II

Making Research Decisions: Questionnaire Design

Chapter Contents

- Understanding Reliability and Validity
- Levels of Measurement and Why They Matter
- Types of Questions and the Information Each Type Provides
- Ensuring Clarity and Avoiding Bias
- Questionnaire Layout and Design
- Handling "Don't Know Responses
- Design Features That Affect Response Rate
- Final Thoughts

Writing a questionnaire seems simple. As the famous survey researcher G. W. Allport once said, "If we want to know how people feel: what they experience and what they remember, what their emotions and motives are like, and the reasons for acting as they do—why not ask them?" (Selltiz, Jahoda, Deutsch, & Cook, 1959, p. 236).

Unfortunately, just asking them is not as easy as it sounds. To avoid obtaining misleading results, questions must be clear, must elicit honest and reliable answers, and must keep the respondent interested in providing answers. The construction of questions must differ according to whether they are being read or heard, and the researchers can ask only as many questions as respondents have the time and energy to answer. It is easy to write bad questions and difficult to write good ones. Guidelines for questionnaire design typically focus on the importance of clarity, simplicity,

and objectivity. Other important considerations include making questions interesting and the questionnaire logical so that respondents feel motivated to answer carefully.

Given the myriad of details that can make or break a questionnaire, the best questionnaires often turn out to be those you would have written once it is too late and you have the answers to the one you already used. To avoid giving postsurvey regrets, questionnaires need to be pretested with attention to every detail and with members of the intended sample of respondents. Important issues to consider when writing questionnaires include the following:

- Validity and reliability concerns as they relate to the sample, the topic, and the client
- Levels of measurement and why they matter
- Ways to ensure clarity and avoid bias
- Types of questions and how the information each type provides differs
- Questionnaire layout and design to ensure logical flow and visual clarity

UNDERSTANDING RELIABILITY AND VALIDITY

What is a "fast" city? If you are the communication manager for the local chamber of commerce, would you want your city rated as "fast" or as "slow"? National and regional rankings of cities, universities, corporations, and other organizations are published all the time. Depending on your interpretation, "fast" could mean exciting or it could mean stressful. Meanwhile, "slow" could mean boring or it could mean mellow and comfortable. When *Psychology Today* published its feature on fast cities, the things that made a city fast included the length of time it took to be waited on in a bank, the speed at which people walked down the street, the number of people who wore watches, the speed with which people talked, and the rates of coronary heart disease in the city population. A ranking of 1 on this list probably would not make city leaders happy. But how can ratings based on objective data be refuted effectively (Sidebar 11.1)?

The answer is that measures need to be *valid* and *reliable*. A valid measure is one that seems to represent a particular idea in a convincing way. If people generally can agree that the things used to measure something such as a fast city are appropriate, the measures are considered valid. A reliable measure has consistency. If virtually anyone can replicate the study using the same measures and come out with similar answers, the measures are considered reliable. The fast city measures could be attacked as invalid by arguing that coronary heart disease has no relationship to bank teller speed, except perhaps for the bank teller.

SIDEBAR II.I
Measuring a City's Pace

To see if there is any relationship between a city's characteristic pace and its rate of coronary heart disease, we looked at four indicators.

Walking speed: We clocked how long it took pedestrians to move 60 feet along relatively uncrowded streets. To eliminate the effects of socializing, we timed only people walking alone. We also excluded children, pedestrians with large packages or obvious physical handicaps, and window shoppers.

Working speed: We timed how long bank clerks took either to give change, in set denominations, for two $20 bills or to give us two $20 in return for change.

Talking speed: In each city we tape-recorded how long it took postal clerks to explain the difference between regular mail, certified mail, and insured mail. We then calculated their actual "articulation" rates by dividing the number of syllables in the response by the total time it took.

The watch factor: As a simple measure of concern with clock time, we counted the percentage of men and women who were wearing wrist watches.
Individually, each of these measures has its weaknesses: They all tap into special groups, not the city's general population; the second two are confounded by skill and efficiency; and the last is affected by fashion as well as concern with time.

Finally, we created an index of the overall pace of life in each city by giving the four scores equal weight and adding them together. The chart below shows how the cities ranked, from 1st to 36th, in each category.

FAST CITIES, SLOW CITIES, HOW THEY RANK

City	Overall Pace	Walking Speed	Bank Speed	Talking Speed	Watches Worn	CHD*
Boston, MA	1	2	6	6	2	10
Buffalo, NY	2	5	7	15	4	2
New York, NY	3	5	7	15	1	1
Salt Lake City, UT	4	4	16	12	11	31
Columbus, OH	5	22	17	1	19	26
Worcester, MA	6	9	22	6	6	4
Providence, RI	7	7	9	9	19	3
Springfield, MA	8	1	15	20	22	7
Rochester, NY	9	20	2	26	7	14
Kansas City, MO	10	6	3	15	32	21
St. Louis, MO	11	15	20	9	15	8
Houston, TX	12	10	8	21	19	36
Paterson, NJ	13	17	4	11	31	1
Bakersfield, CA	14	28	13	5	17	20
Atlanta, GA	15	3	27	2	36	33

(Continues)

SIDEBAR II.I (Continued)

City	Overall Pace	Walking Speed	Bank Speed	Talking Speed	Watches Worn	CHD*
Detroit, MI	16	21	12	34	2	11
Youngstown, OH	17	13	18	3	30	6
Indianapolis, IN	18	18	23	8	24	22
Chicago, IL	19	12	31	3	27	13
Philadelphia, PA	20	30	5	22	11	16
Louisville, KY	21	16	21	29	15	18
Canton, OH	22	23	14	26	15	9
Knoxville, TN	23	25	24	30	11	17
San Francisco, CA	24	19	35	26	5	27
Chattanooga, TN	25	35	1	32	24	12
Dallas, TX	26	26	28	15	28	32
Oxnard, CA	27	30	30	23	7	34
Nashville, TN	28	8	26	24	33	14
San Diego, CA	29	27	34	18	9	24
East Lansing, MI	30	14	33	12	34	29
Fresno, CA	31	36	25	17	19	25
Memphis, TN	32	34	10	19	34	30
San Jose, CA	33	29	29	30	22	35
Shreveport, LA	34	32	19	33	28	19
Sacramento, CA	35	33	32	36	26	23
Los Angeles, CA	36	24	36	35	13	28

*Lower numbers indicate faster speeds, more watches worn, higher CHD rates. CHD indicates rates of coronary heart disease, adjusted for the median age in each city. From "Type A Cities and Your Heart," by R. Levine, 1989, *Psychology Today.* Copyright 1989 by Sussex Publishers, Inc. Reprinted with permission.

For measures to be valid, the *concept* or idea they represent must be clear and the *operationalizations*, the actual measures themselves, must seem appropriate. To a great extent, determining validity is an exercise in persuasion. Reliability is easier to verify objectively.

Threats to validity and reliability sometimes are subtle, making it important to think carefully about the context in which surveys will be answered and interpreted. The measure itself must seem appropriate, and the scale used to measure it must seem appropriate. For example, operationalizing a concept such as a "livable" city might seem fairly straightforward. *U.S. News and World Report* publishes a feature on this issue every year focusing on characteristics such as crime rates, housing costs, and employment rates. Most people would agree that a more livable city would feature a lower crime rate, lower housing costs, and higher employment rates. But wait: rising housing costs make a city more attractive, not less attractive, to those investing in real estate. So a high score on housing costs could

mean less livable to some people and more livable to others. Meanwhile, some might argue that the most appropriate measure of livability would be people's perceptions of safety and satisfaction, rather than more objective measures of crime rates, numbers of parks, and so on.

