The Tools at Hand

In: Survey Questions

By: Jean M. Converse & Stanley Presser

Pub. Date: 2011
Access Date: September 16, 2021
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9780803927438
Online ISBN: 9781412986045
DOI: https://dx.doi.org/10.4135/9781412986045
Print pages: 48-75

The Tools at Hand

Every questionnaire must, finally, be handcrafted. It is not only that questionnaire writing must be “artful”; each questionnaire is also unique, an original. A designer must cut and try, see how it looks and sounds, see how people react to it, and then cut again, and try again. Handcrafting a questionnaire involves successive trials, which we shall consider here in two stages: exploration and pretesting.

Exploration

The crafting of a questionnaire involves intellectual preparation of all sorts, well prior to the exploration we consider here. One must have a clear set of research purposes, knowledge of work on the problem that has already been conducted (“the literature” bearing on concepts and data), and some lively notions of how a survey could shed some new light. How should one start on the preliminary work of actual questionnaire design? We recommend starting out by consulting two kinds of people with special expertise.

Experts and Insiders

Professional experts

An exploratory study should take investigators out beyond their own academic or industrial subculture, to new “experts”—ones with differing counsel if at all possible. If investigators consult only the likeminded, they are likely to constrict the intellectual range of their inquiry and to give the *appearance* of bias. Academic researchers do not often have to contend with the kind of publicity that election pollsters face routinely, but when they do, they are much better prepared if they have already absorbed criticism from cultural strangers or even political “enemies.” The recent attack on the Ladd-Lipset surveys of American academics by Serge Lang, a mathematician, is a complicated matter and doubtless a rare event, but Lang's charges of bias should serve as a cautionary tale to make survey researchers of all kinds more sensitive to varying interpretations of their questionnaire (Ladd and Lipset, 1976; Schuman, 1983). Survey questions, finally, must *seem fair* to people of widely different viewpoints—people one will meet, at the last, in a cross-section sample of the general population.

More important still, differing perspectives and experiences can turn up new information. Consider the finding that on three issues concerning federal government action on social policies, *conservatives* endorsed federal intervention almost as often as the sample as a whole did. The questions were of this form:

- • Do you agree or disagree that the *federal government* ought to help people get medical care at low cost? [Turner and Martin, 1984, I: 82–83.]

The poll directors concluded that conservatives seemed to be moving to the left. The North American

Newspaper Alliance was skeptical, however, and commissioned a split-ballot experiment that asked the “federal government” wording to half the sample, and to the other half substituted “private enterprise.” The results showed that most people also favored intervention by private enterprise, and a comparable majority of *liberals* favored it, too.

Were conservatives moving to the left and liberals moving to the right? The split-ballot experiment suggested another hypothesis entirely: The public might be more concerned with getting action on these problems than with the issue of who took responsibility. This may be another illustration of agree/disagree problems that we noted in Chapter 2, but the important point for our present purposes is that the results were generated by the suspicion of bias. It is useful to be faced with these suspicions while one's questionnaire is still under construction, by intensive interviewing of professional experts on the other side of some intellectual or political fence.

Cultural insiders

Exploratory inquiry can involve “in depth” interviews with members of the target population, and it is the conviction of McKennell (1974), among others, that it should. He is skeptical of the practice of taking attitude items from “the literature,” because they often represent other professionals' impression of what people in general think about things, with almost no validation by work with people in general themselves. At this exploratory stage, there is little prospect of formally *sampling* the target population, but interviewing even a few individuals can enrich the researcher's perspective. Another useful procedure is to assemble somewhat more formally the insiders of a given subculture in a “focused discussion group.” This can be of special value when a target population is likely to have special perceptions, problems, and idioms that may be relatively foreign to the investigator—youth culture, gambling, drugs, prisons, and so on.

Because survey questions are now so abundant, it is the more difficult to proceed in the spirit of McKennell's (1974: 33) advice and undertake one's work “like an anthropologist approaching an alien culture, and regard one's own background and established frames of reference as a positive hindrance.” Unfortunately, most of us are probably all too likely to neglect this preliminary phase of exploration and move quite directly to writing new questions and borrowing others from the survey literature.

Borrowing Questions from Others

Large-scale surveys are now so common in American cultural and scientific life that survey questions have accumulated in a truly mountainous supply. Easy access to many survey questions is possible through various published compilations of survey questions. Early in the process of designing a questionnaire, one should consult these data, for they are very likely to save time and effort. Here we list nine that we have found most useful:

- • Converse, Philip E., Jean D. Dotson, Wendy J. Hoag, and William H. McGee III (eds.), *American Social Attitudes Data Sourcebook 1947–1978* (Cambridge, MA: Harvard University Press, 1980).
- • Gallup, George, *The Gallup Poll: 1935–1971* 9 vols. (New York: Random House, vols. 1–3, 1935–1971; Wilmington, DE: Scholarly Resources, Inc., vols. 4–9, 1972–1981).
- • Hastings, Philip K. and Jessie C. Southwick (eds.), *Survey Data for Trend Analysis: An Index to Repeated Questions in U. S. National Surveys Held by the Roper Public Opinion Research Center* (Roper Public Opinion Research Center, 1974).
- • Martin, Elizabeth, Diana McDuffee, and Stanley Presser, *Sourcebook of Harris National Surveys: Repeated Questions 1963–1976* (Chapel Hill: Institute for Research in Social Science, University of North Carolina Press, 1981).
- • Miller, Warren E., Arthur H. Miller, and Edward J. Schneider, *American National Election Studies Data Sourcebook 1951–1978* (Cambridge, MA: Harvard University Press, 1980).
- • National Opinion Research Center, *General Social Surveys 1972–1985: Cumulative Code Book* (Chicago: NORC, 1985).
- • Robinson, John P., Robert Athanasiou, and Kendra B. Head, *Measures of Occupational Attitudes and Occupational Characteristics* (Ann Arbor, MI: Institute for Social Research, 1969).
- • Robinson, John P., Jerrold G. Rusk, and Kendra B. Head, *Measures of Political Attitudes* (Ann Arbor, MI: Institute for Social Research, 1968).
- • Robinson, John P., and Phillip R. Shaver, *Measures of Social-Psychological Attitudes* (Ann Arbor, MI: Institute for Social Research, 1973, rev. ed.).

