessing" strategy, rather than actual retrieval of information. Finally, for people who do not believe that exposure to radon is harmful, it became very tedious, and sometimes even offensive, to repeatedly ask about the specific harmful effects of radon.

In this case, it appeared that the subject's decision processes were again excessively burdened by the phrasing of the question.

4) Suggested revision:

Do you believe that prolonged exposure to radon is unhealthy, or do you believe that it has little or no effect on health?

(IF radon believed unhealthy:)

[SHOW CARD TO RESPONDENT] Which, if any, of these conditions do you believe can be caused by radon exposure?

___Headaches ___ Lung cancer
___Asthma ___ Other cancers
___Arthritis ___ Don't Know

The revised phrasing provides the respondent with a way to respond, once, that he or she does not believe that radon is harmful. Then, if he/she does believe it to be harmful, the next question simply allows him/her to "pick and choose" the items that seem appropriate. The burden on decision processes appeared to be reduced, using this alternative.

EXAMPLE 7:

1) Original:

What is the primary reason you have not tested your home for radon?

---

2) Probes:

a) Is it hard to think of the main reason?

b) Can you think of any other reasons?

c) How much have you thought about having your home tested?

3) Results:

Although the question is easily enough understood, it was very difficult for subjects to produce a reasonable answer, especially if they had never given the issue much thought. Instead of simply saying "I never thought about it", or "I haven't gotten around to it", subjects tried to think of more "appropriate" answers, that appear to be more defensible. Here both recall and decision processes appeared to be operating.

4) Suggested solution: **--- DELETE QUESTION ---**

The sponsor/client agreed that it was not especially useful to ask the reason that someone <u>had not</u> carried out this activity.

This example demonstrates an important point worth emphasizing; sometimes, there is no obvious "correction" to a survey question. Especially when subjects simply don't have information that we want, it is better to acknowledge that we may not want to ask that question. Thus, one effect of lab testing is to test the boundaries of "what can be asked and what can't."

5. DETECTION OF STRUCTURAL PROBLEMS IN SURVEY QUESTIONS

The discussion above has focused almost completely on cognitive problems in questionnaires; that is, problems involving the comprehension, recall, decision, or response processes necessary to adequately answer the question. However, cognitive interviewing has several overall positive effects, in addition to the understanding of specific cognitive processes:

a) *Learning about the topic*: One can explore the nature of the underlying concepts to be measured in the survey, and the specific topical material, by relying on lab subjects as substantive "experts". For example, no one is more knowledgeable about the topic of illicit drug use than those individuals who have used them, and the basic logic of questions on the use of assistive devices can best be assessed through intensive discussions with individuals who use canes, wheelchairs, walkers, and so on.

b) *Learning about non-cognitive defects in the questionnaire*. An important beneficial effect of lab testing is to detect structural, or logical problems, not normally viewed as relevant to the cognitive approach. Structural problems are those features of the questionnaire, such as erroneous skip patterns, unclear layout, and other elements, that do not clearly involve the cognitive processes of the respondent. This problem category also includes more subtle types of logical problems in survey questions. For example, given the question: "How long have you owned your house?" the subject may simply respond that he is a renter. Here, it should not be strictly necessary to study cognitive processes to make the discovery that the question is flawed, because simple knowledge of the appropriate logical relationships ("some people own, some people rent") should have been sufficient to avoid such a problem. However, survey designers often fail to take into account all of these logical truths when constructing a questionnaire, and the laboratory-based interview allows the subject to spontaneously point out flaws (or for the interviewer to notice them, independently of the subject's behavior).

Of course, many structural problems could be detected by either a careful expert review, or in the field pretest, rather than through the cognitive interview. However, from a practical point of view, the expert review may never get done, or it can be imperfect. The field pretest generally occurs late in the process; it is much better to detect the problems earlier rather than later, and the cognitive interview serves this purpose well. Therefore, the cognitive interview often becomes the means for "fixing the logic of the questionnaire." Note that it takes no special "techniques" to detect the types of problems mentioned above, beyond simply attending to the possibility that they can occur.

6. THE SEQUENCE OF COGNITIVE INTERVIEWING ACTIVITIES

The following sections place these techniques described above into the broader context of conducting this testing within a real-life survey development process. To appreciate this overall process, it is useful to first consider an overview of the general sequence of events that may occur after a questionnaire is designed. Below is a schematic diagram of one such sequence that incorporates cognitive interviewing techniques, as well as other pretesting techniques, into the developmental and testing sequence:

FIRST DRAFT OF QUESTIONNAIRE IS AVAILABLE FOR TESTING

PREPARATION FOR INTERVIEWING	**RECRUITMENT**
EXPERT APPRAISAL: Review questionnaire, and make suggestions for modifications prior to testing	Determine the types of subjects to be interviewed
Develop basic probes to use in first round of interviewing	Develop and implement recruitment strategy (advertisements, flyers, ...)
	Recruit and schedule subjects

COGNITIVE INTERVIEWING ROUND

> Individual interviews are conducted with 5-10 subjects
(In early rounds, emphasis is on general concepts)
(In later rounds, emphasis is on question wording, format)

Interviewers write up comments on each interview done

Interviewers meet (as a group, or with sponsors/clients) to discuss results, propose potential modifications

The questionnaire is revised by designers

< Is another round of cognitive testing to be conducted?