Moreover, the measures chosen to operationalize livability might be criticized as insufficient to measure the concept appropriately. Perhaps factors such as average commute times, numbers of cultural events, and the quality of public schools (as indicated by standardized tests? Availability of special services?) also need to be included to provide a valid measure. The three-quarters of the population who do not have school-age children, however, may not consider public school quality a primary determinant of a city's livability. In addition, some people might consider other factors such as nearby access to parkland as critical to quality of life, whereas others might consider a booming nightlife a higher priority. And what about weather? Is a warm temperature an advantage or a disadvantage? It depends on whether you prefer to snow ski or water ski. Thus, to measure a fairly simple idea such as a livable city, the measures chosen must be the following:

- Sufficient in number to represent enough about the concept
- Appropriate as indicators of the concept
- Unambiguous, so that a high score on a measure clearly represents a high level of the concept

Reliability, too, is an important characteristic for measures, with two primary components. First, the indicator of a concept must be replicable, that is, reusable with a similar result. The second component of reliability is how consistently the various operationalizations of a concept measure it. The operationalizations are the ways the researcher measures an idea, such as by counting the number of adults employed in the community during a single year. Consistency of the measures is important because observers tend to find a group of measures more convincing than any single measure. For example, if a city that scores highly on employment rates and cultural events also scores highly on individuals' reports of perceived safety and happiness with the town, these measures as a group representing "livability" can be called reliable. Some people may think employment rates are more important than personal reports of happiness, others may consider personal reports more important, but they all may accept the others' measures if convinced that the measures "hang together" consistently. This type of reliability can be measured statistically.

Rankings such as "most livable cities" can have major repercussions for organizations that score well or poorly or just better or worse than expected. According to Monks and Ehrenberg (1999), educational institutions that receive a less favorable rating from the *U.S. News and World*

Report annual rankings of colleges and universities end up with a lower-quality pool of applicants and have to use bigger financial incentives to attract desirable students. As shown in Table 11.1, rankings of universities can vary dramatically depending on the methods used to develop the ranking scores. As Gater (2002) has asserted, it is nearly impossible to identify, quantify, and measure characteristics that fairly compare colleges and universities because they have wide-ranging sizes, scopes, missions, and disciplinary emphases. Nevertheless, many parents of potential students, students themselves, faculty considering potential employers, and donors considering potential beneficiaries look for rankings to help them in their decision making.

As Table 11.1 shows, rankings change depending on the types of institutions included in the analysis, how the organizations collect data about "quality" or "value," and when they collect the data. Some raters, for example, analyze institutional-level data such as admission and graduation rates. Some include surveys of university administrators. Some include surveys of alumni or current students. In addition, some surveys rely on volunteer samples rather than representative samples. Table 11.1 shows that even the same organizations rank universities differently depending on whether they focus on a particular issue (e.g., academic excellence or cost) or include a range of issues important to most applicants (e.g., academic excellence *and* cost) in the ranking criteria. Meanwhile, other organizations may focus on criteria important only to some, such as political activism. In addition, rankings can change depending on whether they focus exclusively on schools that meet criteria for "national" universities or mix public and private colleges along with public and private universities. Communication managers at educational institutions must understand validity and reliability issues to deal effectively with these rankings, which frequently receive a great deal of news coverage and attention from important constituents. Universities that wish to publicize rankings that make them look good while de-emphasizing those that portray them less positively have to defend the validity and reliability of one survey credibly while attacking the validity and reliability of another.

LEVELS OF MEASUREMENT AND WHY THEY MATTER

Survey questions fall into four general levels, shown in Table 11.2, that dictate how the questions can and should be analyzed. It is important to know your level of measurement because the level dictates the types of statistical tests you can perform on the data. Usually, clients hope to determine how certain beliefs or behaviors relate to other beliefs and behaviors. If measures have been collected at a low level of measurement, this will seriously limit the analyst's ability to investigate relationships of interest.

TABLE II.I

Rankings of Universities Based on Varying Criteria From Different Organizations

Kiplinger "Best Value" (2003)	Peterson's "Solid Value" (2004)	Peterson's "Academics Plus" (2004)	Mother Jones "Activist Campuses" (2003)	Princeton Review "Most Politically Active" (2004)	Princeton Review "Academic Experience" (2004)	U.S. News & World Report "Top National Universities" (2004)
1. UNC Chapel Hill	1. Barry University	1. Clemson	1. U of Puerto Rico, Mayaguez	1. Claremont McKenna College	1. U of Chicago	1. Harvard U
2. U of Virginia	2. Bryan College	2. William & Mary	2. UCLA	2. US Naval Academy	2. Marlboro College	2. Princeton U
3. College of William & Mary	3. Burlington College	3. SUNY Maritime College	3. Spelman College	3. Hampden-Sydney College	3. Reed College	3. Yale U
4. U of Georgia	4. Carson-Newman College	4. US Air Force Academy	4. California Community Colleges	4. Sarah Lawrence College	4. St. John's College (MD)	4. U of Pennsylvania
5. U of Florida	5. Centenary College of LA	5. US Military Academy	5. Rutgers U	5. Georgetown U	5. Swarthmore College	5. Duke U
6. New College of Florida	6. Eastern Mennonite U	6. UC Berkeley	6. Suffolk U	6. US Air Force Academy	6. Haverford College	6. Massachusetts Institute of Technology
7. Georgia Institute of Technology	7. Freed-Hardeman U	7. UC Irvine	7. U of Massachusetts	7. Eugene Lang College	7. St. John's College (NM)	7. Stanford U
8. U of Illinois at Urbana, Champaign	8. Lenoir-Rhyne College	8. UC San Diego	8. Mount St. Mary's U	8. Hampshire College	8. Harvey Mudd College	8. California Institute of Technology
9. Truman State University	9. Rollins College	9. UC Santa Barbara	9. Arizona State U	9. The George Washington U	9. Sarah Lawrence College	9. Columbia U
10. Virginia Polytechnic Institute and State U	10. Shorter College	10. U of Mary Washington	10. UC Berkeley	10. Wesleyan College	10. Carleton College	10. Dartmouth College
Sources Peterson's database	Alumni feedback and institutional data	Alumni feedback and institutional data	Editors' analysis of news reports	Convenience- and volunteer-sample surveys of students	Combination of convenience- and volunteer-sample surveys of students, counselors, parents, and educators staff opinions	Perceptions of administrators (61% response rate) and institutional data

223

TABLE II.2
Levels of Measurement for Credibility

Nominal Level	Which of the following companies do you find credible?
Ordinal Level	Rank the following companies from most credible to least credible, with the most credible company receiving a 5 and the least credible company receiving a 1.
Interval Level	How credible are the following companies? Not at Very all credible credible 1 2 3 4 5
Ratio Level	How many times in the last year have you wondered whether the following companies were telling you the truth? 0 1 2 3 4 5 or more

To choose the appropriate level, questionnaire designers need to consider how they will use the information gathered. Sometimes, for example, a client may wish to know whether people first heard about an organization from the newspaper or from a friend. Other times, however, the organization may need to know how often newspapers and friends are sources of information or how credible the information received from these sources seems. Each of these needs requires a different level of measurement.

The first level of measurement is called the *nominal* level, meaning names or categories of things. This level of measurement is useful when an organization needs to know how many people fit into particular categories. The possible answers for a nominal variable are mutually exclusive and exhaustive. In other words, they have no overlap and include all possible responses. For example, a question assessing gender of the respondent would include "male" and "female" (social scientists consider "gender" a socially constructed identity rather than a category dictated by chromosomes). A question assessing information sources could include "mass media" and "interpersonal sources." Including "mass media" and "newspapers" would be redundant instead of mutually exclusive because the newspaper is a form of mass media. Eliminating "interpersonal sources" or including "friends" but not "coworkers" or "family" would not be exhaustive. Nominal variables can be useful, but little statistical analysis can be performed using this type of variable. They have little explanatory power, because people either fit a category or do not fit. They cannot fit a little bit or a lot.

The second level of measurement is called the *ordinal* level, indicating some meaningful order to the attributes. These questions have answers that are mutually exclusive, exhaustive, and ordered in some way. A popular type of ordinal question is the ranking question, as in "Please rate the following five publications according to how much you like them, with

the best one rated 1 and the worst one rated 5." It would be possible to know which publications do best and worst, but it would not be possible to know whether Publication 2 is liked a lot better than Publication 3 or just a little bit better.

The ranking question often creates problems and generally should be avoided. It not only provides information of limited use but also frequently confuses or frustrates respondents. Ranking is difficult to do and tends to discourage respondents from completing a questionnaire. Sometimes respondents may consider two or more items to be ranked in a tie, and other times they may not understand the basis on which they are supposed to determine differences. When asked to rank the corporate citizenship of a group of companies, for example, respondents may not feel they have enough information on some companies to distinguish them from others. Respondents often rate several things as the same number, rendering their response to the entire question useless to the analyst. If two or more items are tied, they no longer are ranked. The answers no longer are mutually exclusive, which makes the question of less use than even a nominal variable would be.