Need we pretest these tried and true questions? It's a good idea, for two reasons. First, because language constantly changes, and we may catch some of these changes only if interviewers listen carefully and relay respondent comments. For example, interviewers for the National Election Study reported new meanings given this question:

- • Do you think the people in Washington are smart people who know what they are doing?

This item had been used in the 1960s as one of several indicators of confidence in government. At the time of the Watergate scandals in the 1970s, however, new meaning was given to the question, as certain respondents volunteered wryly, “Oh yeah, those guys know what they're doing, all right—they're plenty smart.”

Were *many* people supplying a cynical new context for “smart people”? There is rarely a means of knowing from pretests. These vignettes from the field nevertheless serve as a reminder that survey questions can weather and age with time, and some should be retired from replication. Pretesting of borrowed items is important, for a second reason, because the meaning of questions can be affected by the context of

neighboring questions in the interview, as we have noted in Chapter 2.

Pretesting: Strategies, Purposes, and Phases

Pretesting a survey questionnaire is always recommended—no text in survey methods would speak against such hallowed scientific advice—but in practice it is probably often honored in the breach or the hurry. There is never the money nor, as deadlines loom, the time, to do enough of it. There is a corollary weakness that the practice is intuitive and informal. There are no general principles of good pretesting, no systematization of practice, no consensus about expectations, and we rarely leave records for each other. How a pretest was conducted, what investigators learned from it, how they redesigned their questionnaire on the basis of it—these matters are reported only sketchily in research reports, if at all. Not surprisingly, the power of pretests is sometimes exaggerated and their potential often unrealized.

Strategies of Design

A given pretest scheme projects a set of expectations for respondents, and variations in these change the character of the pretest. The design can change by whether or not respondents know that it is a pretest and by the role that interviewers play.

Respondents' awareness: participating and undeclared pretests

We term it a “participating” pretest when respondents are *told* that this is a practice run, and are asked to explain their reactions and answers. This design opens some doors of information and closes others. It opens the possibility, for example, of asking very detailed probes about each question, phrase by phrase, even key word by word.

- • “What did the whole question mean to you? How would you say it?”
- • “What did___make you think of?”
- • “What was it you had in mind when you said___?”

Because there is no need to simulate an actual interview, one can also ask respondents to react to different wordings of the same basic question.

- • “Consider the same question this way:___”
- • “How would you answer that question now?”
- • “You said___Would you feel differently if I said___?”

With this “intensive” design, one can examine a few questions in great detail. Not a large number, or the questionnaire as a whole. The participating strategy may limit the range of possible respondents as well, because it is probably of greatest interest to people who are accustomed to surveys, reflective and confident about their own opinions and mental processes, sensitive to nuances of language, as well as willing to give up time and thought to help social scientists. Narrowed down to this subset, investigators may find themselves relying on that familiar source of forced labor—colleagues, friends, and family.

One interesting study of pretesting used respondents from the broader public (convenience samples of store customers and a random sample of households) in a variant on the participating pretest (Hunt et al., 1982). The findings raise important doubts about whether or not the general public should be asked to serve as actual judges of survey questions. In a short written questionnaire, ten questions were designed to represent five well-known faults, such as questions with incomplete alternatives, as in “Do you vote Republican or Democrat?”; and questions with inappropriate vocabulary, such as “Do you think that the current inflation is demand based or cost based?”

The researchers felt that they built in “blatant” errors and asked their subjects “to be critical,” but most of their respondents found nothing much to complain about. The missing alternative error was by far the most visible: a third of the 146 respondents commented on the problem, which is probably about the proportion who needed the main missing alternative, “Independent.” Inappropriate vocabulary was noticeable to a somewhat smaller group, but “loaded words,” “double-barreled” questions, and ambiguous questions were virtually unnoticed by almost everyone. This rare and welcome research on the pretest suggests that respondents are not very critical or sophisticated about survey questions, even when invited to be, and their counsel may not be a very good guide to practice. (Many respondents probably answer survey questions out of basic civility and politeness). Their actual *answers*, nevertheless—their interpretations and misinterpretations of investigators' intent—are likely to be illuminating.

In what we call the “undeclared” pretest the respondent is not told that this is a questionnaire under construction, and the interviewer plays it straight. Here one can indeed probe *some* of the questions for respondents' frame of reference and meaning, but not with the intensity or exhaustiveness permitted in the participating strategy. In this more “extensive” design one can test more questions and do so in a mode closer to the final questionnaire. The best strategy is probably to begin with a participating pretest, then move to an undeclared one.

The interviewers' responsibilities

The role of interviewers in the pretest can also vary, especially in the degree to which the role is structured. Some investigators who do pretesting themselves prefer the freedom of improvising questions on the spot during a first pretest. Very skilled interviewers can be instructed to do the same thing, keeping a close record of the exact questions they asked, or tape-recording the whole encounter. The degree of structure can also vary within the same pretest, with a few key staff researchers free to depart from a structured schedule to

explore leads on the spot, and most interviewers constrained to test the questionnaire as written. In another variant, all interviewers can ask some unstructured questions at the end of a standardized interview (Belson, 1981).

The Purposes of Pretests

The confident comment that a certain question “has been pretested” implies, first, that pretesting is a permanent state of grace—once pretested, always pretested. This in turn implies that a pretested question can be pulled out of one questionnaire and simply patched into another without losing its pretesting credentials. Neither assumption is safe. “Pretested for what?” is the appropriate query, for there are some very specific purposes. Consider these ten, the first four of which are tests for specific questions:

- • variation
- • meaning
- • task difficulty
- • respondent interest and attention

The last six bear more on the questionnaire as a whole:

- • “flow” and naturalness of the sections
- • the order of questions
- • skip patterns
- • timing
- • respondent interest and attention, overall
- • respondent well-being

We shall consider the two groups in turn.

Testing questions

Variation. Testing items for an acceptable level of variation in the target population is one of the most common goals of pretesting, and it is probably this purpose that people usually have in mind when they say that a question “has been pretested.” Questions that show a 95/5 or 99/1 distribution of Yes/No may represent descriptive findings of capital importance if, for example, they mean that 95% of a population has learned to read and write, that 1% is at risk for a certain disease, or the like. Often, however, one is on the lookout for

items showing greater variability that will be useful in detecting subgroups of people or clusters of attitudes of analytic interest. One rarely has enough pretest cases to be at all confident, but very skewed distributions from a pretest can at least serve as warning signals.