YES > Individual interviews are conducted with 5-10 subjects

NO

FIELD PRETESTING PHASE

SURVEY IS ADMINISTERED IN THE FIELD

7. PRACTICAL ASPECTS OF COGNITIVE INTERVIEWING

How long should a cognitive interview be?

Although interviews of up to two hours are possible, a common view is that one-hour interviews are optimal; longer periods make excessive demands on subjects. In general, the interview process should be as flexible as possible, and not require interviewers to cover a certain number of pages of a questionnaire. Questionnaires often have skip patterns that result in widely varying actual questionnaire lengths for different individuals, and subjects vary in their overall speed and the degree to which they respond in detailed ways to either the survey questions, or to probe questions.

Note that even though the interview itself may take only an hour, the interviewing process requires considerably more time. In all, preparation, interviewing, and writing up results of the interview usually take at least three hours, and sometimes considerably more. Because of this, and because cognitive interviewing can be a taxing activity, it is recommended that any individual do no more than three interviews in a single day, if possible.

What types of individuals make effective interviewers?

It is unnecessary to have an advanced degree in psychology to be a good cognitive interviewer (although a behavioral sciences background appears to be helpful). We have found that good interviewers are those people who:

a) Have experience in questionnaire design, and are knowledgeable about both survey practice and about the purpose of the questionnaire to be tested. These skills are essential when the time comes to apply the results of the interviews in revising the questionnaire.

b) Have learned the basic premises of cognitive interviewing, and are familiar with the ways in which fundamental cognitive processes may influence the survey response.

c) Have been exposed to social science research concepts such as bias, context effects, measurement and scale effects, and so on.

d) Perhaps most importantly, have good inter-personal skills, are capable of putting a subject at ease, and remaining non-judgmental in approach. There is no common agreement concerning how "professional" versus "friendly" the interviewer should be during the interview itself, in order to obtain the best quality data (this may in part depend on the personality of the interviewer, as well as the philosophy of the organization).

A common question is whether field interviewers can be taught to perform laboratory cognitive interviews. This may be possible, if interviewers can be induced to "unlearn" some habits that are

very valuable for field interviewing, but that may be counterproductive for cognitive interviewing. In particular:

a) Field interviewers have learned over time "to make a question work", for example, by re-wording it, so that a confused respondent will ultimately provide a codeable response. It must be emphasized that our task in the lab is different; to find, rather than to adjust for, flaws in the questions.

b) Interviewers tend to work as fast as possible in the field, usually in order to complete a very long interview before the respondent becomes uncooperative. Interviewers must be reminded to work at an unhurried pace in the lab.

c) Field interviewers often focus their attention on very detailed formatting and other structural features such as skip pattern errors and spelling errors. They must be instructed that the format of the questionnaire may be very rough, and that it is the questionnaire content that is of primary concern in lab testing.

d) Field interviewers are taught not to deviate from the instructions contained in the instrument. In contrast, cognitive interviewers must be comfortable departing from the questionnaire flow when this appears to be called for. They also must be able to adjust to a situation in which sequencing instructions are incorrect or missing, which often occurs in the testing of a draft questionnaire.

Cognitive interviewer training

Cognitive interviewing is an acquired skill, consisting of a number of separate sub-skills. Optimally, good interviewers can serve as "detectives" who can find problems in survey questions, and as "engineers" who can work toward developing workable solutions to the problems defined. The former skill is generally obtained more quickly than the latter, and that the attainment of mastery is very gradual. Interviewers can be taught in an incremental, step-wise fashion, consisting of as many of the following steps as possible:

a) Trainee interviewers should conduct expert reviews or appraisals of questionnaires to make determinations of structural and potential cognitive problems. They also attend early questionnaire design meetings, as well as meetings where cognitive interviewers discuss the results of cognitive testing.

b) Trainees familiarize themselves with material on the philosophy and purposes of the cognitive aspects of survey methodology and cognitive interviewing techniques.

c) They are taught the specific probing methods for use in the interview, in a lecture-based training program.

d) They are shown examples of the way that probing is used to detect problems in survey questions. This can be in both written form, and through the use of audio- and video-taped recordings of previous interviews.

e) Trainees observe experienced interviewers performing actual interviews. Unless a topic is very sensitive, subjects generally have no objection to being observed by an individual who is described as "in training."

f) Trainees perform one or more interviews while being observed by a practiced interviewer, or compile tape recording of the interviews for review by other staff. The trainee can then be given feedback.

g) Trainees attends questionnaire review meetings, subsequent to the interviews, and attempt to make specific recommendations for solution of the observed problems.

This guide is intended to serve as "training." There is no substitute for experience, however, and interviewers should begin interviewing as soon as they have a fairly good idea of what is involved.