Organizations that want to rank a group of things may find it better to let rankings emerge from the data rather than trying to convince respondents to do the ranking themselves. They can do this by creating a question or a group of questions that can be compared with one another, such as, "Please rate each of the following information sources according to how much you like them, with a 4 indicating 'a lot,' a 3 indicating 'some,' a 2 indicating 'not much,' and a 1 indicating 'not at all.'" The mean score for each information source then can be used to create a ranking.

The third level of measurement is the *interval* level. This is the most flexible type of measure to use because it holds a lot of meaning, giving it a great deal of explanatory power and lending itself to sensitive statistical tests. As with the previous levels of measurement, the interval measure's responses must be mutually exclusive, exhaustive, and ordered. The order, however, now includes equal intervals between each possible response. For example, a survey could ask people to indicate how much they like a publication on a 10-point scale, on which 10 represents liking it the most and 1 represents liking it the least. It can be assumed that the respondent will think of the distances separating 2 and 3 as the same as the distances separating 3 and 4, and 9 and 10.

Most applied research—and some scholarly research—assumes perceptual scales such as strongly agree–strongly disagree or very important–not important at all can be considered interval-level scales. Purists disagree, saying they are ordinal because respondents might not place an equal distance between items on a scale such as *not at all . . . a little . . . some . . . a lot* in their own minds. Fortunately, statisticians have found that this usually

does not present a major problem. Nevertheless, this is a controversial issue (Sarle, 1994). Researchers must construct such measures carefully and pretest them to ensure that they adhere to the equal-distance assumption as much as possible.

The fourth level of measurement is the *ratio* scale, which is simply an interval scale that has a true zero. This means the numbers assigned to responses are real numbers, not symbols representing an idea such as "very much." Ratio scales include things such as the number of days respondents report reading the newspaper during the past week (0–7 days), the number of minutes spent reading the business section, or the level of confidence they have that they will vote in the next presidential election (0–100% likelihood of voting). This type of scale is considered the most powerful because it embodies the most meaning.

TYPES OF QUESTIONS AND THE INFORMATION EACH TYPE PROVIDES

Various strategies exist for eliciting responses at each level of analysis. Keep in mind that respondents will find complex questions more difficult and time consuming to answer. As a result, the survey designer has to make trade-offs between obtaining the most meaningful information and obtaining any information at all. For example, a lengthy and complex mail survey may end up in the trash can more often than in the return mail. Even if the questions are terrific, the few responses that come back may not compensate for the loss of information resulting from the number of nonresponses.

Likewise, people answering a telephone survey will find complicated questions frustrating because they tend to comprehend and remember less when hearing a question than when reading a question. This is the reason telephone surveys often use generic response categories such as the Likert-scale type of response, in which the answer range is: strongly agree, agree, neutral, disagree, strongly disagree. People on the telephone often have distractions in the background and other things they would rather be doing, making them less involved in the survey. This makes it easier for them to forget what a question was, what the response options were, or how they answered a previous question on the survey.

For ease of response and analysis, questions on surveys usually are closed ended, meaning respondents choose their favorite answer from a list of possibilities. Open-ended questions, which ask a query but provide space for individual answers instead of a response list, invite more information but are often skipped by respondents and are time consuming to analyze afterward. As a result, surveys typically limit the number of open-ended questions to 2 or 3 of 50. The primary types of closed-ended

questions are the checklist, ranking scale, quantity/intensity scale, Likert-type scale, frequency scale, and semantic differential scale.

Checklists

The checklist is a nominal variable, providing categories from which re-spondents can choose. They can be asked to choose only one response, or all that apply.

Checklist example:

Please indicate whether you are male or female:
☐ Male   ☐ Female

Please indicate which of the following publications you have read this week (check all that apply):
☐ Newspapers ☐ News magazines ☐ Other magazines ☐ Newsletters

Ranking Scales

Ranking scales are ordinal variables, in which respondents are asked to put items in the order they think is most appropriate. Ranking scales are problematic because they incorporate a series of questions into a single item, requiring respondents to perform a complex and often confusing task. They must decide which choice should come first, which should come last, which comes next, and so on until the whole series of comparisons is completed.

Ranking example:

Please rank the following issues according to how important they are to your decision about a congressional candidate this year. Put a 1 by the issue most important to you and a 5 by the issue least important to you:
☐ Taxes
☐ Economy
☐ Environment
☐ Education
☐ Crime

Questionnaire designers can help respondents answer a ranking ques-tion by breaking it into a series of questions, so that the respondents do not have to do this in their heads. Although this method makes it easier for respondents to answer ranking questions, it uses a lot of valuable questionnaire space.

Among the following issues, which is the most important to your decision about a congressional candidate this year?
☐ Taxes
☐ Economy
☐ Environment
☐ Education
☐ Crime

Among the following issues, which is the *next most* important to your decision about a congressional candidate this year?
☐ Taxes
☐ Economy
☐ Environment
☐ Education
☐ Crime

Among the following issues, which is the least important to your decision about a congressional candidate this year?
☐ Taxes
☐ Economy
☐ Environment
☐ Education
☐ Crime

Quantity/Intensity Scales

The quantity/intensity scale is an ordinal- or interval-level variable, in which respondents choose a location that best fits their opinion on a list of options that forms a continuum.

Quantity/intensity example:

How much education have you completed?
☐ Less than high school degree
☐ High school diploma or GED
☐ Some college (no degree; may be currently enrolled)
☐ Vocational certificate or associate's degree
☐ College graduate (bachelor's degree)
☐ Some graduate work (no degree)
☐ Master's or other graduate professional degree
☐ Doctoral degree

Likert-Type Scale

The most frequently used scale is known as the Likert scale.

Likert scale example:

Please indicate whether you strongly agree, agree, disagree, or strongly disagree with the following statement:
The Bestever Corporation is responsive to public concerns
☐ Strongly agree
☐ Agree
☐ Disagree
☐ Strongly disagree

Other variations on the Likert scale appear frequently on questionnaires. Some popular response ranges include the following:

- Very satisfied/Somewhat satisfied/Somewhat dissatisfied/Very unsatisfied
- Strongly oppose/Oppose/Support/Strongly support
- Very familiar/Somewhat familiar/Somewhat unfamiliar/Very unfamiliar
- A lot/Somewhat/Not much/Not at all
- A lot/Some/A little/None
- Always/Frequently/Seldom/Never
- Often/Sometimes/Rarely/Never
- Excellent/Good/Fair/Poor

Quantity/Intensity example:

Please indicate if the following reasons have been *very important* (VI), *somewhat important* (SI), *not very important* (NVI), *or not at all important* (NAI) to your decision whether to give to the Most important Association in the past.

The tax benefits resulting from giving   VI SI NVI NAI
Because you like being involved with the MA VI SI NVI NAI

Another variation of the Likert scale is known as the *feeling thermometer*, which can be modified to measure levels of confidence, degrees of involvement, and other characteristics. The feeling thermometer as presented by Andrews and Withey (1976) used 10- or 15-point increments ranging from 0 to 100 to indicate respondents' warmth toward a person, organization, or idea.

Feeling thermometer example:
100   Very warm or favorable feeling
85   Good warm or favorable feeling
70   Fairly warm or favorable feeling

60 A bit more warm or favorable than cold feeling
50 No feeling at all
40 A bit more cold or unfavorable feeling
30 Fairly cold or unfavorable feeling
15 Quite cold or unfavorable feeling
 0 Very cold or unfavorable feeling

Yet another variation of the Likert scale uses pictorial scales, which can be useful for special populations such as children, individuals lacking literacy, or populations with whom language is a difficulty. Often, the scales range from a big smiley face (very happy or positive) to a big frowny face (very unhappy or negative), or from a big box (a lot) to a little box (very little).

Frequency Scales

The frequency scale is an interval or ratio scale. Instead of assessing how much a respondent embraces an idea or opinion, the frequency question ascertains how often the respondent does or thinks something.

Frequency example:
How many days during the past week have you watched a local television news program?
☐ 7 days
☐ 6 days
☐ 5 days
☐ 4 days
☐ 3 days
☐ 2 days
☐ 1 day
☐ 0 days

About how many times have you visited a shopping mall during the past month?
☐ 16 times or more
☐ 11–15 times
☐ 6–10 times
☐ 1–5 times
☐ 0 times

Sometimes frequency scales are constructed in ways that make it unclear whether equal distances exist between each response category, which makes the meaning of the measure less clear and the assumption of interval-level statistical power questionable.

