Marketing Bulletin, 1998, 9, 52-66, Article 5

Questionnaire Pretesting Comes of Age

Ron Czaja

Too many researchers still entertain the false ideas that if they get answers to their survey questions or if the questions have been used in other research, there is little reason to pretest them. This paper outlines the goals of pretesting - the problems and questions that can be identified and addressed; describes a number of new pretesting techniques - their purposes, strengths, and at what stage in the questionnaire development process they are most relevant; and concludes with some general guidelines on pretesting and where to go for more detailed information.

Keywords: questionnaire design, pre-testing

Introduction

In the past 15 years, the approach to questionnaire pretesting has changed dramatically, with the development of many new and innovative techniques to improve the reliability and validity of survey questions and questionnaires. This paper outlines a wide variety of problems that pretesting can identify and cites references for many new procedures that are being used to develop and pretest questionnaires. For purposes of this paper, pretesting is generally defined as testing a set of questions or the questionnaire on members of the target population. These activities can take place both in the field and in an office or laboratory setting.

Pretest Goals

A number of researchers have pointed out that there are many purposes for pretesting and that investigators will maximize their results by specifying clear and precise pretest objectives. Dillman (1978) provides seven reasons for a Total Design Method pretest, Frey (1989) discusses 10 purposes, and Converse & Presser (1986) discuss 11 considerations when testing questions and questionnaires.

Following the lead of these authors, Table 1 lists a variety of pretesting goals. While it is not intended to be exhaustive, I have tried to include the most serious problems affecting reliability and validity that a researcher would want to know about and address. The table includes five broad categories: respondent comprehension, burden and interest; interviewer tasks; other questionnaire issues; sampling; and coding and analysis.

The issues in the first two categories are primarily concerned with whether respondents and interviewers can perform their designated tasks. For respondents, we want to know if they understand the words, terms, and concepts being used. Do they understand the question or the task being asked of them and the answer choices from which they are to select? Does the respondent's interpretation of what the question is asking coincide with what the researcher wants the question to measure? Does the respondent use different response categories or choices than those offered in the question? Are respondents attentive and interested in the questions? This last point is important and its purpose may not be obvious. Attentiveness and interest may be indicators of how hard the respondent is working to provide complete and correct answers. While we do not know this with certainty, we feel more confident when the respondent shows an interest. Many respondent-related problems are not directly

Table 1. Goals of a Pretest - Problem Identification and Questions to Address

Respondent Comprehension, Burden, and Interest

- Do respondents have difficulty understanding words, terms or concepts?
- Is the sentence structure too complex? Do respondents understand the question, the task required, and the answer format?
- Do respondents interpret the question as the researcher intends?
- Do respondents use different response categories or choices than those offered?
- Are respondents willing and able to perform the tasks required to provide accurate and complete answers?
- Are respondents attentive and interested in the questions?

Interviewer Tasks

- Do interviewers have difficulty pronouncing words or reading particular sentences?
- Do interviewers leave out words or modify the question wording in other ways?
- Do interviewers read the question and probe in a neutral manner?
- Do interviewers follow skip patterns and other instructions correctly?
- Do interviewers record complete answers? Is adequate space provided?
- Are there any other tasks interviewers have difficulty performing?

Other Questionnaire Issues

- Do the sections of the questionnaire and the questions within sections have a logical flow?
- Are the skip instructions correct?
- Is there evidence of question order effects?

Sampling

- What is the response rate? Does the response rate indicate any potential problems?
- Are the eligibility rates as expected; do these rates indicate any unexpected problems?
- Are there any indications of problems with the completeness and accuracy of the sampling frame?

Coding and Analysis

- Is it difficult to construct code categories for the question or to code responses to open-ended questions?
- Is the level of variation in responses to each question acceptable?

observable, but they can be identified by using specific techniques. Table 2 outlines some innovative approaches for identifying these types of problems.

The ability of interviewers to perform their tasks is equally important. It is important to determine whether interviewers have difficulty pronouncing certain words or reading particular sentences. Do they read the question as written or do they leave out words or modify the question wording? Are they neutral in reading questions and probing? Do they understand the instructions presented in training sessions and those written on the questionnaire? Do they record complete answers? While training sessions are designed to address these problems, we know that after interviewing commences, some interviewers do modify their behaviors.