Other considerations for interviewing

There are several features of laboratory interviewing that are important for cognitive interviewers to understand, and that are useful to express to the subject, before beginning a cognitive interview:

a) The interviewer should stress to the subject that he/she is not primarily collecting survey data on them, but rather testing a questionnaire that has questions that may be difficult to understand, hard to answer, or that make little sense.

b) Make clear that although we are asking the subject to answer the survey questions as carefully as possible, *we are primarily interested in the ways that they arrived at those answers, and the problems they encountered*. Therefore, any detailed help they can give us is of interest, even if it seems irrelevant or trivial.

c) If think-aloud responding is desired, tell subjects, at the least, to "think out loud to the extent possible, so we can tell what you are thinking about when you answer the questions." Be warned that this introduction generally does not produce a great amount of think-aloud, however. Eliciting a spontaneous flow of verbalization often requires subject practice with the technique.

d) It also is somewhat helpful to add: "I didn't write these questions, so don't worry about hurting my feelings if you criticize them- my job is to find out what's wrong with them". This helps to "bring out" subjects who may otherwise be sensitive about being overly

critical.

8. INTERVIEWING LOGISTICS

See Willis (1994) for a very detailed description of the operation of a permanent Federal-level cognitive laboratory (the Questionnaire Design Research Laboratory at NCHS).

Recruitment

In order to test a questionnaire, recruitment of the appropriate subjects is vitally important. One initially needs to identify and recruit volunteers from appropriate sub-populations for testing the survey questionnaire, taking into account several considerations:

a) Subjects either have characteristics of interest for the survey (a particular status with respect to health, work, age, sex characteristics), or they may be "general" subjects, for questionnaires that are asked of the general population. However, even for a questionnaire that is intended for special populations, it is worth testing the initial screening sections on people who do not exhibit the characteristic(s) of interest. This practice allows the interviewers to ensure that the questions do not create problems in the majority of cases in which the questionnaire will be administered (where the respondent does not have the characteristic). As an example, a questionnaire that is intended to identify individuals with Pediatric conditions might be tested only on individuals who answer an advertisement for "people with foot problems." However, failure to test the screening questions on individuals without foot problems could be catastrophic. If, for example, virtually everyone answers initial screening questions (in effect asking: "Do you have any foot problems") in the affirmative, a large number of inappropriate respondents might wind up "passing" the screener and be subjected to a series of completely irrelevant follow-up questions. As a general rule, questionnaires that seek to identify a particular population should be tested to determine that they adequately 1) screen in people having the characteristic of interest (that is, they exhibit sensitivity), and also 2) screen out those who do not (they also exhibit specificity).

b) Subjects are recruited through newspapers, fliers, service agencies, and support groups. If payment will be involved, flyers and newspaper ads should clearly emphasize this feature (monetary incentives tend to be very effective).

c) Statistical sampling methods are *not* normally used in obtaining laboratory subjects. At most, we use a "quota" sample, in which one attempts to obtain a range of ages, genders, and socio-economic levels, if possible.

Payment

As of 1999, the industry standard appears to be $25 - $50 for a one-hour interview, depending mainly on how difficult it is to induce individuals to participate. This amount is sufficient to pay for the subjects' travel time and for basic inconvenience involved in traveling to the location of the interview, if this will be necessary.[5] Further, this payment is enough that it is not simply a token remuneration; this way, we are less likely to only recruit individuals who are practiced volunteers, and who may participate mainly out of interest in the survey topic, or in surveys in general (and who may therefore be very different from the usual household survey respondent). However, the amount of payment should be determined by considering a number of issues, such as the general demographic status of the types of subjects required, difficulty of getting to the interview location, difficulty of the task, and so on.

Administration mode of the cognitive interview: face-to-face versus telephone

The "generic" cognitive interviewing procedure consists of the conduct of the cognitive interviews in a face-to-face mode, within a cognitive laboratory environment. However, it is also possible to conduct these interviews over the phone, once an interview has been scheduled. It is rare that researchers will call "out of the blue" to someone selected randomly, as in a Random-Digit-Dial telephone survey. Telephone-based interviews to be useful for several specific purposes (see Schechter, Blair, and Vande Hey, 1996):

a) When the questionnaire is intended for telephone administration.

b) When the subjects to be interviewed are unable to travel to the interviewing location (e.g., the elderly disabled), and where it is infeasible or costly to travel to them. In particular, limitations in mobility due to disability should not be a factor in determining whether an individual is eligible to be interviewed, and it makes sense to provide flexibility in this regard, through reliance on the telephone.

Generally, in-person interviews may be preferable, overall, because this allows observation of non-verbal cues, and provides a more natural type of interchange between the subject and the interviewer than may be possible over the phone. However, we advocate the imaginative use of many different testing modes (for example, one may even conduct telephone interviews within a cognitive laboratory environment, in which the interviewer and subject are placed in different rooms).