Frequency example:

In the past 6 months, how many times have you done the following things?

	Never	1–2 times total	3–4 times a month	1–3 times a month	1 time a week	More than once a week
Been offered an alcoholic beverage						
Attended a party where alcohol was served						
Drank an alcoholic beverage						
Had four or more drinks in a row						
Rode with a driver who had been drinking alcohol						
Got sick from drinking alcohol						

Semantic Differential Scales

The semantic differential scale is an interval-level variable, on which respondents locate themselves on a scale that has labeled end points. The number of response categories between the end points is up to the questionnaire's designer, but it is useful to have at least four response options. More options make it possible for respondents to indicate nuances of opinion; beyond a certain point, which depends on the context, a proliferation of categories becomes meaningless or even confusing. An even number of response categories forces respondents to choose a position on the issue or refuse to answer the question, whereas an odd number of response categories enables respondents to choose the neutral (midpoint) response.

Semantic differential example:

Please rate your most recent experience with the Allgetwell Hospital staff:

Incompetent	_ _ _ _ _ _ _ _ _ _ _ _ _	Competent
Impolite	_ _ _ _ _ _ _ _ _ _ _ _ _	Polite
Helpful	_ _ _ _ _ _ _ _ _ _ _ _ _	Unhelpful

Semantic differential questions can provide a lot of information in a concise format. Written questionnaires especially can include a list of semantic differential items to assess the performance of an organization and its communication activities. Because this type of question includes information as part of the answer categories themselves, some consider these

items more valid than Likert-scale items. For example, a Likert-scale question asking if the staff seemed competent could bias respondents who do not want to disagree with the statement, whereas a semantic differential question that gives equal emphasis to "competent" and "incompetent" as end points may elicit more honest answers. Psychologists have demonstrated that agree/disagree question batteries can suffer from acquiesence (Warwick & Lininger, 1975, p. 146), which occurs when people hesitate to express disagreement.

Measuring Knowledge

Often, an organization wants to determine what people know about a topic. One option is to give a true/false or multiple-choice test. The advantage of the multiple-choice test is that, if carefully written, it can uncover misperceptions as well as determine the number of people who know the correct answers. The wrong answers, however, must be plausible. A second option is to ask open-ended questions in which people must fill in the blanks. This requires a lot of work from the respondent but potentially provides the most valid answers. A third option is to ask people how much they feel they know, rather than testing them on what they actually know. This technique seems less intimidating to respondents. Finally, follow-up questions can ask people how sure they are of a particular answer.

ENSURING CLARITY AND AVOIDING BIAS

Wording can affect the way people respond to survey questions. As a result, it is important to pretest for clarity, simplicity, and objectivity. Using standardized questions that have been pretested and used successfully can help prevent problems. Of course, because every communication issue has unique characteristics, standardized batteries of questions suffer from lacking specific context. Often, a combination of standard and unique items serve the purpose well. When designing questions, keep the following principles in mind:

1. *Use words that are simple, familiar to all respondents, and relevant to the context*. Technical jargon and colloquialisms usually should be avoided. At times, however, the use of slang may enhance the relevance of a questionnaire to a resistant target public. For example, asking college students how often they "prefunk" could elicit more honest responses than asking them how often they "use substances such as alcohol before going out to a social function," which is both wordy and could have a more negative connotation to the students than their own terminology. When using specialized terms, it is important to pretest them to ensure the respondents understand them and interpret them as intended. Try to choose words that

will not seem patronizing, class specific, or region specific. Choosing to ask about "pasta" instead of "noodles" when assessing audience responses to messages about an Italian restaurant could alienate some respondents who think "pasta" seems pretentious.

2. *Aim for precision to make sure the meaning of answers will be clear.* Avoid vague terms. For example, the word often may mean once a week to some people and twice a day to others. *Recently* could mean "this past week" or "this past year." Terms such as here and there do not set clear geographic parameters.

Do not leave room for interpretation. People responding to a question about how often in the past year they have donated to a charitable organization may consider each monthly contribution to a church a separate donation. The sponsor of the survey, however, may have intended for respondents to indicate to how many different organizations they have made donations during the past year. Avoid hypothetical questions because people often are not very good at, or may have trouble being honest about, predicting their own behavior. Direct questions about cause or solutions also may be difficult for respondents to answer validly (Fowler, 1995). It is better to let the reasons for things emerge from the data analysis by looking at the associations between attitudes and behaviors instead of asking respondents to make those associations for the researcher.

Finally, because the use of negatives in a question can result in confusion, use positive or neutral statements, providing respondents with the opportunity to disagree. For example, instead of asking, "Do you think the Neverong Corporation should not change its partner benefits policy?" a survey can ask, "Do you think the Neverong Corporation's partner benefits policy should change or stay the same?"

3. *Check for double-barreled questions*. Each question must cover only one issue. Asking if respondents rate staff as "polite and efficient," for example, makes it impossible for respondents to choose "polite but inefficient" or "impolite but efficient" as their answer. Sometimes a double-barreled question is subtle, and the problem occurs because a phrase requires respondents to embrace an assumption they may not hold. For example, asking "How likely are you to use this service on your next visit to Funpark?" assumes there will be a next visit.

4. *Check for leading or loaded questions*. A leading question prompts the respondent in one direction instead of treating each possible response equally. Asking the question, "How much did you enjoy your visit?" leads respondents in the direction of a positive answer, whereas the question, "How would you rate your visit?" allows enjoyment and disappointment to be equivalent answer categories, making it easier for respondents to choose the negative answer.

A loaded question biases the answer through the use of emotionally charged words, stereotypes, or other words that give a subtle charge to a

phrase. Loading can occur in the question or in the answer. For example, the question given earlier asking respondents to indicate which issues are most important to them in an upcoming election mentions only some of the possible alternatives about which voters may care. Health care, abortion, social security, welfare, agricultural issues, and race/gender equality are among the many issues not even mentioned. In addition, loading can occur by using words that have positive or negative connotations, such as "unwed moms" versus "single mothers." Loading also can occur in frequency scales. Asking people whether they had 0, 1, 2, 3, 4, 5, 6, 7 or more alcoholic drinks during the past week, for example, gets more people to acknowledge having 3 or 4 drinks than asking people whether they had 0, 1, 2, 3 or more alcoholic drinks during the past week (Fowler, 1995). People often feel marginalized by picking what seems like an extreme response.

5. *Check for social desirability effects*. Some people find it difficult to express an opinion or report a behavior they think is inconsistent with what most other people think or do. Some also find it difficult to give a response they think the surveyor disagrees with or disapproves of. It is well documented, for example, that a higher percentage of people claim to have voted in an election than actually turn out at the polls. Try to write questions so that people find it easy to give a negative response.

One technique for reducing social desirability bias is to include an introduction to a sensitive question that makes any response seem normal and acceptable. For example, Fowler (1995) noted that asking people if they own a library card can seem threatening because a "no" response could be perceived as a lack of interest in reading, which might seem socially unacceptable. As a result, Fowler suggested the following introduction: "Many people get books from libraries. Others buy their books, subscribe to magazines, or get their reading material in some other way. Do you have a library card now, or not?" (p. 36).

6. *Provide enough context to enable people to respond realistically or remember accurately*. On the whole, questions should be as brief as possible so that they can be digested with the least effort. Nevertheless, the goal of questionnaire design is to construct questions such that answers will provide the most meaningful information possible. As a result, adding some context can be useful. It helps, for example, to ask people to recall behaviors over a limited time or from a recent time, such as during the past week.

In general, questionnaire designers must avoid yes/no items. Besides providing information of limited usefulness for statistical analysis (dichotomous questions are nominal variables), this type of question leaves no room for a respondent to answer "maybe" or "well, it depends." Answers to dichotomous questions, as a result, can be misleading. Similarly, questions usually need to avoid "always" and "never" as categories. "Almost always" and "almost never" give people the opportunity to be more context specific.

QUESTIONNAIRE LAYOUT AND DESIGN

Most discussions of survey design focus on how to construct the questions themselves, but other aspects of design, such as how items appear on a page or the order in which questions appear, also can make a difference to respondents.