Pretesting should also address other question and questionnaire issues. For example, do the sections of the questionnaire and the questions within sections have a logical flow? Do the skip patterns make sense and are they correct? It is crucial to insure that the correct respondents are answering intended follow up questions and that others are not being asked inappropriate questions. When pretests are a reasonable size, at least 30 completed interviews, one can also check for obvious indications of question order effects. By this I mean that a prior question or answer is influencing the response to a later question. Keep in mind however that the typically small sample sizes for pretests can only give indications of these effects for the total sample but not for subgroups.

Pretesting can also address coding and data analysis issues. Constructing code categories and coding responses to open-ended questions needs to be examined. Can meaningful categories be constructed from respondents' answers? Are respondents' answers within the framework of expected responses? What types of responses are being given to questions with an "other (please specify)" category? Is the level of variation among response categories acceptable?

Field pretests can take one of two forms: declared or undeclared. In a declared pretest, respondents are told that at times during the interview they will be asked to do more than just answer the questions. They may be asked to rephrase a question in their own words, to think out loud while they are trying to formulate their answer, or to do other things which will be discussed shortly. The goal is to get respondents' immediate thoughts and reactions to a question or task. In an undeclared pretest, the interview is conducted in the same manner as intended for the main study. The purpose is to get a sense of the dynamics of the entire interview, that is, how well the questions flow, whether the skip patterns work, the amount of time it takes to conduct the interview and so forth.

Post-interview probe questions may be asked about individual questions or responses, but the number and scope of these questions is much fewer and more limited than in a declared pretest. Declared pretests are usually part of the pre-pretesting activities while undeclared pretests are usually used to field test questionnaires and to get a sense of sampling, coding and analysis issues.

Most field pretests are carried out with the target population using the procedures planned for the main study. Generally, this includes conducting the pretest in the same geographic area as the main study and testing the intended sampling frame. The consensus among most researchers is to use experienced interviewers (Converse & Presser 1986; Fowler 1993; Czaja & Blair 1996). The rationale is that experienced interviewers are better able to spot errors

and identify problems. This is especially important when testing the sampling frame for completeness, errors, missing information, and so on. In addition, when screening of the population to find eligible respondents is required, it is important to determine or test the rates of eligibility and ineligibility and to estimate response rates for the main study. Encountering unexpectedly low eligibility and/or response rates during main data collection usually has a disastrous effect on a study's time schedule and budget.

Pretesting Techniques

In most instances, if you are not looking for problems or ambiguities you will not find them. In the past decade or so, a number of new techniques and reconstituted versions of old procedures have been used to identify problems with survey questions and questionnaires. Table 2 lists many of these techniques and the types of problems they can identify. The procedures are organized into two groups: pre-pretesting activities which take place in a laboratory or group setting prior to field pretesting and those which occur in a field pretest and afterward. The following paragraphs briefly describe each technique, highlight its advantages and disadvantages, and provide references where more information can be found.

Pre-pretesting Activities

Focus Groups. Focus groups have a long history. In recent years, however, they have been used as input for question and questionnaire development, rather than as the primary means to answer a research question. Regardless of the purpose, the same considerations concerning group composition apply (Krueger 1994). For example, a skilled group moderator is important; group size depends on how familiar the topics are to respondents and whether discussion will be readily forthcoming or sporadic; homogeneity or heterogeneity of group members depends on whether any of the topics will make participants feel uncomfortable or reluctant to discuss their thoughts; and so forth.

Focus groups work best in the early phases of question and questionnaire development and when a set of objectives and tasks to be accomplished is specified before the groups meet. It is important to list the terms, recall tasks, reporting tasks, and other matters for which feedback is required (Fowler 1993). Focus groups are very good for determining respondents' understanding of key terms and concepts; how respondents recall or retrieve information; whether behavioral frequencies are being counted, estimated, or "calculated" using some other strategy; if respondents understand the task required based on the current question wording; and what frame of reference or interpretation respondents bring to the question as worded.

An advantage of a group meeting is that the members can use the thoughts and comments of others to help stimulate and formulate their own thoughts. In addition, participants' comments and reactions can often provide valuable insights into approaches for revising questions and questionnaires (Royston et al. 1986). For example, is one question adequate or must the subject be asked about in a series of questions?