Staffing

It is helpful to develop a cadre of staff members who have a "history" of cognitive interviewing

[5]Note that interviews can also be conducted 'off-site', such as in appropriate clinics, offices of service agencies, at churches or libraries, or even in subjects' homes. The location of the interview is not nearly as important as the nature of the activities that are conducted. In determining the interview location, the focus should be mainly on "what do we have to do to interview the people we need."

experience. As mentioned above, interviewing skill is an acquired capacity, and interviewers tend to improve with time. It also helps to have a particular staff member who can be responsible for subject recruitment: placing advertisements, developing flyers, making phone calls, scheduling, and generally monitoring interviewing operations. Further, staff should have experience in relating to clients or sponsors of questionnaires, in order to communicate the findings from laboratory interviews. Finally, and very importantly, staff must have the questionnaire design experience necessary to translate laboratory findings into realistic and practical solutions.

An issue that sometimes arises is that of how many cognitive interviewers should be employed for a particular project, or testing round. Even if the size of the interviewing sample is small (9 or fewer), it is useful to use several interviewers, in order to have a variety of interviewer opinions. That is, it seems more useful to have three interviewers each conduct three interviews apiece, than to have one interviewer conduct nine. However, there is little direct evidence on the efficacy of the various interviews-per-interviewer ratios that might be used, so this is another issue that is open to debate.

Physical requirements of the cognitive laboratory

Although organizations that conduct relatively large numbers of cognitive interviews, such as NCHS, BLS, the Census Bureau, and RTI, have dedicated laboratory facilities containing video and audio equipment, and remote observation capability, cognitive interviewing does not require special physical environments, or sophisticated recording equipment. In fact, as mentioned above, many interviews have been conducted outside the cognitive laboratory, such as in service organization offices or homes. Therefore, any quiet room, such as a conference room or empty office, can serve as a "laboratory" in which to conduct interviews. Equipment needs are also minimal; it is helpful to have a tape-recorder, as it is useful to record interviews (most subjects do not object, as long as privacy and confidentiality requirements are met[6]). Video-taping is also commonly done by the permanent laboratories. If respondents are to be videotaped, it is necessary to hide the camera, or to make it minimally unobtrusive (although informed consent from the subject for taping is of course still necessary). Some organizations also make use of one-way mirrors for observation; these might also affect the interchange, however, especially when the questions that are asked are sensitive or potentially embarrassing.

[6]Although this guide does not discuss the issue in detail, organizational Institutional Review Board (IRB) requirements must often be met prior to the conduct of cognitive interviewing. Typically, IRBs have tended to be sympathetic to such efforts, as long as methods for ensuring privacy and confidentiality are established and rigidly adhered to.

Compilation of results: processing "data" from the cognitive interview

There are a variety of methods used for compiling the results from cognitive interviewing (see Willis, 1994), and no one way is necessarily best. Some organizations may instruct interviewers to carefully listen to a taped recording of each interview, whereas others will work only from written notes. Some will utilize a report for each interview that was conducted, in order to maintain the "case history" integrity of each interview; others will produce one written report which aggregates the results across interviews, in order to provide a more summarized version of the results. Finally, one may sometimes be very interested in compiling the exact responses that were given by subjects to the questions, and other times the response data will be much less relevant, as purely qualitative information may be of greater relative importance.

For readers desiring a specific recommendation in this regard, a fairly efficient means for processing "data," representing an reasonable trade-off between completeness and timeliness, appears to be the following:

a) After the interviews are completed, each interviewer summarizes their findings on a question-by-question basis, by entering comments directly under each question, using an electronic form of the survey questionnaire. For example, one tested survey question, with comments, may appear as follows:

A1. How far do you routinely travel to get health care? Would you say less than an hour, one to two hours, or more than two hours?

Comments:

Of the four subjects I tested, all had problems answering this question. Three of them objected that this really varied, depending on the type of provider they're visiting. The fourth one stated that the answer to "how far" would be five miles; note that the question is internally inconsistent, because the question implies a distance, while the answer categories are all represented by amounts of time.

Finally, it wasn't really clear what the reference period is. One subject had been to the doctor once, in the past year or so, and so didn't know how to handle the "routine" part, or how far back he should go in thinking about an answer. We really need to re-think whether we want to know how long it takes people to see the provider they saw the most, during the past X months, or how long it takes them when they go for a routine check-up (assuming they do), or something else entirely.

Note that this comment is fairly involved, points out several problems, and instead of simply suggesting a re-wording, explicitly brings out the issue of the need for better specification of question objectives. Such a result is very common, through the conduct of cognitive interviewing.

b) Comments of the type illustrated above are then be further aggregated, over interviewer, and over interview, for a complete review of a particular draft of the questionnaire.

c) The final annotated questionnaire then becomes the main section of a cognitive interviewing outcome report, which is prefaced with a description of the specific purposes of testing, the nature of the subject population, a description of recruitment methods used, the number of subjects tested, the number of interviewers used, and description of specific procedures used during the interviews. The "protocol" consisting of the tested questionnaire, along with scripted probes, can also be included as an appendix (in which case the main report may only contain those questions for which interviewers had specific comments). Alternatively, the probes can be included along with question-specific comments. It is also sometimes useful to provide an overall written summary of the most significant problems that were found, prior to the detailed question-by-question listing.