Ease of Reading

It helps to give respondents "chunks" of questions at a time. A series of questions without a break can become boring and confusing. People may get lost in a written questionnaire that has 10 items in a row, for example, checking off their responses to Question 6 on the line for Question 5. Assessing respondents' interest in different types of university-related news, for example, is difficult to follow in a continuous format.

The following are topics that might be covered in a publication from Central State University. For each one, please tell me whether you are *very interested* (VI), *somewhat interested* (SI), *not very interested* (NVI), or *not at all interested* (NAI) in receiving information about each topic*:

1. The university's branch campuses	VI SI NVI NAI RF/DK
2. Student accomplishments	VI SI NVI NAI RF/DK
3. The financial needs of the university	VI SI NVI NAI RF/DK
4. The work of the administration	VI SI NVI NAI RF/DK
5. How donations are being used	VI SI NVI NAI RF/DK
6. News about teaching	VI SI NVI NAI RF/DK
7. Athletic accomplishments	VI SI NVI NAI RF/DK
8. News about university research	VI SI NVI NAI RF/DK
9. University nostalgia and history	VI SI NVI NAI RF/DK
10. News about alumni	VI SI NVI NAI RF/DK
11. News about campus life	VI SI NVI NAI RF/DK

*RF/DK = refused or don't know.

Questions are easier to answer in chunks. Generally, chunks of three or four items at a time work well.

The following topics that might be covered in a publication from Central State University. For each one, please tell me whether you are *very interested*, *somewhat interested*, *not very interested*, or *not at all interested* in receiving information about each topic*:

12. The university's branch campuses	VI SI NVI NAI RF/DK
13. Student accomplishments	VI SI NVI NAI RF/DK
14. The financial needs of the university	VI SI NVI NAI RF/DK
15. The work of the administration	VI SI NVI NAI RF/DK
16. How donations are being used	VI SI NVI NAI RF/DK
17. News about teaching	VI SI NVI NAI RF/DK

18. Athletic accomplishments   VI SI NVI NAI RF/DK
19. News about university research VI SI NVI NAI RF/DK
20. University nostalgia and history VI SI NVI NAI RF/DK

21. News about alumni   VI SI NVI NAI RF/DK
22. News about campus life   VI SI NVI NAI RF/DK
* RF/DK = refused or don't know.

Respondents on a telephone survey also can become fatigued by a long list of items and will benefit from a break during which the surveyor gives a new introduction, even if the questions in the next section do not focus on anything new.

Example from a phone survey:

OK, now I need to know if you [READ SLOWLY] *strongly agree* (SA), *agree* (A), *disagree* (D), or *strongly disagree* (SD) with each of the following statements about politics and the media.* [REPEAT CATEGORIES AS NECESSARY.]

	(5)	(4)	(3)	(2)	(1)	(9)
19. The media rarely have anything new to say.	SA	A	n	D	SD	RF/DK
20. I'm interested in campaigns and election information.	SA	A	n	D	SD	RF/DK
21. The news media only pay attention to bad news about political issues and candidates.	SA	A	n	D	SD	RF/DK
22. Candidates for office are interested only in people's votes, not in their opinions.	SA	A	n	D	SD	RF/DK
23. My vote makes a difference.	SA	A	n	D	SD	RF/DK

This won't take much longer and we really appreciate your help. These next few questions also are about politicians and the media. Do you *strongly agree* (SA), *agree* (A), *disagree* (D), or *strongly disagree* (SD) that:

	(5)	(4)	(3)	(2)	(1)	(9)
24. Politicians are out of touch with life in the real world.	SA	A	n	D	SD	RF/DK
25. I pay attention to campaign and election information.	SA	A	n	D	SD	RF/DK
26. There's often more to the story than you hear in the news.	SA	A	n	D	SD	RF/DK
27. Political campaigns are too mean spirited.	SA	A	n	D	SD	RF/DK
28. I actively seek out information concerning the government and politics.	SA	A	n	D	SD	RF/DK

29. I have a say in what the   SA A n D SD RF/DK
    government does.
* n = neutral; RF/DK = refused or don't know.

It also is important to remember how easily people answering telephone surveys can get lost. Because they do not see the questions, they can forget what a series of questions is about or what the response options are. In addition, changing response options frequently will slow them down and may make it difficult for them to keep track of what they are supposed to be doing. Forcing them to slow down can help improve the validity of answers by making sure they think carefully about their answers, but it also can hurt validity by causing utter confusion. Pretesting, over the phone, is essential.

Clarity of Graphics

Recent work by Christian and Dillman (2004) has shown that respondents to self-administered surveys pick up important cues from the visual design of survey questions and answers. For example, Christian and Dillman demonstrated that respondents become confused if a scale is broken up into two columns instead of being presented in a single row or in a single column.

Clear:

Overall, how would you rate the quality of education that you are getting at WSU?
☐ Excellent
☐ Very Good
☐ Good
☐ Fair
☐ Poor

More confusing:

Overall, how would you rate the quality of education that you are getting at WSU?
☐ Excellent   ☐ Good   ☐ Poor
☐ Very Good   ☐ Fair

Directionality and Response Set

Another issue that can affect validity is known as *directionality* and refers to the order in which response categories are presented. It helps both

respondents and analysts to associate negative opinions with lower numbers and positive opinions with larger numbers. Some questionnaires, for example, ask respondents to choose a numbered response instead of checking a box. This can make a questionnaire easier to read—it makes the middle of a scale more obvious—and it also makes data entry easier because the computer usually has to receive numbers.

Example of numbered responses on a written survey:

How important are the following for helping you choose your preferred candidates or issues?

	Not at all important						Very important
Newspapers	1	2	3	4	5	6	7
Radio	1	2	3	4	5	6	7
Television	1	2	3	4	5	6	7
Magazines	1	2	3	4	5	6	7
Friends	1	2	3	4	5	6	7
Family	1	2	3	4	5	6	7

Questionnaire designers also need to pretest for response set, which is a form of response bias. If response categories always have the negative response first (on the left if on a written questionnaire) and the positive response last (on the right), people may begin to answer questions too quickly, without thinking them through. Response set can make a group of questions seem related simply because they appear near each other on the questionnaire. In other words, people may have marked similar answers on the questions out of convenience or habit, instead of because they thought deeply about each question and agreed that a similar answer applied to each.

Using "Skip Patterns" Effectively

Sometimes a question will not apply to all respondents. In this case a screening question, or a series of screening questions, is used. This series of questions is known as a *skip pattern*. On a written questionnaire, it is important to make it clear when and which questions should be skipped. Often instructions such as "GO TO Q. 6" appear next to the appropriate response to the screening question. Sometimes, questionnaires include graphic elements such as boxes or arrows to guide respondents. Such devices can make a questionnaire look cluttered and confusing, however, so skip patterns need to be pretested carefully. Christian and Dillman (2004) found that directional arrows help prevent confusion, as does placing instructions

to skip a question if it does not apply *before* the response options instead of after them. Their example looked something like this:

Clear:

A. Have one-on-one meetings with professors contributed significantly to your WSU education?

If you haven't had many one-on-one meetings, just skip to Question 9.
☐ Yes
☐ No

Confusing:

A. Have one-on-one meetings with professors contributed significantly to your WSU education?
☐ Yes
☐ No
If you haven't had many one-on-one meetings, just skip to Question 9.

HANDLING "DON'T KNOW" RESPONSES

One subtle but major issue in the use of surveys is how to handle people who do not have an opinion on a particular question. One of the problems with survey research is that most people do try to offer an opinion for the researchers, even if they must manufacture an opinion on the spot. For example, the average American respondent queried about the economic situation in Mozambique probably knows little about Mozambique's economy, unless it has been in the news. Few people knew much about Kosovo until Slobodan Milosevic decided to "ethnically cleanse" the province. Nevertheless, if asked for an opinion many people offer one, even though some may decline to answer the question. Such opinions, based on little or no information, mean little because they are unstable. They are pseudo-data, not real data.

Researchers must be ready to handle respondents' potential lack of opinion. The most common way is to include a "don't know" response among the answer categories. The drawback of making it easily available is that respondents may be tempted to use it. Making it subtly available is easy to do on a telephone survey because the interviewer can be instructed not to read that option. On a written survey, the respondents will see the option if it is available. Respondents will use the "don't know" option for one of two reasons: either they truly do not know, or they do not feel like answering the question. To prod people who are not motivated to think about an issue to report an opinion, even if it is top of the mind, surveyors eliminate the

"don't know" option, forcing respondents to leave the question entirely blank if they do not want to answer it.