A great deal of information can be obtained from 90 minute focus groups. Typically, the groups are audio recorded or videotaped. The drawbacks are that the results are difficult to work with; time consuming to interpret; and only a limited number of terms, topics and issues can be addressed in a typical 90-minute session.

Table 2. Pretesting Techniques and Objectives

Technique	Objectives
Pre-pretesting Activities	
Focus Groups	Determine how respondents define key words, terms, and phrases Determine whether respondents interpret phrases and questions as the researcher intends Obtain a general assessment of respondents' ability to perform required tasks (e.g. recall relevant information, estimate frequency of specific behaviors, etc.) Obtain ideas for question wording
Intensive Individual or Cognitive Interviews	Identify words terms or concepts that respondents do not understand, interpret consistently, or interpret as the researcher intends.
Thinkalouds	Identify questions that respondents can not answer accurately (e.g., recall problems, inability to estimate frequencies accurately)
Special or follow-up probes: Comprehension probes Information retrieval probes Probes to evaluate response choices	Assess close-ended response choices Obtain suggestions for revising questions and/or the questionnaire
Response Latency	Identify questions that are too complex or that are difficult to understand Measure attitude strength
Computer-Assisted Coding of Concurrent Protocols	Identify respondent comprehension, retrieval, judgement, and response problems

Technique	Objectives
Expert Panel Review of Questionnaire and/or response problems	Identify potential respondent comprehension Identify potential interviewer problems Identify potential data analysis problems Obtain suggestions for revising questions and/or the questionnaire
Questionnaire Appraisal Coding System	Identify potential respondent comprehension, retrieval, judgment, and/or response problems Identify potential interviewer problems Identify potential data analysis problems Obtain suggestions for revising questions and/or the questionnaire
Field Pretesting	
Respondent Debriefing	Identify words terms or concepts that respondents do not understand, interpret consistently, or interpret as the researcher intends. Identify questions that respondents can not answer accurately (e.g., recall problems, inability to estimate frequencies accurately) Assess close-ended response choices Obtain suggestions for revising questions and/or the questionnaire
Interviewer Debriefings Group discussion of field pretest experiences and questions	Identify faulty interviewer instructions, incorrect skip patterns, inadequate space to record answers, and typographical errors
Use of rating forms	Identify questions that are awkward or difficult to read

Technique	Objectives
Group discussion of behavior coding results	Assess respondent interest Obtain suggestions for revising questions and/or the questionnaire Identify sampling problems
Vignettes	Assess whether different question wording affects respondents' interpretation of a question Identify terms and concepts that respondents interpret differently from researchers
Behavior Coding	Identify problem questions based on the frequency of occurrence of specific interviewer and/or respondent behaviors

Intensive Individual or Cognitive Interviews. These interviews are conducted one-on-one by an interviewer with a respondent from the target population, usually at the facilities of a research organization. One technique is called a thinkaloud. Respondents are asked to think out loud or verbalize their thoughts as they attempt to understand the question, retrieve relevant information from memory, and formulate their response. Interviewers are trained to use scripted probes to clarify ambiguous thoughts and statements.

Thinkalouds can be concurrent (probe questions are asked after the respondent answers the question) or retrospective (probe questions are asked at the end of the interview). The sessions are usually audio recorded so nonparticipating staff can listen to and analyze interviews. The purpose of thinkalouds is to identify comprehension, interpretation, and recall problems. A major goal is to gain an understanding of the cognitive processes that respondents use to formulate their answers.

A thinkaloud interview does not emulate the flow of a normal interview, and thus, gives no indication of interviewing and questionnaire problems. Thinking out loud and probing specific answers breaks the question flow and the relationship between questions. Also, the degree to which respondents can verbalize their thoughts may be a function of their level of education. Concern has been raised that this is not the usual way that survey respondents answer questions, and thus, it may affect the answers given (Fowler 1993).

A second form of intensive individual interviewing is the use of follow-up probes designed to identify comprehension, information retrieval, and other problems. In this type of intensive interview, follow-up questions are asked after the respondent answers a question or a series of related questions. Comprehension probes include asking respondents to paraphrase a key phrase, define a term used in a question, elaborate on an aspect of their response, rate the clarity of a phrase or concept, and identify words or phrases that are difficult to understand. The purposes of these probes are to identify terms or concepts that respondents do not

understand or that they interpret differently than intended and to determine if respondents miss important words or qualifiers that are part of the question, e.g., reference period.

