

Appendix A: Commonly Used Measurement Dimensions

In this section, we are going to critically evaluate some common approaches to asking questions about commonly measured subjective and objective phenomena.

1. Measuring frequency
2. Measuring quantity
3. Measuring feelings
4. Answering evaluative questions
5. Measuring satisfaction
6. Rating agreement
7. Measuring priorities
8. Dealing with people who “don’t know”

Obviously, the very best question will depend on the research context and what the researcher is trying to do. One useful step is to look at how other researchers have measured these dimensions or addressed these issues. That is always good practice. However, the goal of Appendix A and Appendix B is to help readers think about, and choose among, the various alternative questions that could be used.

Measuring Frequency

Certainly asking people questions about the frequency with which they do various things is one of the most common tasks in survey research. Respondents are asked how frequently they consume alcohol, use marijuana, go to the bank, eat meat, go to the doctor, and have sexual relations.

Consider the following set of response categories:

- More than once a day
- Almost every day
- A few times a week
- About once a week
- Two or three times a month

About once a month
Less than once a month
A few times a year
Once a year or less

The response categories in this scale cover a wide continuum. Which of the categories in the scale are appropriate depends, of course, on what is being asked. For things that a person might do frequently, such as eat or drink common foods, categories at the frequent end of the scale might be used. A different part of the scale would likely be appropriate for recreational activities such as going to the opera or going to a movie.

A fundamental problem with a question oriented toward this response task is the assumption of some kind of regularity. Although some sorts of behaviors lend themselves to regularity, irregularity is probably more common. Moreover, even for activities that are done regularly by some people, such as going to church or drinking alcohol, there are other people whose patterns are more erratic.

Generally speaking, measures of frequency are done better by asking people to estimate the number of times they have actually done something in an appropriate period of time. The period of time will vary depending on what behavior is being measured.

Example A.1: In the last 14 days, on about how many different days did you have any wine, beer, or liquor to drink at all?

Example A.2: In the last 30 days, on about how many different days would you say that you did physical exercise for a period of at least 20 minutes?

Example A.3: In the past 12 months, how many different times were you admitted to a hospital as a patient overnight or longer?

The most important criticism of questions that cover a short time span, such as a week or two, is that the time period may not be representative. Hence, if the past two weeks have been atypical, the person may not be characterized well as, for example, a heavy or a light drinker. On the other hand, asking people to report frequency of behaviors, even if they have to be estimates, solves the problem of trying to impose some kind of regularity or pattern on people's behaviors.

Another approach to frequency might be something like the following:

Very often	Often
Fairly often	Sometimes
Occasionally	Rarely
Rarely	Never
Never	

Response scales such as the above are somewhat more difficult to interpret than responses that use numbers. A great advantage of asking people to report frequencies over some period of time is that it avoids the problem of different definitions of what constitutes "often." Respondents who differ in their conception of what constitutes "often" may give different answers for that reason rather than because of their differing behaviors. However, if a question is being asked that seems to defy quantification, and the general purpose is to sort people into very broad categories, a question like the above might work fine.

Frequency is particularly difficult to ask about when the question includes two things that could vary. For example, if one wanted to know about popcorn eating at the movies, the absolute frequency would be the product of the number of times a person went to the movies multiplied by the percentage of times the popcorn was purchased. For such questions, a proportional sort of question might be best such as:

Always	Almost always	Always
Usually	More than half the time	Usually
A good bit of the time	About half the time	Sometimes
Some of the time	Less than half the time	Rarely
Rarely	Rarely	Never
Never	Never	

The above three scales are all approaches to measuring something as a proportion of time or events. They also might be the right way to quantify feelings such as pain or fatigue. Talking in terms of the proportion of time that a person was in pain or felt fatigue might be a better way to ask the question than asking about the number of days on which fatigue was felt. In the same way, if a student was asked how frequently homework papers were produced on time, proportion might be the right way to ask the question.

Several differences among the response tasks above are worth noting.