It is important to consider two issues here. The first is that "don't know" can be a meaningful response, of great usefulness to the communication manager. For example, if a large proportion of participants respond "don't know" to a question about corporate reputation, the organization can conclude that it does not have a bad reputation even if it does not have a good reputation. In other words, the company may learn that instead of running a persuasion campaign, it needs to launch an awareness campaign. This frequently is the case, but organizations must be skilled at collecting and interpreting "don't know" responses to make the appropriate diagnosis.

Another important issue about "don't know" responses is that "don't know" cannot be interpreted the same way as "neutral." Likert scales, for example, often feature a neutral category, which can tempt people to avoid taking a position on an issue. Nevertheless, "neutral" does not necessarily mean the respondent lacks information on which to base an opinion. A neutral opinion is an opinion. Responding "neutral" to a question about providing child care in the workplace, for instance, may mean "I don't care; this doesn't affect me," or "I am satisfied either way," rather than "I have no information on which to base an opinion." Both options can be made available to respondents to avoid misinterpreting the findings. Another way to handle the possibility of "don't know" responses is to provide information in the introduction to a question that gives the respondent background on which to base an opinion. This has important advantages and disadvantages. For example, three members of the state legislature in Washington state included the following questions on a survey of their constituents:

About 9 in 10 Washington voters in a recent nonpartisan survey said education was their number one issue this year. About 2 in 3 people surveyed said education was underfunded and worse than it was 4 years ago. How would you address this?

Divert funding from other areas of state government and put it into higher education?
YES   NO

Increase enrollment and on-campus opportunities at our state's colleges and universities?
YES   NO

Increase enrollment opportunities at branch campuses?
YES   NO

Increase the number of courses available to students at off-campus sites via television, e-mail, and the Internet?
YES   NO

Build more classrooms, laboratories, and other facilities for off-campus instruction?
YES   NO

The value of this strategy is that it gives a client the opportunity to see how respondents will react to an issue of emerging importance, about which they may not yet know much. This can help the client respond to the issue effectively.

The risk associated with this strategy is that it can bias the question in the direction of the information selected for inclusion. Some organizations do this intentionally on bogus questionnaires sent out as fund-raising appeals or to attract media attention through "created" opinion. This is a blatantly unethical practice that is denounced by the American Association of Public Opinion Researchers, and it violates the principles of the PRSA code of professional ethics (Sidebar 11.2).

SIDEBAR 11.2
PUSH POLLS
(Not to be confused with legitimate polling)

What is a push poll?

- ➢ A push poll is an insidious form of negative campaigning disguised as a political poll that is designed to change opinions, not measure them.

How do push polls differ from legitimate political polls?

- ➢ Legitimate polls measure existing opinion among representative samples of the public and voters.
- ➢ Push polls contact large numbers of voters in order to change their opinions.
- ➢ Legitimate polls accurately describe candidate characteristics in order to understand voter reactions.
- ➢ Push polls frequently distort candidate characteristics in order to influence voters.
- ➢ Push polls go beyond the ethical boundaries of political polling and bombard voters with problematic statements about candidates in an effort to manufacture negative voter attitudes.

For example, push polls mostly ask questions like:
"Would you be more or less likely to vote for (NAME OF RIVAL CANDIDATE) if you knew he had avoided the draft / falsified his resume / been arrested / gone through bankruptcy / patronized prostitutes / failed to pay child support / failed to pay income taxes?

(Continues)

SIDEBAR II.2 (Continued)

How do you spot a push poll?

> ➤ The organizations conducting these "polls" are not usually recognized as professional pollsters.
> ➤ Push polls typically call thousands of people. The people called are not a representative sample of voters. Instead, they're people who are targeted because they're thought to be undecided voters or supporters of a rival candidate.
> ➤ The truth of the questions is stretched or fabricated.
> ➤ Usually people's answers are not tabulated; the intent is to create a negative effect on potential voters.

What is the position of the American Association for Public Opinion Research on push polls?

> ➤ AAPOR Councils have repeatedly warned members and the public about the iniquity of push polls.
> ➤ The 1996 and 2000 Councils issued formal condemnations of push polls.
> ➤ AAPOR has reacted to complaints about suspected push polls and made investigations.
> ➤ AAPOR urges its members and the media to uncover push-polling and help us alert the public.

How can you help in combating push polls?

> ➤ Attempt to get the name and location of the organization doing the "interviewing."
> ➤ Ask about the sponsors, the number of people called, the questions asked, and how the information from the poll is being used.
> ➤ Contact AAPOR at AAPOR-info@goAMP.com.

Some organizations do this in hopes of educating the public; the problem is that only a small sample of the public is answering the survey. Surveys are opportunities to gather unbiased information to guide a campaign. Still others use them to see if providing certain information on a survey can change people's responses. This type of survey is known as a "push poll" and often is used by political campaigns. Not only is the information provided on such "surveys" biased in favor of the sponsoring candidate but also the survey often ends with a fund-raising appeal. Turning surveys into vehicles for persuasion and fund-raising defeats the purpose of collecting objective data and compromises the reputation of surveyors trying to do authentic research. It is no wonder, with such unethical practices going on, that a declining number of people agree to answer market research (or any other) surveys.

DESIGN FEATURES THAT AFFECT RESPONSE RATE

Every survey requires an introduction to inform respondents of the study's purpose. The introduction also represents an opportunity to motivate respondents to cooperate. Mail surveys and telephone surveys have different introductions, specially suited to the context. Nevertheless, the primary information to include in an introduction remains the same and includes the following:

- What the study is about
- Who the sponsor is
- How the results will be used
- Why respondents should be a part of the study (how they were chosen and why they should care about this)
- The extent to which their responses will be confidential or anonymous.

Anonymous means that no one, including the researchers, will know respondents' identities. *Confidential* means that the researchers will have identifying information for some purpose, such as matching respondents from the same household or calling back to gather more information later, but their identities will be kept secret from everyone else. Some organizations have "human subjects committees" or "institutional review boards" that must approve of study materials and may require additional elements as well.

The Mail Survey Introduction

Mail surveys must have a cover letter introducing the study. It should be brief (never longer than one page), on letterhead (to identify the sponsoring institution and lend credibility to the study), include a contact person for questions or concerns, indicate how long the survey will take to fill out (be honest), and make it clear how to return the questionnaire (providing a self-addressed, stamped, return envelope is essential). The respondents must be thanked for cooperating and told how important their participation is. It also helps to give respondents a deadline for returning the questionnaire, of no more than a week, so that they do not forget about it. Finally, the letter may include information about an incentive provided for cooperating with the study.

The Telephone Survey Introduction

Introductions by telephone need to be even shorter than those by letter. If possible, keep the introduction to two sentences. In those two sentences,

identify the sponsor and the purpose of the study, the way respondents have been selected, the length of time required for answering the survey, and the importance of the respondent's part in the study. It also helps to have the callers identify themselves by name so that respondents do not think the caller has anything to hide. At the end of the introduction, proceed directly to the first question on the survey. Do not ask permission to begin explicitly because this invites participants to tell you that you have called at an inconvenient time. Instead, solicit their permission to begin through the explanation of the survey and the assurance that their answers are important. This introduction makes it obvious that you plan to ask questions and that you hope they will answer. Because they probably will be skeptical of your motives and more interested in whatever they were doing before you called, you want to get right to the first question on the survey to gain their confidence and capture their interest. You want to make the first question nonthreatening, general, and interesting so that the respondents will think the questionnaire is easy and worth their time (Sidebar 11.3).

Nuances can make a difference. For example, identifying the organization at the outset gives the project more credibility and can make it clear that this is not a sales call. In addition, calling the project a *study* instead of a *survey* makes it sound less threatening and makes it clear that this will be not be one of those sales calls masquerading as a survey. In addition, saying you have called them long distance (if you really are) or that your study is being sponsored by a university or respected institution can make their selection and the study itself seem more special (Dillman, 2000).