Information retrieval probes are aimed at assessing respondent's confidence in their answers; determining if a question caused any uncertainty about how to answer; or obtaining information about how respondents arrived at answers to numeric, rating or recall questions. Follow up probes are also used to evaluate the adequacy of close-end answer categories. If respondents use different words or choices than those offered by the question, it is necessary to find out why.

Response Latency. This is a less common technique that can be used in conjunction with intensive individual interviews. The length of time it takes before each respondent begins to answer a question is measured from audio recorded interviews. Unusually long time lags may signify that the question is too complex or that respondents are having difficulties recalling the information they need to formulate their answers. Conversely, unusually quick responses may indicate a lack of understanding (Royston, et al. 1986). Bassili (1996) has proposed this methodology as a measure of attitude strength or opinion crystallization.

Computer Assisted Coding of Concurrent Protocols. Bolton (1991) developed this technique which uses thinkaloud interviews and codes respondents' verbal and nonverbal behavior into categories that measure respondents' cognitive difficulties in answering questions. From a series of studies, Bolton has classified words and word strings used by respondents into categories that indicate different types of difficulties in the question-answering process. This technique was used in a study dealing with low involvement activities that compared different survey questions on the same topics (Bolton & Bronkhorst 1996).

Coders listened to audio tapes of interviews and coded respondents' thinkaloud thoughts into segments by noting cues (short pauses and changes in intonation) that indicate the end of one thought or the beginning of another. The segments were then analyzed using software which examines precoded language transcripts. Words and word strings were assigned to one of sixteen categories, each of which signifies either a comprehension, retrieval, judgment, or response problem. This approach requires a considerable amount of time for coding and analysis, however, it does seem to be a good technique for identifying cognitive difficulties with attitudinal, as well as, behavioral questions.

Expert Panels. Expert panels can detect problems not found by other techniques and have the added advantage of being relatively inexpensive. Expert panels are usually a small group of people (3 to 8) that critique the questionnaire from multiple perspectives. Ideally the panel includes subject matter experts and survey professionals experienced in survey design, data collection, coding, and data analysis. In a group session, the panel reviews the questionnaire question by question. The strength of this approach comes from the diversity of expertise and the interaction that takes place in the group meeting (Czaja & Blair, 1996).

Expert panels are typically used prior to conducting a field pretest and, again, during the questionnaire revision process after a field pretest. In a study by Presser & Blair (1994) that compared four pretesting methods, expert panels identified the largest number of problems; expert panels and cognitive interviews were the only methods that diagnosed a number of analysis problems; and expert panels and behavior coding were the most consistent methods of identifying many problems and different problem types. About 50 percent of the problems

identified by expert panels were classified as "respondent semantic" and 30 percent were "respondent task" problems.

Questionnaire Appraisal Coding System. This is a technique in which trained coders perform a systematic, question-by-question analysis of a survey questionnaire using a detailed taxonomy of item characteristics which indicate potential response accuracy problems (Lessler & Forsyth 1996). Each question or instruction is rated on multiple factors that are grouped into four major categories that correspond to key stages of the response task: comprehension, information retrieval, judgment, and response generation.

For example, within the comprehension category, coders examine each question for undefined or ambiguous terms, technical terms, hidden questions, undefined or poorly defined reference periods, mismatchs between questions and their response choices, vague response choices (e.g., "always; nearly always, sometimes," etc), and a host of other potential problems. Summary statistics such as the number of codes assigned to each item and the proportion of items assigned specific codes, are used to identify and prioritize questions that need revision.

This is a low cost technique that can be used at any stage of questionnaire construction to identify problems with question wording, instructions, and questionnaire format. On the other hand, Lessler & Forsyth (1996) caution that "more research is needed before this coding scheme can be used as a general purpose tool for analyzing survey items." They point out that code categories need to be refined, that relationships between various codes need to be investigated, and that research to test the refined scheme should be conducted.

Field Pretesting

Most researchers believe that no matter how much developmental and pre-pretesting work is done on a questionnaire, the instrument must still be tested under field conditions (Converse & Presser 1986; Oksenberg et al. 1991; Fowler 1993; Czaja & Blair 1996). Field testing generally means administering a questionnaire to respondents selected from the target population using the procedures that are planned for the main study. Respondents can be selected by probability or convenience sampling and the number of completed interviews is usually between 20 and 70. Field pretesting may involve any or all of the techniques listed in table 2 and discussed below.