Meaning. For this purpose, the fact that an item has been used in a published study may not tell whether it was ever fully pretested for meaning; or whether the meaning intended by the investigator was shared by most respondents at the time; or if so, whether it *still* is. If the original source was a written questionnaire administered to college students, it should be considered with special suspicion, and duly pretested for its applicability to a general population. The literature offers some colorful examples of confusion—“profits” taken for “prophets” is a classic—and other pretest vignettes provide other merry examples: “heavy traffic in the neighborhood” meaning trucks to investigators and drugs to respondents; “family planning” meaning birth control to investigators and saving money for vacations to other people (Berckmans, 1985). One woman of whom we recently asked family social-class level—“poor, working class, middle class, etc.”—told us that her family wasn't really very sociable. How many people brought quite a different meaning to “social class” than the one we intended? Very few, as we happily learned from pretest results, but it was worth checking.

Testing the meaning of questions is probably the most important pretesting purpose. It may nevertheless be the most neglected because it can require such extensive probing, as in Belson's ambitious tests of questions. He administered, first, a short questionnaire to four small quota samples of approximately 50 each on the subject of television, using questions of a sort that he found in frequent use by a set of market researchers. The day after the interview, a specially trained interviewer returned to conduct a second intensive interview on the meaning of the questions.

On the face of it, the results are appalling. In no case did all respondents bring to every part of the question the approximate meaning intended by the investigator. For the highest scoring question, 50% of the sample interpreted all parts of the question within acceptable meanings. In the lowest-scoring question, *nobody* did. The average, overall, was an unimpressive 29%. The fact that Belson selected problem questions, however—types that he expected to confuse respondents—makes his findings a little less grim.

Belson's findings, nonetheless, carry two very important messages. First, the meaning that investigators intend for many questions actually used in surveys is often not the meaning that respondents apprehend. Respondents do not necessarily even *hear* every word of the question, much less assume the definitions that the investigator has in mind, or fully understand the concepts. Although a question presents the word “impartial,” for example, perhaps the “im” gets lost and the respondent hears a synonym for biased, or perhaps the word “impartial” itself is an unfamiliar one (Belson, 1981: 76–86).

Second, respondents nevertheless answer most questions because, as Belson (1981: 371) writes,

> When a respondent finds it difficult to answer a question, he is likely to modify it in such a way as to be able to answer it more easily.

This is the first of 15 hypotheses that Belson distills from his results. For our reading, it is the most important

and the most general of them all, and in fact many of the others could be subsumed under it. Respondents probably modify questions, for the most part, not because they are lazy or want to hide their ignorance, or even because surveys project a demand to be opinionated, as Riesman and Glazer (1948) cogently observed over 35 years ago. People seem to answer questions, most importantly, because they expect survey questions—or at least survey interviewers—to be sensible; people *think* they heard the question properly; or the part of the question that is meaningful takes on more salience or vividness and they answer that part; or they take the parts that are meaningful and reassemble a different question from them and answer that. In sum, respondents probably *transform* obscure questions into ones that seem sensible from their standpoint as they strain for meaning.

Task difficulty. A question can be hard to answer even though the meaning is entirely clear if the respondent has not previously packaged the information in the way the question demands. “How many pounds of coffee have you consumed this past year?” may be an answerable question for some very methodical shoppers, but most of us do not total up our consumption of coffee by the year, or even by the pound, and our estimate in these terms would probably be quite unreliable. The literature about “non-attitudes” should help us avoid asking respondents things they know little about (Converse, 1970; Smith, 1984a)—but it continues to be easy, nevertheless, to ask people questions we want to know rather than ones they are able to answer.

Take a recent DAS question:

- • How important has it been to you to have more money than your parents had—very important, somewhat important, or not very important?

From a few follow-up probes, we learned that the question triggered people's ruminations about the importance of money itself (“money doesn't buy happiness,” and so on) often without reference to their parents' fortunes, as if the question had read:

- • How important has it been to you to have more money?… Very important, somewhat important, or not very important?

It seems to have been an evocative question for people who had been very poor as children; for others who had no such vivid memories, however, the whole idea may have been quite new. A few of our respondents volunteered just that. “Why I never thought of it before—I guess I'd have to say Not Very Important.”

There are two morals to the money story. One turns on retrospective measurement itself, which is rarely recommended, because it is known that recall even of dramatic events can be fragile: People may forget the event itself or they may misremember the time at which the event occurred. And yet retrospective measures are sometimes essential. We asked a number of retrospective questions in this particular study, without illusion that we were getting exact measures of the past; for reasons connected to the analysis we were willing to take the risk of rough approximations.

There are, indeed, such times when any “rules” of question writing will be bent, because at the time it seems

rather a matter of Hobson's choice. In this case, however—and this is the second moral—we did not pretest this question well enough. It was only on the final questionnaire that we picked up a smattering of evidence that many people had not organized the past or compared it with the present in quite the way we were assuming. Had we gotten the cues earlier, we might well have added an open-ended probe to the question on the final questionnaire to distinguish systematically the people for whom the question was a meaningful one. Or dropped the question.

Respondent interest and attention. This is an aspect of pretesting that ordinarily seems to get rather short shrift. Investigators are rather prone to forgetting that not everyone brings the same fervent interest to their topic that they do. Interviewers usually know whether respondents seem interested or stimulated by the questionnaire, and should be asked to report systematically on this dimension, at least to note questions that respondents found especially interesting and especially dull.

There is little research to support such caution. In fact, the proposition that survey data deteriorate when respondents' attention and interest flag markedly is one we take largely on faith and the evidence of fatigue effects in learning experiments. There is some recent suggestive evidence that such things hold in surveys, too. In a self-administered questionnaire, Herzog and Bachman (1981) asked some fairly large sets of questions in the same format; as many as 23 items were grouped together with a four-point choice ranging from Strongly Agree to Strongly Disagree. This format would probably have been more monotonous and deadly in a personal interview than in this paper-and-pencil questionnaire, but there was evidence of a “fatigue” effect toward the end of the question sets, as respondents tended to check the same alternative, no matter what the question.