Prenotification Cards or Letters

The use of prenotification cards to tell respondents that a survey will be happening soon helps boost response rates. If they understand the purpose

---

SIDEBAR 11.3
A Standard Telephone Introduction With a Screening Question

Hello, my name is _____. I'm calling long distance from _____. We're conducting a study of _____ and I have a few questions for a registered voter 18 years of age or older.

Are you registered to vote?

[IF YES: BEGIN SURVEY. **DO *NOT* STOP TO ASK PERMISSION TO BEGIN.**]

[IF NO: Is someone else in your household registered to vote? May I speak with that person please? BEGIN AGAIN WITH NEW RESPONDENT.]
IF NO: Thanks for your time. Good-bye.

of the study ahead of time, they will be less likely to throw away an enve-
lope that looks like a direct mail solicitation or to hang up on a call from an
individual who sounds like a telephone solicitor. Letters or postcards both
suffice, but postcards offer the advantage of being readable at a glance and
less costly to mail. Envelopes may go unopened into the trash.

Follow-up Mailings

Most mail surveys achieve a wide-ranging 5% to 40% response rate with
a single mailing (Wimmer & Dominick, 2006). Research has found, how-
ever, that reminders can boost response rates to 75% or better. It may take
four reminders to achieve a 75% return, with each reminder netting fewer
responses. As a result, the research manager needs to decide the extent to
which the cost of reminders is worth the additional responses. Generally,
each reminder garners half the number of responses that the previous mail-
ing achieved. As a result, a single reminder can increase a 30% response
rate to a 45% response rate or a 40% response rate to a 60% response rate.

Incentives

It is difficult to give telephone respondents a concrete reward for their par-
ticipation, although credit card companies have tried promising a bonus
credited to the respondent's account in return for answering a survey.
Mail surveys, however, frequently include incentives. Usually the incen-
tive is provided ahead of time to motivate the person to respond, instead
of afterward as a reward. Monetary incentives ranging from a penny to
$5 are especially popular, with amounts over $10 rare. Other incentives
can include a gift certificate or product samples. Organizations sometimes
promise donations to a charity in hopes of appealing to respondents' sense
of altruism, as illustrated in Figure. 11.1. Questionnaires returned in per-
son can include a raffle ticket that can be deposited in a separate container
alongside the questionnaire container to preserve the anonymity of the
questionnaire.

Sensitive Questions

Sensitive questions should never appear at the beginning of a survey. In-
stead, the most sensitive questions come at the end so that respondents
who may quit the study because of a particularly offensive question will
already have answered most of the questionnaire. This is why demographic
questions almost always appear at the end of a survey. Many people es-
pecially dislike identifying their income levels, ethnic background, and
age.

General Manager

200 Liberty Street
New York, NY 10281
Subscriber Services 1-800-JOURNAL
Product/Service Information 1-800-832-1234
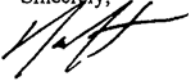
THE WALL STREET JOURNAL.

September 20, 1999

DOWJONES

Dear Wall Street Journal Subscriber,

We would like to better understand your business news and financial information needs. Would you please complete this brief survey about your use of THE WALL STREET JOURNAL and your professional and personal interests? Your answers will help us do a better job of developing products and services that meet your needs.

For your convenience, we have provided a postage-paid envelope in which to return this survey. As a token of our appreciation we have enclosed a dollar bill, which may brighten the day of a child you know. We appreciate your willingness to participate in this survey and thank you in advance for your cooperation.

Sincerely,

General Manager

P.S. Occasionally, a select group of Wall Street Journal advertisers wish to make special offers (such as discounted services or previews of new products) specifically to subscribers who respond to this survey. Please check this box if you *do not* wish to receive any of these offers: ☐

FIG. 11.1. Sample cover letter with incentive. Used with permission from *The Wall Street Journal*. Copyright © *The Wall Street Journal*.

Encouragement

Because respondents to telephone surveys cannot see the questionnaire, they will be worrying about how long the interruption will take. If the questionnaire seems to go on too long, they will lose interest. As a result, it helps to thank them every so often for continuing, as well as assuring them that the end will come soon. For example, one phone survey about the media and politics includes the phrase, around Question 23, "This won't take much longer and we really appreciate your help. These next few questions.... " Then, anticipating increasing impatience from the respondent, the survey includes more encouragement at Question 30, "I have just a few more questions about your use of the media ... " before beginning the introduction for the next set of questions. Before the demographic questions

at the end of the survey, a last bit of encouragement ensures respondents that these are the final set of queries.

Response Set

When constructing a questionnaire, it is important to scatter questions that measure a concept instead of clustering them together. The reason for this is that people's answers can suffer from *response set*, which means they answer a set of questions similarly because they are answering too quickly or not thoughtfully enough rather than because they think similarly about each question in the set. Response set can make measures look statistically reliable but can render them meaningless or invalid. When commissioning a survey from a research firm, managers should make sure the outfit's findings provide both reliable and meaningful information.

Length

No perfect length exists for a questionnaire, although a good bit of research focuses on the topic. For example, people seem more likely to participate in a mall-intercept survey limited to a 5″ × 8″ card. It is difficult to keep people on the phone longer than 5 to 8 minutes, which means a telephone survey of more than 40 to 50 questions will suffer from attrition as people begin to hang up. Questions appearing at the end of the survey will end up with a smaller sample of responses than questions appearing at the beginning of the survey. In mail surveys, a longer mail survey will receive less cooperation than a shorter survey, but the number of pages is not always the deciding factor. For example, respondents may prefer the ease of reading and feeling of accomplishment that comes from having fewer questions and more white space, even if that necessitates using more pages. Some survey designers find a two-column format makes questionnaires more reader friendly. Disagreement exists about the use of single-sided versus double-sided printing. Double-sided printing cuts the number of pages but can confuse respondents, who may end up accidentally skipping pages. As with everything else, it helps to pretest surveys using different formats to see which works the best.

Interviewer Directions

Untrained or confused interviewers can ruin a survey. Interviewers must sound enthusiastic, polite, and confident. They also need to present questions clearly, which means they need to enunciate carefully and speak slowly. To prevent interviewer bias from affecting the validity of the responses, they need to present the surveys in a consistent way. As a result, interviewer directions must be clear. Chapter 12 discusses interviewer training in some detail.

CHILD PROFILE

English

IF YOU HAVE TROUBLE UNDERSTANDING THE MATERIALS, OR IF YOU NEED HELP FINDING HEALTH CARE, YOU CAN CALL HEALTHY MOTHERS, HEALTHY BABIES AT 1-800-322-2588. WHEN YOU CALL, TELL THE PERSON THE NAME OF YOUR LANGUAGE AND STAY ON THE LINE UNTIL YOU ARE CONNECTED WITH AN INTERPRETER.

Amharic

ከዚህ ጽሑፍ ውስጥ ያልገባዎ ነገብ ካለ ወይም የህክምና አገልግሎት ካስፈልጎዎ፣ ጤናማ እናቶች፣ ጤናማ ህፃናት (ሄልዚ ማዘርስ፣ ሄልዚ ቤቢስ) በስልክ የጣታውወን የህክምና ከፍላ በ 1-800-322-2588 በመደወል መጥየቅ ትችላላችሁ። በምትደውሉበት ጊዜ ለምታነጋግሩት ሰው የቋንቋችሁን ስም ብትነግሉ፣ አስተርጓሚ ሊያነጋግላችሁ ስለሚችል የስልክን መስመር በመያዝ ጠብቁ።

Cambodian

បើលោកអ្នកមានការពិបាកយល់ទូរសារអ្នាះទាំងឡាយ ឬក៏ត្រូវការជំនួយក្នុងការស្វែងរកមន្ទីរព្យាបាលជំងឺ លោកអ្នកអាចទូរស័ព្ទទៅកន្លែ Healthy Mothers និង Healthy Babies តាមយោះលេខ 1-800-322-2588 ។ នៅពេលលោកអ្នកទូរស័ព្ទទៅ សូមប្រាប់អ្នក លើកទូរស័ព្ទនូវភាសាដែលលោកអ្នកនិ យាយហើយរង់ចាំទាល់ដែលគេបន្ថូរទូរស័ព្ទទៅអ្នកបកប្រែ។

Chinese

如果您無法完全理解這些資料，或者您需要幫助尋找醫療保健服務，請打電話至 "健康母親、健康寶寶" 熱線 (Healthy Mothers, Healthy Babies)，號碼是 1-800-322-2588。當您打電話時，請告訴接線員您說什麼語言，並請耐心等候直到接線員將您與一位翻譯接通。

Korean

만약 귀하에서 자료를 이해하시는데 문제가 있거나 건강관리자를 찾는데 도움이 필요하시면 무료 전화인 1-800-322-2588 건강한 엄아, 건강한 아기(Healthy Mothers, Healthy Babies)로 전화를 하십시오. 전화를 하실 때 전화 받는 분에게 귀하가 사용하시는 언어를 말씀하시고 통역원과 연결이 될 때 까지 전화를 들고 계십시요.