Respondent Debriefing. In general, the goals of respondent debriefing are the same as for intensive individual or cognitive interviews: to identify comprehension and information retrieval problems and to assess close-ended response choices. This technique is similar to the use of special or follow-up probes in intensive individual interviews except that respondent debriefing takes place in conjunction with a field pretest. Depending on the goals of the pretest, respondent debriefing can be part of a declared or undeclared pretest and the debriefing questions can be interspersed throughout the questionnaire or asked at the conclusion of the interview.

The most common approach is to conduct an undeclared pretest and ask the debriefing or follow-up probes at the end of each interview. The interviewer uses a script to direct the respondent's attention back to specific questionnaire items and then probes for specific information. For example, in one study, respondents who had more than one job had been asked a series of questions about their "main" job. In the respondent debriefing, these

respondents were asked: "You mentioned earlier that you had more than one job. How did you decide which job was your MAIN job?" (Esposito et al. 1991).

A major strength of this procedure, compared to some of the others, is that it can provide insights into the nature of the problem and, in many situations, suggestions for dealing with the problem (DeMaio & Rothgeb 1996). The limitations involve how much follow-up questioning and probing can be accomplished without unduly burdening respondents. Only a subset of the questions can be probed with any one respondent.

To evaluate the entire questionnaire, the questions are usually grouped by topic and random subsamples of respondents are asked follow-up probes for only one or a few groups of questions. The exception is for questions that apply to only a small proportion of the respondents; then, all respondents are asked the follow-up probes. Another drawback is that inserting follow-up questions and probes changes the dynamics of the interview, as well as its length.

Interviewer Debriefing. Until recently, interviewer debriefing - a group meeting to discuss interviewers' experiences administering the questionnaire and related procedures-was the primary pretesting technique for obtaining feedback on respondent, interviewer, and sampling problems. The typical agenda includes four topics: overview of the pretest to identify serious problems that occurred, question-by-question problem identification, suggestions for revising questions, and summary comments (Czaja & Blair 1996).

In far too many cases, conventional pretesting, which relies on interviewing a small sample of respondents from the target population followed by an interviewer debriefing, does not catch serious problems with questionnaire items. At fault is the structure of these activities. An interviewer's role is to read each question as instructed and to mark the appropriate response choice(s) or record the respondent's verbatim response. It is not easy for an interviewer to be both an interviewer and an observer of problems. Also, except in very obvious or blatant situations, determining whether or not a respondent is having difficulty with a question is a subjective call. Interviewers are very good at finding problems related to their responsibilities, such as typographical errors, faulty instructions, poorly worded questions, incorrect skip instructions, inadequate space to record answers, and so forth.

Interviewer debriefing sessions, while worthwhile, can also be misleading. Interviewers typically complete only a handful of interviews for a pretest. The atypical or extreme situations stand out to them and often color their evaluations. It is important to keep a proper perspective on the frequency of particular problems so that they do not appear more severe than they actually are. For example, when vocal interviewers have encountered one or a few atypical respondents, they are likely to dominate the debriefing discussion especially if it is not highly structured or controlled. One is likely to come away with the impression that the questionnaire needs serious rethinking or revision when in fact the problem has been greatly magnified.

There are, however, ways to maximize the usefulness of these meetings. One approach is to have interviewers complete a rating form in which they answer the following three questions about each question in the questionnaire (Fowler 1993):

1. Did you have any difficulty reading the question exactly as worded?

2. Does the question contain words or concepts that respondents do not understand?

3. Do respondents have difficulty retrieving information or providing an answer to the question?

This procedure allows each interviewer to contribute equally to the evaluation and provides a systematic and quantifiable method of identifying problems. The results can be tabulated prior to the group meeting and then, during the meeting, more time can be spent exploring the possible nature of the problems and how they might be fixed. The cost of using rating forms is minimal and the gains are well worth the expenditure. (Fowler 1993).

Another approach to maximize the information from an interviewer debriefing is to report the results of behavior coding (discussed below) to get interviewers' reactions, interpretations, and suggestions for each problem.