Analysis of taped interviews

At mentioned above, some researchers prefer to rely on standardized analysis of tape recordings of interviews. Be cautioned, however, that this is a very time-consuming activity, and the appropriateness of this activity depends on the nature of the testing. For production work, in which revisions are made at a fairly quick rate, it is often not possible to devote the resources necessary to transcribe or analyze taped interviews. In this case, reliance on written outcome notes alone may be sufficient. Tape-recording is still valuable, however, where project staff or a sponsor/client may want to listen to the tape to get a first-hand impression of how the questionnaire is working. Transcription or analysis of these tapes can also be valuable for purposes of research, in addition to strict questionnaire evaluation and development.

Again, cognitive interviewing outcome data tends to be qualitative, rather than quantitative. Qualitative trends worth focusing on include:

a) Dominant trends across interviews (problems that seem to emerge repeatedly).

b) "Discoveries": Even if they occur in only a single interview, there are some problems that prove to be very important, because they can severely threaten data quality in a few cases, or because these problems are expected to be fairly frequent in the actual survey.

Reliance on clinical judgment in assessing interview outcomes

Especially because of the generally small samples involved, there is very little in the way of "truth in numbers." That is, one must rely heavily on the interviewer's "clinical judgment," in determining the implications of cognitive interview findings, as these have ramifications for the fielded survey. For example, one might conclude that a particular interview was very idiosyncratic, and should be ignored. Or, it may be found that the set of subjects tested was more highly educated, on average, than the population to be surveyed. In this case, even relatively

modest levels of documented comprehension problems might motivate the designers to attempt a simplification of the questionnaire. In general, it is dangerous to conclude, for example, that if problems are found in 25% of lab interviews, then they are to be expected in 25% of field interviews. One must always be careful to apply a type of “subjective correction factor" to the interview findings, based on knowledge of the likely differences that exist between the subjects that were tested, and the respondents who will be surveyed. The capacity of the interviewing and questionnaire design staff for applying judgment, adjustment, and subjective corrections is basic to the practice of cognitive interviewing.[7]

Meetings and subsequent modification

Because the focus of cognitive interviewing is the detection of questionnaire problems, there is often a tendency to "get into testing quickly", and then deal with the problems that emerge. It is imperative, however, that initial meetings be conducted prior to interviewing, to make clear the objectives of the questionnaire, and that interviewers conduct some type of technical review or appraisal of an initial draft. The placement of an Expert Appraisal step prior to cognitive interviewing may be a particularly effective practice (Forsyth and Lessler, 1991). In fact, experienced cognitive interviewers can often anticipate the types of difficulties that may be expected, prior to interviewing. Once an initial review, and perhaps a modification, has been conducted, interviewing can be initiated. After a suitable number of interviews are completed, and interviewer notes are compiled, one can convene a group meeting to discuss findings.

The determination of what a "sufficient" number of interviews is depends on several factors:

a) If it becomes obvious after several interviews that there are major problems to be rectified, then there is little benefit in conducting more interviews before modifications are made to the questionnaire. Especially in the very early stages of development, as few as four interviews may be sufficient to constitute a "round" of interviewing.

b) Even if it appears that more interviews should be done, it is seldom necessary, to conduct more than 12 - 15 interviews before meeting or delivering comments concerning that round of interview results[8], unless one is mainly interested in obtaining quantitative data related to the answers to the survey questions themselves. In that case, though, it might be better to conduct a small-scale field pretest, as opposed to cognitive

[7]The emphasis on subjectivity, clinical judgment, and opinion may strike some readers as undisciplined and fundamentally indefensible. Note, though, that the usual alternative that has typically characterized questionnaire design (in effect, the armchair crafting of survey questions) exhibits these same problems, but on a much greater scale. The recommendation made here is not to ignore empirical evidence, but to put it in an appropriate context when making decisions about what is likely to be the best questionnaire design practice.

[8]Researchers who are subject to OMB Clearance restrictions will be limited to the conduct of nine or fewer interviews.

interviewing.

At any post-interview design meeting, interviewers should discuss their findings in detail with any questionnaire designer who has not participated in the interviewing process. *As a general rule, it is beneficial if everyone who is actively involved in the questionnaire design process, including clients, participate in cognitive testing, even if simply as an observer*. Clients or sponsors should be encouraged to observe interviews, or to listen to tape recordings; the impact of a question that is confusing or long-winded is very difficult to ignore when such evidence is presented. Very often, where abstract discussions concerning the flaws contained in a questionnaire are unconvincing, the evidence from only a few laboratory interviews can have a potent impact. This is a point worth stressing; beyond its strength in *identifying* problems, a major positive feature of the cognitive laboratory approach is in the relative *persuasiveness* of the information it collects.