Still, research into fatigue or boredom effects in surveys is rare—and it is a more complicated matter than the sheer length of time required by the questionnaire. Sharp and Frankel's (1983) recent study shows that length of the interview is experienced by respondents as a negative factor but is not in fact predictive of their willingness to participate in subsequent interviews. In the design of questions, investigators usually try to avoid the wearing repetition. They make of their surveys something of a pastiche, one kind of questioning and then another; a bit of so-and-so's format, and something in another style. This is an aspect of survey design that is guided by mysterious matters: the “art” of writing and especially arranging survey questions to keep respondents' attention and interest. Varying the format is important. In personal interviews, one can do this not only by using different kinds of oral questions but also by using some “show cards” or a small booklet that lists answer alternatives and some self-administered questions, but on the telephone one's choices are more restricted. Here, especially, one must pretest carefully for a desirable balance between the benefits of question variety and the “start-up” costs of explaining a new question format.

The first four objectives that we have discussed bear especially on the testing of *questions.* Another set of purposes that we turn to now bears more on the *questionnaire as a whole.* We should not put too fine a point on this distinction, for in every phase of pretesting, one is tinkering with the wording and form of specific questions and also trying to cast the overall shape of the questionnaire. Still, there is a difference in emphasis. One cannot really test the questionnaire as a whole until the basic sections are chosen and arranged in a

given order, with the wording of many questions cast in more or less final form. So the pretesting objectives for the questionnaire tend to come second.

Pretesting the questionnaire

"Flow" and naturalness. Testing the "flow" of the questionnaire is such a matter of intuitive judgment that it is hard to describe or codify. One can at least be guided by one crucial caveat: Reading is not enough. One must listen to the questionnaire, over and over, hearing it as interviewers actually deliver it, trying to hear it as respondents do, always mindful that they will *not* have the print in front of them to review and clarify the meaning. What respondents hear is what they get, and every question probably comes anew to respondents, with a certain "surprise" quality.

The interest and clarity of the questions and a "sensible" arrangement probably contribute more to a coherent flow than any very elaborate transitions from section to section. Transitions can be simple and brief, such as "Now I am going to ask you some questions about your job…" or even the vague preparation, "Now I have some questions on a different topic…" These are not very elegant transitions, but in our experience they do not have to have nor should there be many of them. One wants a few (redundant) words that slow things down enough for the interviewer and the respondent to turn a corner to another subject—now and then, when without a transitional phrase changes seem too sharp.

The order of questions. The positive guides to the order of questions are few, and bear only the credentials of common sense. The proposition that one should open the questionnaire with "interesting" questions, for instance, seems like a good idea—assuming of course that one has tested questions for their interestingness.

Frey (1983: 103–105) offers more specific counsel that in telephone interviewing, especially, the initial questions should be items directly related to the topic of the interview expressed in the introductory statement, as in this example:

> Hello. This is___calling from the Telephone Survey Center of the University of Nevada, Las Vegas. We are conducting a survey of Nevada residents on their opinions and perceptions of the quality of life in the state….
>
> First, I would like to ask how you feel about Nevada as a place to live. Do you consider it Very Desirable, Somewhat Desirable, Somewhat Undesirable, or Very Undesirable?

Frey suggests following this with an open-ended "Why" question of this sort: "What specifically do you find (desirable/undesirable) about living in Nevada?" to allow the respondent "to find his or her 'telephone voice.'" His basic point about the opening question is that if the introduction has successfully aroused the respondent's interest, it can be just as quickly deflated by a question unrelated to the announced topic, such as "Did you vote in the last national election?" or "What is your race?" Because we know that respondents' answers to open-ended questions tend to be more terse on the telephone than they are face-to-face (Groves and Kahn, 1979), finding a "telephone voice" may be of some special value. We would nevertheless urge

caution in using open-ended questions with any frequency at the beginning *if* the questionnaire is basically closed, for one is “training” the respondent in the questionnaire from the very outset. In addition, open questions can sometimes be quite demanding, and tasks at the beginning of the interview should be easy ones that do not tax or discourage the respondent.

It is common practice to put “background” questions at the end of the interview. This ritual reflects, most of all, the sensitivity of income questions, which are the most vulnerable to refusal. If respondents are offended by being asked their income, at least their negative reaction appears late in the questionnaire. Most background questions are not really sensitive, however, and they are usually fairly easy for the respondent to answer and can be something of a welcome set of questions if earlier questions have been fairly demanding.

It may not be necessary, however, to move all background questions to the end of the questionnaire just because the income questions may belong there. Putting some background questions up front can be of use if the survey topic bears closely on the individual's own life history and experience, and the approach of the questionnaire is largely chronological. Moreover, if the interview is broken off before completion, some basic background information has been obtained. NORC's General Social Survey divides background questions between the beginning and the end of the interview for these reasons.

Skip patterns. Questionnaires must be pretested not only for the usual typographical errors, but for the logic and format of skip patterns, which can be very complex. If the skip patterns are incorrect or ambiguous, interviewers may vault over various questions or even whole sections and leave unanticipated holes in the data. Defective skip patterns should be caught before pretesting so that pretest interviewers can concentrate on the questions and the respondents' reactions rather than having to struggle against bad skips to get the questionnaire read at all.

The best way of proofreading the skip patterns is to turn the task over to several individuals, each of whom follows the route for a certain “scenario,” such as these:

> (A) The respondent was born in Mexico in 1947 and came to the U.S. in 1966. This is her first marriage; her husband had two children by a previous marriage, who are now living with R and her husband. She first voted in the 1976 national election…
>
> (B) The respondent, age 56, has worked at an automobile assembly plant for 11 years as manager of the shipping department. He is married, without children. He and his wife have recently bought into an investment partnership that is buying real estate in Florida, and they hope to retire there when he is 62…

Dividing up the labor in this way is useful for finding logical errors in the skip patterns. People who are sophisticated about surveys in general but uninformed about this one are especially valuable for this assignment, for they will duly follow the road signs of the questionnaire. Staff members who have been involved in designing the study may tend to go where they are “supposed” to go whether the skip directs them there or not.

It is important to think of the questionnaire as a road map, and consider its graphics accordingly. If the skip pattern is at all ambiguous visually, an interviewer may take the wrong route; and then confident that this is the correct way, never carefully read that instruction again. The clearest instructions are strong arrows, lines, outlines, very boxy boxes, with no fine print and no extraneous instructions at all.