Behavior Coding. This technique was developed by Charles Cannell and his associates at the University of Michigan (Fowler & Cannell 1996). It is based on the assumption that any deviation from the ideal model, in which an interviewer reads a question exactly as written and the respondent provides a complete answer, indicates a problem with the question. Behavior coding involves tape recording interviews conducted in an undeclared field pretest and then coding, for each question, the number of times any of the following interviewer or respondent behaviors occurred:

1. Interviewer made minor wording change when reading the question
2. Interviewer made significant wording change when reading the question
3. Respondent interrupted the question reading to give his or her answer
4. Respondent asked for clarification
5. Respondent qualified his or her answer
6. Respondents initial answer was inadequate
7. Respondent gave a "don't know" response
8. Respondent refused to answer the question

Oksenberg et al. (1991) suggest that when one of these behaviors occurs in 15 per cent or more of the pretest interviews, it is likely that there is a problem with the question.

Behavior coding is a simple, low cost technique for analyzing conventional pretest interviews and to systematically identify problem questions. Reporting the results of behavior coding for each question in the interviewer debriefing session affords an opportunity to get the interviewers' interpretations and suggestions for the problem areas. The major limitation of behavior coding is that it does not identify the source of the problem. However, research cited by Fowler & Cannell(1996) has attempted to link behavior codes with specific types of problems. These authors summarized the preliminary generalizations that emerged from this research as follows:

"1. Questions that are not read as worded are awkwardly phrased or include words that are difficult to pronounce.

2. Questions that are misread and frequently interrupted often offer dangling explanations at their conclusion.

3. Questions that lead to requests for clarification often require response tasks that do not fit respondents' experience or frame of reference.

4. Questions that require clarification often are vague or contain a poorly defined term or concept.

5. Questions that stimulate inadequate answers often ask for a level of detail that is greater than the respondent can provide." (pp27-28)

Vignettes. Vignettes are hypothetical scenarios that are used to determine whether respondents understand and apply a phrase or key concept as the investigators intend. They have been used in the redesign of the National Crime Survey and the Current Population Survey (CPS) (Martin & Polivka 1995). The purpose is to assess respondent comprehension--specifically, how they define and apply key phrases or terms in answering questions.

For example, in pretesting for the Current Population Survey, vignettes were used to measure the effect of alternative lead-in questions on respondents' interpretations of various marginal "work" situations. The objective was to determine whether and how different versions of a question change respondents' interpretations of the question. A limitation of this approach is that it requires that the investigator know which terms or phrases are likely to cause difficulties, so that appropriate vignettes can be developed to test alternative question wordings (Esposito et al. 1991).

Pretest Strategies and Self-Administered Questionnaires

While most of this discussion refers to interviewer-respondent situations, many of these techniques can be used to examine self-administered questionnaires. Focus groups and thinkaloud interviews can be used in the same way that they are used for interviewer-administered studies. Respondents can be asked to complete a draft questionnaire and then asked about their interpretations of words, phrases, the entire question and any difficulties they may have performing the required tasks.

In terms of field pretesting, Fowler (1993) makes an important distinction between observing and not observing respondents completing the questionnaire. If respondents are not observed, post-interview interviews or respondent debriefing can be performed. Follow-up questions and probes are two likely procedures to be used. Time, which leads to memory problems, can be a major problem for follow-up interviews if they are not done immediately following the completion of the questionnaire. When respondents are being observed completing self-administered questionnaires, response latency measurements can be used to identify potential problems.

Also, follow-up interviews with probes and respondent debriefings are ideally suited for observed sessions. Expert panels and the questionnaire appraisal coding system are two other approaches which are suited for evaluating self-administered questionnaires. Probably the only techniques not appropriate for self-administered questionnaires are interviewer debriefing, behavior coding, and computer assisted coding of concurrent protocols

Summary and Conclusions

We have yet to see the prefect questionnaire. Recently, however, a number of new and revamped pretesting techniques have been developed to improve the reliability and validity of survey instruments. As we have noted, these techniques have different goals and strengths and the prevailing wisdom is that multiple pretests should be the norm.

How much and what type of pretesting is needed depends, of course, on the goals and complexities of the survey and the number of new questions. Several researchers recommend using a variety of techniques to evaluate survey instruments at different stages (Esposito et al. 1991; Oksenberg et al. 1991; Fowler 1995; Converse & Presser 1986; Czaja & Blair 1996).