1. People taking scholastic aptitude tests are taught seldom to answer "always" or "never." For some questions, "always" or "never" are in fact reasonable answers that researchers want to identify. "Always" using birth control is importantly different from "almost always." "Never" fainting is importantly different from "rarely."
2. The scale that uses proportion of time, such as more than half the time or less than half the time, may be somewhat less subject to individual variability in interpretation of adjectives than the other scales. Scales that are quantifiable, as compared with those that rely purely on adjectives, also have some advantages in translations across languages.
3. Scales of frequency pose some difficulty in defining the middle or intermediate categories. "Usually" or "always" on the positive end and "rarely" or "never" on the negative end are fairly well defined and commonly

understood. Terms like “some of the time,” “a good bit of the time,” “sometimes,” and “fairly often” are harder terms to work with in the middle of a continuum of frequency. The scales above probably do a fairly good job of dealing with that problem, but it is a continuing challenge for survey researchers to do a good job of labeling spots on a frequency continuum in and around the middle of the scale.

Measuring Quantity

Many survey questions also try to measure quantity or size. Typical examples are “How much of a problem do you have?” “To what extent are you affected by something?” or “How much do you care?” Two common approaches to labeling a continuum of quantity include:

A lot/a great deal	Big
Some	Medium
Only a little	Small
Not at all	None at all

It turns out that it is difficult to break the continuum of quantity into many more than four categories using adjectives. One could add “very small problem” between “small” and “none.” One could add “almost not at all.” However, four or five adjectives is about all that can be sustained.

Subjective sense of quantity is a measure that lends itself to a numerical scale.

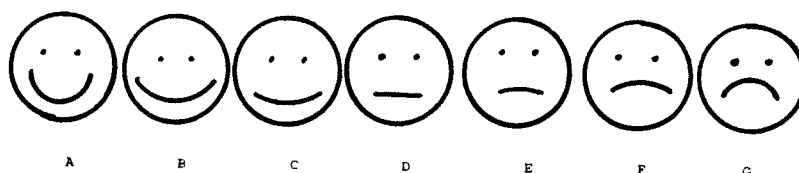
Example A.4: Think of the scale from 0 to 10, where numbers 8, 9 and 10 stand for a very great amount; numbers 0, 1 and 2 stand for none or a little; and the numbers in the middle, 3 through 7, are somewhere in the middle. What number would you assign to the amount of pain you have been experiencing over the last month?

Measuring Feelings

Rating how respondents feel about things, from positive to negative, is another of the most common survey tasks that people are given. A common scale for measuring feeling is:

Very positive
Generally positive
Mixed: About equally positive and negative
Generally negative
Very negative

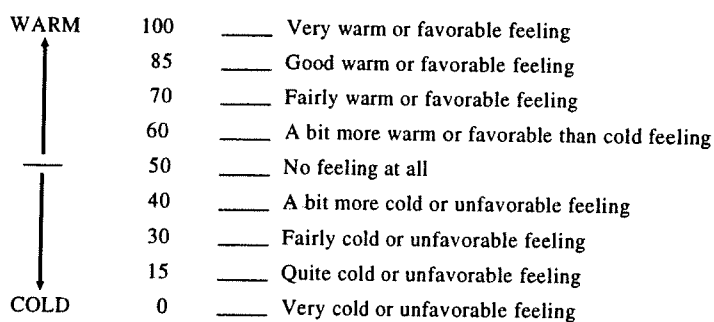
Here are some faces expressing various feelings. Below each is a letter.



Which face comes closest to expressing how you feel about _____?

Figure A.1. Using Faces to Measure Feelings

SOURCE: Andrews and Withey (1976). Reprinted with permission.



Where would you put _____ on the feeling thermometer?

Figure A.2. Using the Feeling Thermometer to Measure Feelings

SOURCE: Andrews and Withey (1976). Reprinted with permission.

Another approach is the "Delighted to Terrible" scale:

Delighted
Pleased
Mostly satisfied
Mixed
Mostly dissatisfied
Unhappy
Terrible

Two other approaches that use visual aids are presented in Figures A.1 and A.2.