Timing. It is ordinarily useful to ask interviewers to time each part of the questionnaire, section by section. That face-to-face survey interviews should average no more than a scant hour is a norm of practice with little grounding in experiment, though it has a long tradition in the 50-minute hour of clinical therapy and college classroom instruction. Beyond that time, one begins to worry about respondent fatigue, interview break-off, and initial refusal if respondents know the expected length. For telephone interviewing, the norms are for shorter times.

Respondent interest and attention, overall. Interviewers should be encouraged to notice and report on respondents' interest in the study. The problem may indeed be that the whole questionnaire is simply too long, and, if so, there is probably no better remedy than going back to the cutting board. Two other prescriptions should be considered, as well: new content and task variety.

Investigators will be loath to give any room to “throwaway” questions (without prospective analytic value) just because respondents might like to talk about them. With field costs so high, this is luxury beyond the typical research purse. Yet it is also foolish to expect respondents to have high motivation and sustained attention for a questionnaire that does not take much account of their interests. We do well to consider the questionnaire now and then from the respondents' perspective: Where is it lively and responsive? Where is it slow going? Can we go a little further to meet them halfway? Can we add anything that serves our own analytic direction and also brightens the way for respondents? At the least, we may have to place some questions strategically to perk up lagging interest.

The liveliest questionnaires in our ken move respondents from one activity to another, as they proceed through the questionnaire. Respondents are asked a set of Yes/No questions, perhaps; then a group of questions involving one choice from a list of, say, five options; then they are asked several questions in which they rate their own feelings on a scale; then they choose between one idea or the other. And so on. The variety of the answering task has been designed to engage the respondent's active attention. This characteristic can be pretested, by asking interviewers to make systematic observations of respondents' interest, and by asking respondents to report their reactions for themselves. Task variety cannot captivate respondents if the subject matter itself is irretrievably difficult or dull, for content itself appears to be far more telling. Questionnaires bearing on people's own experiences, life histories, and health are predictably more interesting to most people than an exclusive focus on attitudes or information. But when investigators want to explore topics that are not likely to be of widespread appeal, task variety may well be of special importance.

Respondent well-being. In our time surveys have burgeoned far beyond the original realms of political polling. Survey interviewers are now admitted into realms of privacy and sensitivity, to ask questions about alcoholism, drugs, crime, heterosexual and homosexual experience, marital satisfaction, divorce, abuse, the

death of children, loneliness, mental illness, depression, suicide, physical handicaps, widowhood, terminal illness, religious experience, anxiety, and faith. The litany is long. Survey researchers have proved a voracious lot, with a huge appetite for information about people's lives and experiences.

Are people sufficiently protected from possible injury in this research process? Requirements that surveys offer “informed consent” to their respondents and undergo “human subjects review” by their institution represent an effort to protect respondents from any untoward effects. There are real problems and dilemmas in these efforts, however. In surveys contracted by and for the federal government, for example, questionnaires themselves are reviewed, and this means that they must be put in final form months before the field work is started, which is a major block to the research process. In the university, on the other hand, a “human subjects review” of the overall research proposal is usually undertaken months before a questionnaire is even designed; it is our impression that there is unlikely to be any review of the final questionnaire at all.

In the academic setting, who can say that a questionnaire is distressing respondents, leaving them feeling worse about themselves or their lives than the interviewer found them? If that happened, would the investigators know? Who would tell them? Is it anyone's responsibility to find out?

In our own experience, interviewers have sometimes taken that role. In a recent project, long-term and very skilled professional interviewers insisted that the questionnaire needed major redesigning. The topic was psychological depression, and interviewers reported that the pretest was *depressing* people with its exclusive focus on life's bad news and symptoms of mental illness. The design had included only those life events hypothesized to predict depression—the tragic and troubled and stressor events. The interviewers' intelligent reports and impressive experience finally prevailed: The questionnaire was duly redesigned to allow some upbeat features, such as the inclusion of some good life events and questions about how respondents had coped with their problems. The happy ending is that after the document went through *five* pretests, both respondents and interviewers registered great satisfaction with the study: a real favorite.

We have no hard data to support these moves. We were personally convinced that the problem was a real one by the particular interviewers' experience, arguments, concern—and talent. The experience pointed up the fact that some professional interviewers bring a personal involvement and a professional commitment to the well-being of respondents. When this is combined with personal self-confidence, broad professional experience, and good judgment, their advocacy on behalf of the respondent can be of enormous value to the quality of a survey. Some student interviewers, even though less experienced, bring equivalent talents and sensitivities. But temporary interviewers, who are on deck for a single study, even if they have advanced training in other fields, will usually lack that special advocacy of the respondent.

Phases of Pretesting

Pretests represent a “qualitative” stage in the quantitative survey enterprise, with N's nothing like what they should be for quantitative evidence. For small pretests, numbering 25–75 cases, is it really worth coding the

data and hand-tallying or grinding out marginal distributions? Yes—at the upper ranges, anyway. On rare occasions, marginal distributions can provide a measure of support for certain “hunches,” and diminish one's confidence in others. And they can provide some corrections. The experience of interviewing even a single real and unforgettable character, for instance, can be so vivid and compelling that staff members may come back from the field with very broad, entirely unwarranted conclusions about the questionnaire based on the infamous N of 1. A larger set of pretest interviews per person is desirable—5, for instance, would be much better. If this cannot be undertaken, quantitative data may help pretesters realize that their own experience was not entirely representative or general. And sometimes the live testimony of other interviewers who have just as vivid, entirely contrary views can help, too. In any case, both the colorful vignettes and some numbers can be useful. If interviewers' reports are often more instructive than the marginal frequencies (and in our experience, that has usually been the case), it is another indication of pretesting's qualitative hallmarks: small Ns, samples of convenience, hypothesis testing by hunch and judgment. Pretests would seem to be absolutely necessary even if almost never sufficient.

Just as most pretesting Ns are regrettably small, the number of pretest trials is often sharply limited—often to one. For a new study, to which investigators bring no previous hands-on experience, a minimum of *two* pretests is indispensable, in our view. For as we have already stressed, in a first trial the wording of the questions themselves is still uncertain enough—they are the focus of the testing—that the questionnaire does not yet have a very coherent shape. In the discussion that follows, we assume this minimum of two pretests. We will consider them in the framework of these three topics, developmental pretest I, evaluation, and polishing pretest II, and set forth some useful properties of each phase.