Meetings should be used both to point out identified problems and to suggest resolutions to these problems. An advantage of the cognitive approach is that, if one understands the basis for the failure of a particular question, a resolution to the problem may be readily suggested. For example, if a term is clearly not understood, the designer(s) may search for an easier-to-understand substitute. Likewise, if it is found that a reference period for a question is far too long for subjects to recall information with any confidence, the use of a shorter interval is in order.

Subsequent cognitive testing rounds

After the questionnaire has been revised, based on the comments from meeting, and on any discussions with clients or sponsors, a new round of interviewing can be conducted to test the changes made, and to provide additional testing of questionnaire segments that were not yet changed. Several issues are pertinent at this stage:

a) *The number of interviewing rounds to conduct*. In one sense, a questionnaire could be tested forever, and still have problems (the perfect survey question may not exist). Optimally, one would test until all the major problems have been detected and satisfactorily addressed. Usually, however, surveys are subject to strict developmental time-frames, and limited time is available for testing. Thus, the usual procedure involves conducting as many iterative rounds is practical (usually up to three to four), prior to a field pretest, or to actual administration. Though limited lab testing may not produce a perfect questionnaire, the final product should be markedly better than if not subjected to any testing of this type (although this is always subject to subsequent decisions influencing whether or not the recommended changes are made, a very complex issue that is beyond the scope of this guide).

b) *Changes in the nature of interviewing*. As noted earlier, the nature of the interviewing rounds tends to change over the course of development of a questionnaire. Early in the process, findings relate not only to particular question wordings, but to more global issues, such as the appropriateness of the survey measurement of major concepts that the questionnaire is attempting to cover. It may be determined that a class of information is

simply not available through reliance on respondent knowledge and memory (for example, in the absence of immunization records, parents appear to have appallingly bad knowledge of their toddlers' immunization histories). Or, it may be determined that a concept is much too complicated to be measured in a few short questions, and that a long series would actually be required to adequately cover this level of complexity.

Once major conceptual problems have been ironed out, later rounds of interviewing tend to be focused more exclusively on the appropriateness of individual questions (as in the examples presented previously). Still, the unit of analysis is not necessarily the particular survey question, apart from its context; relevant findings may cover a series of questions, and relate to clarity, appropriateness of the series, biases due to question ordering, and so on. Again, the focus of testing is always on both the question, and the questionnaire. One of the challenges of engaging in a useful cycle of testing and development activities is that we must be cognizant of all of these levels, both small- and large-scale, simultaneously.

The value of limited cognitive interviewing efforts

Readers of cognitive interviewing guides, or audience members in training sessions, sometimes object that "I don't have the resources to put together a cognitive laboratory, so I won't be able to do cognitive interviews." Although it is of course beneficial to have a full-scale cognitive laboratory, this is by no means necessary. Based on testimonials from individuals who have attempted these techniques, small-scale, informal cognitive interviews of friends, colleagues, and family members do appear to be effecting in "bringing out" issues and problems in questionnaires that the designer has completely overlooked. From a common-sense perspective, it is logical to expect that before one attempts to field an interview, it pays to "try out" the questions on several individuals who can serve as reasonable respondents, and to "ask more questions" (to probe) to ensure that the concepts intended are being expressed effectively, and that the researcher has a reasonable chance of minimizing response error to the extent that the objectives of the questionnaire can be fulfilled.

9. Evaluation of the effectiveness of cognitive interviewing

At some point, researchers and others contemplating the use of cognitive interviewing ask the very reasonable question - "how do I know it works, or that it's really worth the trouble?" This is not an easy question to answer, and as such, it is becoming the focus of a considerable amount of attention. First, there are several logical issues that can be argued, in the absence of any type of quantitative evaluation data:

Do laboratory subjects differ from survey respondents? If so, does this render the results invalid?

Volunteers for cognitive interviews are by definition self-selected for participation, and are therefore clearly not representative of the survey population as a whole. Most importantly, unless explicit steps are made to counter the effect, laboratory volunteers may tend to be higher in level of education than the average survey respondent. This could have important ramifications, in that one might miss problems that occur in "real life", and the laboratory findings therefore underestimate the severity of problems.

This possibility is not usually seen as a serious flaw. In general, a set of cognitive interviews does identify a significant number of problems; it is not often reported that "the questions worked fine - we didn't find anything wrong." Further, note that if a question does not "work" in the cognitive interviews, with presumably more highly able subjects, it will almost certainly be expected to cause problems in the field, so the significant, and often serious, problems that surface through cognitive interviewing appear to justify the use of the technique. In any event, proponents of cognitive interviewers do not argue that such interviewing alone be used to evaluate questionnaires. Instead, evaluators must also rely, if possible, on field pretests with more representative respondents, and on additional forms of pretesting (such as the coding of the interaction between interviewer and respondents, as described by Oksenberg, Cannell, and Kalton, 1991).

Does it matter that the "cognitive laboratory" environment is different from that of the field?