In addition to testing questions informally on colleagues, students, or others, early stages of questionnaire development can involve focus groups, intensive individual or cognitive interviews, expert panels, and computer assisted or questionnaire appraisal coding systems. In the later stages, field pretesting can include declared and undeclared respondent debriefings, usually with special probing; interviewer debriefings in combination with rating forms or behavior coding; and/or the use of vignettes. The final stage should be a pilot study with a sample selected from the target population and which emulates as closely as possible the procedures to be used in the main study.

Pretesting is good insurance against making mistakes in the main study. To this end, there are a number of simple guidelines to keep in mind when deciding how much testing is necessary:

1. Do not assume that questions taken from a previous survey are necessarily good questions or good measures with your population.

2. Look for problems. If you do not look for problems, you are likely to find only a few of the most blatant ones.

3. Always plan on doing more than one pretest, especially if you are writing a fair number of new questions. Do additional pretesting work with members of the target population each time you make significant changes in the questionnaire. Remember, the revisions you make to fix one problem can create unforeseen new problems. The only way to find out if changes have fixed a problem or improved a question is to test them.

4. Do not use the main study as your default pretest. It is far too costly and time consuming to find that respondents do not understand certain questions or that interviewers are having problems with the questionnaire at this late stage in the research.

References

Bassili JN (1996). The how and why of response latency measurement in telephone surveys. In Schwarz, N & Sudman, S (eds). *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*. San Francisco: Jossey-Bass Publishers, 319-46.

Bolton RN & Bronkhorst TM (1996). Questionnaire pretesting: computer-assisted coding of concurrent protocols. In Schwarz, N & Sudman, S (eds). *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research.* San Francisco: Jossey-Bass Publishers, 37-64.

Bolton RN (1991). An exploratory investigation of questionnaire pretesting with verbal protocol analysis. *Advances in Consumer Research, 28,* 558-565.

Converse JM & Presser S (1986). *Survey Questions: Handcrafting the Standardized Questionnaire*. California: Sage.

Czaja R & Blair J (1996). *Designing Surveys: A guide to Decisions and Procedures*. California: Pine Forge.

DeMaio TJ & Rothgeb JM (1996). Cognitive interviewing techniques: in the lab and in the field. In Schwarz, N & Sudman, S (eds*). Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*. San Francisco: Jossey-Bass Publishers, 177-196.

Dillman DA (1978). *Mail and Telephone Surveys: the Total Design Method*. New York: John Wiley & Sons.

Esposito JL; Campanelli PC; Rothgeb J & Polivka AE (1991). *Determining Which Questions are Best: Methodologies for Evaluating Survey Questions*. In Proceedings of the American Statistical Association (Survey Research Methods Section). Alexandria, VA: American Statistical Association, 46-55.

Fowler Jr. FJ & Cannell CF (1996). Using behavioral coding to identify cognitive problems with survey questions. In Schwarz, N & Sudman, S (eds). *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*. San Francisco: Jossey-Bass Publishers, 15-36.

Fowler Jr. FJ (1993). Survey Research Methods (Second Edition). California: Sage.

Frey JH (1989). *Survey Research by Telephone* (Second Edition). California: Sage.

Krueger RA (1994*). Focus Groups: A Practical Guide for Applied Research* (Second Edition). California: Sage.

Lessler JT & Forsyth BH (1996). A coding system for appraising questionnaires. In Schwarz N & Sudman S (eds). *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*. San Francisco: Jossey-Bass Publishers, 259-292.

Martin E & Polivka AE (1995). Diagnostics for redesigning questionnaires. *Public Opinion Quarterly*, *59* (4) 547-567.

Oksenberg L; Cannell CF & Kalton G (1991). New strategies for pretesting survey questions. *Journal of Official Statistics, 7* (3), 349-365.

Presser S & Blair J (1994). Survey pretesting: do different methods produce different results? In Marsden PV et al. (eds.) *Sociological Methodology* Volume 24. Cambridge, MA: The American Sociological Association, 73-104.

Royston P; Bercini D; Sirken M & Mingay D. (1986). *Questionnaire Design Research Laboratory*. Proceedings American Statistics Association, 703-707.

Ronald Czaja is an associate professor, Department of Sociology & Anthropology, North Carolina State University, Raleigh, NC. 27695-8107. Portions of this paper were written while he was a visiting academic, Department of Marketing, Massey University, Palmerston North, NZ.