Developmental Pretest I

Most questions should be closed. Closed questions should generally be given the lion's share of the first pretest simply because they must generally be given the lion's share of the final interview schedule—and this is a pretest of it. To make that statement is to face one of the sternest limitations of survey research. Closed survey questions inevitably simplify and stylize the life and thought of individuals, even for the most routine of measures.

Try marital status, for instance, or number of times married. Reliably, the boxes (precoded responses) will not capture everyone's condition in a way that guides the selection of questions (skip patterns) in the rest of the questionnaire. For example, in most cases, this will be a straightforward sequence:

> (A2) First, I would like to know your current marital status—are you now married, separated, divorced, or widowed, or have you not married?
>
> (IF RESPONDENT IS DIVORCED, TURN TO PAGE 18)

But what of the man who is divorced from his first wife but lives with her and considers her his wife? He is certainly not Separated or Widowed, but is he Married, Divorced, or both? Legally, of course, he is divorced,

but that does not solve our problem. Shall we ask him the questions to be asked of divorced people, or the questions to be asked of married people, or all of both?

Anyone who has conducted or coded a fair number of survey interviews knows that human experience is much too unruly in its diversity to be fully contained by the precoded responses of closed questions. When this richness thrusts out of the boxes, like so much jungle growth, we hastily set up another box, the residual "Other," and by relegating these wondrously oddball situations into this miscellaneous junk box, we lose entirely the vividness and "life" of individual character and unusual circumstance. There is little for it. Open-ended questions are far better for capturing those details and idiosyncrasies; entirely unstructured interviews conducted by master interviewers are better still; biographies and novels, of course, are probably best of all.

One repairs to closed questions for several good reasons. The first reason, and best, is that some boxes can be built to accommodate almost all cases. Our codes for marital status clearly did not cover everyone, but they covered *almost* everyone, which is not only good enough but absolutely splendid.

Another good reason is cost. To oversimplify the choices, one can pay interviewers to spend two hours with 250 respondents or less than one hour each with 500 persons. With the two-hour version, interviewers can ask and transcribe much more "in depth" detail, but an N of 250 will be too small to support any very compelling analysis. (The sampling error will be large for an N of 500, at that.) To undertake any statistical analysis at all, one will inevitably pare away much of the detail gathered in those two hours back to the bone of gross code categories. This risks a kind of triple squander of survey materials: one will pay to gather the detail (sacrificing a larger N in the process), then shear much of it away, paying for the time that takes too.

An even better reason for closed questions bears on validity. One may seek open-ended material in the very quest for greater validity, but the choice can sometimes work the other way around when the frame of reference for open questions is ambiguous. It is difficult to keep open questions free of the "tacit assumption" that Paul Lazarsfeld noted 50 years ago, with an example that remains one of our favorites. When school children in Austria were asked in an open-ended questionnaire item what they would most like to have in life, they wrote down economic and psychological goals such as big farms, good jobs, money, happiness. Nobody mentioned intelligence. When this alternative was included in a closed version of the question, it was a great favorite. The children had apparently assumed that the open question included only the goods that might be had for the striving, not the gifts of personal endowment (Jenkins, 1935: 355). There are other strengths and weaknesses about both open and closed questions that we have considered in Chapter Two. Suffice it here to say that one is well-advised to include many closed questions in Pretest I and probe their meaning in open ended follow-ups. Because open questions will usually constitute at most a small part of the interview schedule, they must be selected and trimmed with great care.

Rough codes for open questions should be designed in advance. In the press of field deadlines, it is usually difficult to find time to anticipate codes for open questions; but even if this preliminary code construction is done in a rather informal and approximate fashion, it enhances the realism of the first pretest and the particular value of any open questions used. It has the great value of helping investigators face up to what

their objectives are in asking the question at all. And explaining the question objectives is essential for the interviewer, for without this information, the interviewer cannot know when a question has been answered.

The first pretest interview should be less than twice the final expected length. This is merely to say that the first pretest can be rather outsize, as long as one can also conduct a second pretest. A first pretest running as long as two hours can be useful, but this should be considered an absolute maximum. Beyond this length, the labor of cutting back to an hour or less will be formidable, and even if one can find respondents willing to spend two hours, the test of some questions may be doubtful if the length of the interview has tired them and dulled their attention and interest. Such a long trial may be testing personal endurance more than anything more cogent. If there is only one pretest in view, it should be much shorter: one should strive to make it an hour and a quarter as a maximum. Remember, too, that interviewers conducting a pretest may have to do most of their work in the evenings, and a long pretest may restrict them to one interview per evening; and in that case, the time scheduled for pretesting will have to be stretched.

Respondents should resemble the target population. A probability sample of the survey's target population would make an ideal pretest. But this is ordinarily much too expensive. One must nevertheless take a pretest out beyond the small worlds of colleagues, friends, or family, who offer much too thin a slice of life. Interviewers should not be left to their own devices. In a recent pretest conducted by DAS students, we found that almost half of our respondents were graduate students, hardly a cross section of the population.

There are two likely routes out of the small world. One way is to take advantage of group character or neighborhood stratification of importance to the study: One can go door-to-door in neighborhoods that are visibly ethnic, or elderly, or young and noisy with children, and so forth. A second is to interview strangers, by knocking on doors close to home or work. This seems an absolute minimum for pretesting. Going this route will ordinarily not yield as heterogeneous a collection of respondents as the first way, but it will certainly be more useful than interviewing one's own friends and relatives. (It seems fair to insist that interviewing in that inner circle should simply *not count* as a pretest.) If one is pretesting by telephone, one, of course, has no visual cues to heterogeneity, but one can achieve something of the same effect by selecting central office codes (the first three digits following the area code) that are known to span a range of neighborhoods, though one will unfortunately not be able to identify the neighborhoods themselves. One can also make a random selection of telephone numbers within those codes, staying within a close radius to save money and relaxing selection rules to save time.

A pretest N of 25–75 is reasonable. The Magic N for a pretest is of course as many as you can get. We see 25–75 as a valuable pretest range, which can vary first with the experience and talent of the interviewers. With student interviewers, one may have to settle for a yield of 2–3 interviews each. This is not an optimal number per individual but at least the task is manageable and each interviewer's share of the variance is appropriately small, both features of special value when the interviewers are inexperienced.