Assuming that one makes use of a cognitive "laboratory," the physical environment will be different than it is for a field interview. This is another reason why the cognitive lab is not seen as a substitute for the field test. To see how the questionnaire works in "real-life" circumstances, it has to be tested under field conditions, and this is worth doing, even for a small survey, with a few test respondents. However, the extent to which the differences in question-answering contexts between lab and field matters may depend greatly on the type of question administered. For example, comprehension processes appear not to differ greatly between the lab and the household; if someone does not know the location of his or her abdomen in the lab, it is doubtful that he or she will know this when at home. Retrieval processes, similarly, will be different between lab and field to the extent that the home environment provides cues that affect the nature of the retrieval process. This again does not appear to be a great problem, given the experience of researchers who have used cognitive interviewing techniques extensively.

The case may be much different, however, for survey questions that ask about sensitive topics. Here, environmental variables appears to be critical, and in fact often overshadows the other, more basic cognitive processes. Therefore, one should not generally use laboratory cognitive interviewing techniques to attempt to <u>directly</u> assess how likely people will be to answer survey questions about such activities as drug use or sexual behavior. Rather, one might use the lab only as a context for more indirect, experimental studies, in which we interview individuals about their understanding of questions, or about conditions they believe would be more or less likely to prompt them to answer truthfully in a fielded survey.

Are the sample sizes from cognitive interviewing large enough?

It is sometimes argued that the cognitive approach is deficient, compared to the field pretest, because the samples used are too small to make reasonable conclusions. There are at least three faults in this argument:

a) *The purpose of laboratory interviews is not statistical estimation*. One does not desire sample sizes large enough to supply precision in statistical estimates. Rather, we strive to interview a variety of individuals.

b) *The nature of laboratory interviews is qualitative, not quantitative*. As discussed previously, one does not evaluate problems in survey questions simply by counting the number of interviews in which a problem occurs. Of course, if every interviewer reports a problem with a particular question in every interview, that is significant. However, a finding can be based on one interview; that is, an interviewer may say that "I had a person with a particular disease for which this question does not work...". This points out a potential problem which does not need to be verified by finding a large number of other individuals with the same situation; the problem is there, and needs to be addressed.

c) *The apparent increased sample size of the field pretest is often illusory*. As discussed previously, questionnaires often contain initial screening questions, and then long follow-up sections that apply only if one "passes" the screener. However, in cases where it is relatively infrequent that respondents receive the follow-up questions, the general-population-based field pretest tends to provide only a few cases in which these follow-up questions are actually tested. For example, Willis (1989) found that one pretest of 300 households routed less then 12 individuals on to an important section on "use of assistive devices" (canes, wheelchairs, etc.). On the other hand, prior laboratory testing of the same questionnaire had specifically incorporated recruitment of subjects who would naturally screen-in to the follow-up questions, and so the "effective sample size" of the lab sample turned out to be significantly larger.

Demonstration of the effectiveness of cognitive interviewing:

The points made above are generally argumentative in nature, rather than truly evaluative; no data are presented to support the contention that there is any relevance of "what happens in the cognitive interview" to "what happens in the field environment." Although there are few studies that purport to directly measure what Lessler, Tourangeau, and Salter (1989) term the degree of "carry-over" from the laboratory to the field environment, Lessler et al. did demonstrate, in an initial study of cognitive interviewing, that the types of problems found in such interviews appeared to be consistent with findings from a field situation. More recently, a controlled experiment designed to more directly assess the degree of carry-over from the cognitive laboratory to the field environment was done by Willis and Schechter (1997). The study procedure involved the production of alternative versions of survey questions, based on the results of cognitive testing, where these results also imposed explicit predictions concerning the

nature of the data estimates that would occur from fielding of those questions. For example, for a question on physical activity, on the basis of cognitive interviewing results, it was predicted that the inclusion of a screener question asking about whether the respondent engaged in *any* physical activity would result in lower overall levels of reporting than would one question which simply asked *how much* activity the person engaged in, as the latter appeared to reduce the expectation that a non-zero amount should be reported. In three subsequent field studies of various size (ranging from 78 to almost 1000 respondents), involving a range of survey types (face-to-face household, telephone, and clinical environment), these predictions were borne out in striking fashion. Although these results are clearly not definitive, they buttress the contention that cognitive interviews are similar enough to field interviews that similar mental processes are involved, and that the directionality of qualitative results from cognitive interviews can reasonably be expected to be maintained in the environment to which they are intended to generalize to.

More extensive discussion of the general issues related to the evaluation of the usefulness of cognitive interviewing, especially as it relates to other pretesting methods such as interaction coding, are discussed by Campanelli (1997), Lessler and Rothgeb (1999), Presser and Blair (1994), and Willis, et al. (1999). One ramification of these discussions appears to be that it is fruitless to determine which particular pretesting technique is superior, as they are “stacked up” against one another. Rather, they may be best applied for different purposes, and at different points in the developmental sequence-- one does not conduct a field pretest on an embryonic questionnaire, and on the other hand, the final "dress rehearsal" of a forms-designed questionnaire is not usually tested with a round of cognitive interviews. The challenge to pretesting is to utilize a cohesive developmental plan that takes advantage of the strengths of each method. If applied properly, cognitive interviewing is likely to be an effective means for identifying potential problems, *before* the problems are encountered repeatedly in the fielded survey.