Laotian

ຖ້າທ່ານວ່າທ່ານມີບັນຫາບໍ່ເຂົ້າໃຈຄຳອະທິບາຍ ຫຼືທ່ານຕ້ອງການຄວາມຊ່ວຍເຫລືອທາງບໍ່ບຮັບການປິ່ນປົວ ທ່ານອາດຈະໂທໄປ ຫາ ໂຮງການແມ່ມີບຸບແດ່ກາໄມ່ບຸບ (Healthy Mothers, Healthy Babies) ເບີໂທລະໄທ 1-800-322-2588 ເມື່ອ ທ່ານໄດ້ໂທໄປທາງແລ້ວ ຈົ່ງບອກຜູ້ຮັບສາຍວ່າທ່ານປາກາພາສາໝັງຮຮໍ້ຕ້ຖ້າຈົນຖ້ຽງມີນາຍພາສາມາຮັບສາຍ.

Russian

ЕСЛИ У ВАС ТРУДНОСТИ В ПОНИМАНИИ МАТЕРИАЛА, ИЛИ ВАМ ТРЕБУЕТСЯ ПОМОЩЬ В ПОИСКЕ МЕДИЦИНСКОГО ОБЕСПЕЧЕНИЯ, ТО ВЫ МОЖЕТЕ ПОЗВОНИТЬ «ХЕЛСИ МАЗЕРС, ХЕЛСИ БЭЙБИС» ПО ТЕЛ. 1-800-322-2588. КОГДА ВЫ ЗВОНИТЕ, НАЗОВИТЕ ГОВОРЯЩЕМУ С ВАМИ ЯЗЫК, НА КОТОРОМ ВЫ ГОВОРИТЕ, И ЖДИТЕ ПОКА ВАС СОЕДИНЯТ С ПЕРЕВОДЧИКОМ.

FIG. 11.2. Healthy Mothers introduction translation.

248

Somali

HADII AY DHIBATO KA'HAYSATO DAAWOOYINKA IYO QALABKA DARYEELKA
CAAFIMAADKA AMA AAD U BAAHANTAHAY CAAFIMAADKA JOOGTADA ALI IYO
WIXII LA MID AH WAXAAD U YEERTAA TELEFOONKA 1-800-322-2588 MARKA
CIDDI KULA HADASHO U SHEEG LUUQADDADA KUNA SII JIR LAYNKA
TELEFONKA SI LAGUUGU XIRO TURJUBAAN.

Spanish

SI TIENE USTED CUALQUIER PROBLEMA PARA ENTENDER ESTOS MATERIALES, O
SI NECESITA AYUDA PARA ENCONTRAR CUIDADO PARA LA SALUD, PUEDE
LLAMAR A HEALTHY MOTHERS, HEALTHY BABIES AL 1-800-322-2588. CUANDO
LLAME, DIGALE A LA PERSONA EL NOMBRE DE SU IDIOMA Y QUEDESE EN LA
LINEA HASTA QUE SEA CONECTADO CON UN INTERPRETE.

Tigrigna

ብዚኦም ዝስዕቡ ዘለዉ ጽሕፋት ዘይተደላእም ነገር እንተሎ ወይ ደማ ናይ ሕክምና እገዛገሎት
ተደልዮ እንድጨር ዮይንኽም እብ ናይ ጦዖኑት እደታጎን ቆልዑን (ሄል ጦዝቢ ሄል
ቢቢስ) ዝዓየል ሕክምና ግልጋል ብ 1-800-322-2588 ብጽደዋ እትሓተ ትኽእለ እኹም። እብ
ትድመሉ ጊዜ ነቲ ተዛርቦም ሰብ እየናዩ ቋንቋ ትዛረቡ ክተዛኽም ዖለ ነገርኽም ጸንጽኹም
ዝዛረብ ሰብ (ትርጉም እገዛገሎት) ክትረልኩም ስለዝኽእል እብ መስመር ዮይንኽም ሓለዉ።

Ukrainian

ЯКЩО ВИ ІСПИТУЄТЕ ТРУДНОЩІ В РОЗУМІННІ МАТЕРІАЛУ, АБО ВАМ
НЕОБХІДНА ДОПОМОГА В ОТРИМАННІ МЕДИЧНОГО ЗАБЕЗПЕЧЕННЯ, ТО ВИ
МОЖЕТЕ ПІДЗВОНИТИ „ХЕЛСІ МАЗЕРС, ХЕЛСІ БЭЙБИС" (HEALTHY MOTHERS,
HEALTHY BABIES) ПО ТЕЛЕФОНУ 1-800-322-2588. КОЛИ ВИ ДЗОНИТЕ, НАЗОВІТЬ
РОЗМОВЛЯЮЩЕМУ З ВАМИ МОВУ, НА ЯКІЙ ВИ РОЗМОВЛЯЄТЕ, ТА ЧЕКАЙТЕ
ПОКИ ВАС СОЄДІНЯТЬ З ПЕРЕКЛАДЧИКОМ.

Vietnamese

NẾU BẠN GẶP KHÓ KHĂN TRONG VẤN ĐỀ THÔNG HIỂU CÁC TÀI LIỆU, HOẶC
NẾU BẠN CẦN GIÚP ĐỠ TRONG VIỆC TÌM DỊCH VỤ Y TẾ, BẠN CÓ THỂ GỌI CHO
HỘI HEALTHY MOTHERS, HEALTHY BABIES TẠI SỐ ĐIỆN THOẠI 1-800-322-2588.
KHI BẠN GỌI, BẠN HÃY NÓI CHO NGƯỜI NGHE ĐIỆN THOẠI BIẾT BẠN NÓI
NGÔN NGỮ GÌ, RỒI CẦM ĐIỆN THOẠI CHỜ ĐỢI, VÀ HỌ SẼ KIẾM NGƯỜI
THÔNG DỊCH ĐỂ NÓI ĐIỆN THOẠI VỚI BẠN.

FIG. 11.2. (Continued)

Generally, instructions to the interviewer appear in all capital letters or
in brackets to set off the information from the parts that are read aloud.
Words that require special emphasis can be italicized or put in boldface.
Information read by an interviewer needs to sound conversational in tone
rather than formal.

Sample interviewer instruction:

As I read the following list of information sources, please tell me
whether the source is **[READ SLOWLY] very important, important,
unimportant, or very unimportant** to you: [REPEAT CATEGORIES AS
NECESSARY.]

Cultural/Language Sensitivity

Knowing the target public well can aid design and secure a better response. For example, question wording may need to change depending on the age of the participants. Translators may be needed if a phone survey aims to interview people who do not speak English. Mail surveys, too, can benefit from making translations available. As shown in Figure 11.2, Washington state made a big effort to reach recent immigrants with its surveys and information intended to help ensure good health care for babies and toddlers.

FINAL THOUGHTS

Clearly, a myriad of design issues contribute to the effectiveness of a questionnaire. Research managers need to write questions in a way that will provide the most meaningful and unbiased information. The order of questions must seem logical and interesting to the respondent, and the directions to respondents and interviewers must prevent confusion. Pretesting needs to check for the existence of subtle bias in word choice.

No perfect survey exists, but various strategies can boost response rates and increase the validity of responses. The details can make the simple task of asking questions seem overwhelmingly complicated, but the use of pre-existing instruments or batteries of questions can provide useful models to follow. For the manager working within a tight budget, the principles laid out make it possible to run a reliable, valid, and useful survey. In addition, various books and even online software such as Survey Monkey (*www.surveymonkey.com*) and Zoomerang (*info.zoomerang.com*) can guide managers through the process. Meanwhile, for managers able to hire experts to do the job, the principles presented here can make it easier to monitor a survey research firm to ensure top-quality results.