With experienced professional interviewers, the N usually has to be smaller because of costs, and in the best circumstances we think it safely can be. NORC's interviewing staff, for example, is ordinarily not augmented

by student trainees, and the recommendation by Sheatsley (1983) that the N can range around 10–25 probably reflects that fact. SRC's practice is similar: Pretest Ns in recent years have averaged about 30, with a half dozen interviewers each conducting about 5 interviews. Even modest pretests of this order of magnitude have value for undergirding intuitive judgments with at least a jot of data and with the informed impressions of experienced pretest interviewers.

From the sizes we have been considering—an N of 25 with professional interviewers; an N of 50 or more with students—one might infer that professionals can be twice as effective as students for pretest work. That is not implausible, if the professional interviewers are selected for their pretesting skill and are an experienced, motivated elite of their group. A student group may well be the more highly educated, to be sure, but it is also likely to vary more in the interest, motivation, and talent for survey interviewing per se. Some students turn up in a survey practicum, after all, because they have to fulfill an academic requirement, not because they have any special interest, gift, or stamina for field work. Experienced professional interviewers who survive the winnowings are, to some extent at least, self-selected for just those qualities.

Two issues about who should conduct pretests: The first issue is whether or not pretest interviewers should represent the best of the professional staff, or the full range of talent that will ultimately work on the study. The DeMaio research group (1983), among others, favors using the full range. They feel that the most able interviewers can make even a poor schedule “work,” and may not reveal the problems that a flawed schedule will present for some of the less able interviewers on the staff.

There are no data that we know to support one view or another, but our own experience argues for choosing the most talented pretesting interviewers. They tend to be strong advocates for their respondents' right to clear, sensible, interesting questions, and they spot questionnaire defects with a good deal of zest, yea, even dedicated nit-picking. The best professional interviewers are not loath to teach investigators, and in many instances their spirited counsel has been very valuable indeed. It should be noted that even among the most competent, experienced interviewers, pretesting is not everyone's cup of tea. Those with a special interest and flair for pretesting can not only catch poor designs, but also make extraordinarily good suggestions for revision (Flanagan, 1985). But this luxurious choice will be beyond the reach of many survey investigators, who will not have easy access or sufficient funds to hire professional interviewers for pretesting. In that case, everyone will be pressed into pretesting duty, whether especially talented for interviewing or not.

The second issue turns on whether or not investigators should be among them. Some writers feel that the research staff should not participate in pretesting unless they are skilled in standard interviewing techniques. Others feel that the field experience can be of value even without that training. We side with the latter. If investigators are poor interviewers, they do not need to conduct the questionnaire themselves; they can see how their questionnaire works in the field by going along with a better interviewer. Or at the very least, they can listen in on telephone interviews. Direct pretesting experience can make investigators more sensitive and sympathetic to the rigors of the interviewers' task, as well as knowledgeable (and humble) about the frailty of their questions. It is our impression that not enough researchers get their own feet wet and weary in the field.

Evaluating pretests

We have sketched out some design features for Pretest I, but these alone do not suggest how to make the best use of the pretest experience. The following six procedures all can offer something of value:

- • marginal comments on the schedule
- • oral debriefing
- • written reports (section by section of questionnaire)
- • written questionnaires (specific facets or problems)
- • field observation of the questionnaire in action
- • coding of answers and tallying of marginal frequencies

It is only the last two that do not depend directly on the observations of interviewers. The DeMaio group (1983: 119) has noted that:

> Interviewers are a key and often underrated element in the practice of survey research. They constitute the link between respondents and researchers, and in their direct contact with respondents, they can pick up valuable information which may be of interest to questionnaire designers.

As they say, “the systematic exploration of an interviewer's knowledge has been seriously neglected in the literature.” In our own experience, that knowledge has been indispensable. Except where pretests are large in size and experimental in design, thereby making possible some statistical analysis, interviewers have a virtual monopoly on the prime information.

Copious comments written in the margins of the schedule should be encouraged. At this stage, the more the better. Interviewers can be asked to give a running account of their own impressions of the interview and of all respondent comments. Interviewers vary a good deal in their ability to do this—at the very least, it takes fast writing and the ability to seize instantly upon dialogue—so one cannot expect a rich running record from everyone. But from those who can do this kind of detailed transcription, investigators can mine evidence of problems and misinterpretations, and interviewers can go back to these marginal notes for the preparation of more systematic reports.

Oral debriefing, a group discussion with interviewers very soon after the pretest, has the advantage of immediacy: Interviewers can report while their memory of the experience is quick and their interest, usually, high. (The hazards of debriefing are those of any undisciplined discussion—if the best raconteurs or the more dominant personalities swamp the meeting and inhibit some of the less voluble interviewers from reporting their experiences at all; and the meeting should be structured accordingly.) It is not inexpensive to bring a

set of professional interviewers together for what may well be a half-day's discussion, but we have found it money well spent.

Written comments may be fruitful, alone or in conjunction with a debriefing. When we have required written comments from student interviewers, we have sometimes collated the comments by topic, duplicated them, and given the entire set back to the students. The purpose has been to economize on academic class time; the schedule of the practicum was crowded enough, on occasion, that we had little time for debriefing; and given that constraint, it seemed important to provide feedback about *everyone's* experiences, so that individual students would not overgeneralize from their own. In any case, written reports have made for easy and systematic reference in revising the questionnaire.

We have also sometimes used both written reports and oral debriefing together. This has actually proved rather repetitive, for interviewers have tended to offer in discussion the same points that they make in writing. This may be a minor disadvantage, however, if the two reporting forms, in conjunction, enhance interviewers' interest and morale, while also providing useful pretest evaluation. And the combination has seemed to work that way.

A *questionnaire* can be a useful way to communicate with a far-flung interviewing staff. For example, questions such as these can serve as a useful focus for interviewer comments:

Questionnaire for Interviewers

Please make out a separate questionnaire for each pretest interview you conduct. For all “yes” answers, please specify the *question numbers* or section and *explain* what the situation or problem seemed to be.

(1) Did any of the questions seem to make R uncomfortable?

(2) Did you have to repeat any questions?