References

Campanelli, P. (1997). Testing survey questions: New directions in cognitive interviewing. Bulletin de Methodologie Sociologique, 55, 5-17.

Campanelli, P., Martin, E., & Rothgeb, J.M. (1991). The use of respondent and interviewer debriefing studies as a way to study response error in survey data. The Statistician, 40, 253-264.

Campanelli, P., Rothgeb, J.M, & Martin, E. (1989). The role of respondent comprehension and interviewer knowledge in CPS labor force classification. Proceedings of the Section on Survey Methods Research, American Statistical Association, pp. 425-429.

DeMaio, T. J., & Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz and S. Sudman (Eds.), Answering questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research, pp. 177-195. San Francisco: Jossey-Bass.

Dippo, C. (1989). The use of cognitive laboratory techniques for investigating memory retrieval errors in retrospective surveys. Proceedings of the International Association of Survey Statisticians, International Statistical Institute, pp. 323-342.

Ericsson, K.A., & Simon, H.A. (1980). Verbal reports as data. Psychological Review, 87, 215-250.

Esposito, J., & Hess, J. (May, 1992). The use of interviewer debriefings to identify problematic questions on alternative questionnaires. Paper presented at the annual meeting of the American Association for Public Opinion Research, St. Petersberg, FL.

Forsyth, B., and Lessler, J.T. (1991). Cognitive laboratory methods: A taxonomy. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman (eds.), Measurement errors in surveys. New York: Wiley.

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.

Jobe, J.B., & Mingay, D.J. (1991). Cognition and survey measurement: History and overview. (1991). Applied Cognitive Psychology, 5, 175-192.

Jobe, J. B., Tourangeau, R., & Smith A. F. (1993). Contributions of survey research to the understanding of memory. Applied Cognitive Psychology, 7, 567-584.

Lessler, J.T., & Rothgeb, J.R. (1999). Integrating cognitive research into household survey design. In M. Sirken, T. Jabine, G. Willis, E. Martin, & C. Tucker (Eds.), A New Agenda for Interdisciplinary Survey Research Methods: Proceedings of the CASM II Seminar. National Center for Health Statistics, pp. 67-69.

Lessler, J.T., & Sirken, M.G. (1985). Laboratory-based research on the cognitive aspects of survey methodology: The goals and methods of the National Center for Health Statistics study. Milbank Memorial Fund Quarterly/Health and Society, 63, 565-581.

Oksenberg, L., Cannell, C., and Kalton, G. (1991) New Strategies for Pretesting Survey Questions, Journal of Official Statistics, 7, 349-365.

Payne, S.L. (1951). The art of asking questions. Princeton, NJ: Princeton University Press.

Presser, S., & Blair, J. (1994). Survey pretesting: Do different methods produce different results? In P. V. Marsden (Ed.), Sociological Methodology: Vol. 24, 73-100. Washington, DC: American Sociological Association.

Royston, P., Bercini, D., Sirken, M., & Mingay, D. (1986). Questionnaire Design Research Laboratory. Proceedings of the Section on Survey Methods Research, American Statistical Association, pp. 703-707.

Schecther, S., Blair, J., & Vande Hey, J. (1996). Conducting cognitive interviews to test self-administerered and telephone surveys: Which methods should we use? Proceedings of the Section on Survey Research Methods, American Statistical Association, 10-17.

Sirken, M., G., Herrmann, M. G., Schechter, S., Schwarz, N., Tanur, J. M., & Tourangeau, R. (1999). Cognition and survey research. New York: Wiley.

Tourangeau, R. (1984). Cognitive sciences and survey methods. In T. Jabine, M. Straf, J. Tanur, & R. Tourangeau (Eds.), Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines, pp. 73-100. Washington, DC: National Academy Press.

Willis, G. B. (1989). Evaluation of 1990 NHIS Pretest Findings. Unpublished manuscript, National Center for Health Statistics, Hyattsville, MD.

Willis, G., DeMaio, T., & Harris-Kojetin, B. (1999). Is the Bandwagon Headed to the Methodological Promised Land? Evaluation of the Validity of Cognitive Interviewing Techniques. In M. Sirken, D. Herrmann, S. Schechter, N. Schwarz, J. Tanur, & R. Tourangeau (Eds.), Cognition and Survey Research. New York: Wiley.

Willis, G.B., Royston, P., & Bercini, D. (1991). The use of verbal report methods in the development and testing of survey questionnaires. Applied Cognitive Psychology, 5, 251-267.

Willis, G.B., & Schechter, S. (1997). Evaluation of Cognitive Interviewing Techniques: Do the Results Generalize to the Field? Bulletin de Methodologie Sociologique, 55, 40-66.