(3) Did R misinterpret any questions?

(4) Which questions were the most difficult or awkward for you to read? Have you come to dislike any specific questions? Why?

(5) Did any of the sections seem to drag?

(6) Were there any sections in which you felt that the respondent would have liked the opportunity to say more?

And so on. The specific questions of interest will depend, of course, on the survey and its particular problems.

Field observation of pretest interviewers, which comes highly recommended by the DeMaio group, allows an interviewer to give his or her entire attention to the conduct of the interview itself, while another person is free to listen to and observe how the questionnaire is working. Observation is sometimes practiced by

field supervisors but it can be valuable for survey investigators (DeMaio, 1983: 101). Another variant is organizing pairs of interviewers (professionals or students) who work together during a pretest, taking turns as interviewer and observer. They can submit a joint report, concentrating their comments exclusively on properties of the questionnaire, without relaying any evaluation of each other as interviewers. Inexperienced interviewers may be more comfortable working this way in the beginning. They may learn more about the questionnaire and contribute more to its revision if they feel more confident interviewing in the presence of a peer observer than they would with a field supervisor or survey investigator. Berckmans (1985) reports some beneficial training experiences from the pairing of experienced and inexperienced interviewers.

Coding of responses, and preparation of marginal frequencies can provide a quick summary of variation. With a small pretest N, hand tallying of the closed question responses may be perfectly serviceable, and the open-ended answers can be quick coded for just a few gross categories, or even just an indication (yes/no) that the respondent interpreted the question as intended. With more substantial Ns, it may be more practical to edit the interviews quickly for direct data entry into a microcomputer.

Coding also can be used in a more ambitious way, as a part of interviewer training, to sensitize interviewers to the problems of coding and analysis that are created by poor interviewing. We usually have students code two or three pretest interviews just as a part of the practicum, to familiarize them with coding itself, and to get the work done. With the hope that the coding experience will sharpen sensitivity and skill in interviewing, we assign to student interviewers the coding of each other's pretest questionnaires. At least some then experience the plagues that poor interviewing visits upon coders, such as illegible handwriting, careless editing, poor probing; and we hope they take away the appropriate moral. For gathering data on occupation, familiarity with coding has come to seem essential. We suspect that many interviewers continue to ask occupation questions quite poorly until they have some experience or exercises in occupational coding. Experience reported by Hauser and Featherman (1977) supports that view.

Pretest II: the polishing pretest

The second pretest should be a “dress rehearsal” of the questionnaire as a whole. Pretest II is not a time to repair gross errors, or to make new exploration. It is rather a time for cutting, trimming, splicing, rearranging, and filling in new skip patterns, formatting for clarity—polishing.

As a dress rehearsal, the polishing pretest must necessarily be an undeclared one, which will be handled as a real interview. One *aspires* to produce a Pretest II schedule that is ready for the printer or final form on the computer (for Computer Assisted Telephone Interviewing). It has never worked quite that way in our experience—one always learns new information from a polishing pretest and revises accordingly—but one learns more from Pretest II by trying to make it as close as possible to the final questionnaire.

Because Pretest II is now a slimmer model, one can now ask outsiders to criticize a draft of it before it goes into the field, just as one asks colleagues to criticize academic papers in draft. The first pretest questionnaire is usually too fat to circulate (and may risk embarrassment, at that). It may be an imposition to ask colleagues

to read it, and they probably cannot be much help at this point anyway, for the sheer bulk of the schedule may blur its main intellectual lines and certainly its formatting structure. (One can hardly expect to format completely a questionnaire that still has to be cut by some 50%.) But one can ask knowledgeable colleagues to read and criticize a draft of a polishing pretest, especially if one presents it along with the set of research objectives that are being operationalized. (Would that colleagues needed nothing more than the questions we present to deduce perfectly our research purposes.)

We have found our best critics to be colleagues interested in the subject matter who have, themselves, done some survey work in the area, and field or coding directors who have experience with a variety of survey questionnaires. The latter especially have shown a fine eye for the design properties of a questionnaire that affect production interviewing and production coding. The problem is time: finding a hole in the survey schedule that is big enough to get a draft out to colleagues and back with comments. When we have not found that time before Pretest II, we have asked colleagues to criticize a draft of the final questionnaire. Earlier is of course better when changes are less disruptive.

Collegial criticism is a new feature of preparing Pretest II. In other respects, the characteristics that were desirable for Pretest I are now almost mandatory: for example, most questions should be closed; code categories for the few open questions should be at least loosely sketched out in advance; respondents should be strangers who resemble the target population; the N should be around 25–75, and more if financially and humanly possible; study staff (including students, if any) should participate in conducting the pretest, along with the best and the brightest of a professional field staff.

To evaluate the second pretest and prepare the final interview schedule, we recommend only one additional procedure. It is to keep the barrier high to any new questions. This is hard discipline. Inevitably, after two pretests, study staff will have some new brainstorms. To fend off these last-minute inspirations, we have tried to stick by a rule that any new question must be given an independent test with some minimum number of respondents. This is not, finally, a thorough test of the question, because the trial is shorn of context. But at least the rule helps to discourage the less-serious people who are unwilling to find the requisite number of respondents. One may lose some splendid ideas by keeping this barrier in place but–given the problem of incorporating untried questions into an otherwise finished questionnaire—the risk seems acceptable.

The complex matter of how the final questionnaire is put into action in the field is beyond the scope of this book. We need only remind ourselves that a questionnaire is not writ in stone. It is *merely* a design, a plan for action and interaction; its execution depends on other directors and actors—able and motivated interviewers and field supervisors, effective procedures of quality control. This means that if investigators do not keep in touch with what is happening in the field, they can lose control of their questionnaire. But at that stage of a survey, responsibilities are always shared and sometimes diffuse, and how they are exercised will vary greatly by particular organizations. As Davis (1964: 231) has pointed out, if survey analysis is an art, it is more like architecture than sculpture or painting. The image seems apt for survey questionnaire design as well. Much painting and sculpture can finally be achieved by a single artist. Architects can make their drawings in solitary confinement, but their buildings take shape only as scores—sometimes hundreds—of other artists,

craftsmen, technicians, take up the task. Survey questionnaire architects are no less dependent on all the others to carry their design into concrete form.

http://dx.doi.org/10.4135/9781412986045.n3