All of the above have been used in numerous surveys. Andrews and Withey (1976) did a systematic evaluation of different ways of measuring how people felt. Their conclusion was that the Delighted to Terrible scale was the best in terms of its general utility and other psychometric properties, such as how well it spread people and the results of construct validity assessments. However, the faces and the 10-point thermometer proved to work almost as well. One great advantage of both of the latter scales is that they do not depend on adjectives. Hence, although the strength of the Delighted to Terrible scale may be the effective way in which the scale points are communicated consistently to people by way of language, a weakness is the difficulty in translating that across languages. The Delighted to Terrible scale is very long for use on the telephone, whereas the faces and thermometer require respondents to be exposed to visual stimuli. Of course, a numerical scale, such as 0 to 10, without visual aids, can be used to measure feelings just as it can be used to measure quantity.

One final note is in order about using numerical scales for measuring feelings. Although every point on a continuum does not need to be defined, it probably is important to define a midpoint, the place where positive feelings turn into negative feelings. That can be done adjectivally, as it is done on the Feeling Thermometer above. Another way of doing it is to give people a numerical task of rating their feelings from -5 to $+5$, with 0, obviously, being the midpoint, when positive feelings turn to negative.

Asking Evaluative Questions

The feeling questions outlined above are for use when people are asked to rate how they feel about things. The questions in this section are for use when people are asked to say how they evaluate something. The two may be correlated, but they are not identical.

The most commonly used evaluative scale may be:

Excellent
Very good
Good
Fair
Poor

An almost equally common rating scale is 0-10, where 10 is as good as it can be and 0 is as bad as it can be. The excellent-to-poor scale is a very easy one to use. Because people are familiar with the words, it is not difficult to remember. The disadvantage is that it has only five categories, and for many ratings it turns out that people tend toward the positive end of the scale. Therefore, it may be valuable to use rating scales that have more points. It is hard to think of adjectives to add to the continuum between good and excellent that have

meaning and do not sound redundant. For that reason, the numerical scale approach from 0 to 10 may in fact be a preferred solution for many rating tasks.

Measuring Satisfaction

Measuring satisfaction is conceptually different from measuring evaluation, though they often would be judged to be highly related. Theoretically, satisfaction is the relationship between what people want and what they get.

Very satisfied	Completely satisfied	Perfectly satisfied
Generally satisfied	Mostly satisfied	Slightly dissatisfied
Somewhat dissatisfied	Mixed—equally satisfied and dissatisfied	Somewhat dissatisfied
Very dissatisfied	Mostly dissatisfied	Very dissatisfied
	Completely dissatisfied	

One question people would raise about the first scale is the fact that the words are not symmetrical. However, one problem with satisfaction as a label is that "somewhat satisfied" may be negative. "Generally satisfied" is almost certainly a more positive state than "somewhat satisfied," and the first scale is probably more symmetrical (cognitively, if not verbally) than if "somewhat satisfied" was the second category.

It also should be noted that there is no reason why the adjectives along a continuum that go through zero need to be exactly the same on each side of a midpoint. The goal is to find adjectives that consistently communicate different spots on a continuum. The characterization of satisfaction may be one for which this is best done by using somewhat different words on the opposite sides of the midpoint.

Whether or not a midpoint is necessary is another issue. Conceptually, the notion of balancing satisfaction and dissatisfaction may be wrong for some tasks. Satisfaction is really about what people get. Being less than completely satisfied and not getting exactly what a person wants are deviations from perfection. There is a good case to be made that we are simply scaling degrees of dissatisfaction as we move further away from "completely satisfied." That orientation leads to the righthand column of responses.

Finally, researchers should be careful to consider the possibility that respondents do not care about some aspect of their life. A satisfaction question should be asked only about things with which respondents come in contact and for which they have some kind of expectations. For example, if a person walks to work, a rating of satisfaction with available parking may be meaningless. If it is not reasonable to think that all respondents have expectations about something, it might be better to have them do a rating task than to have them report on their satisfaction.

Rating Agreement

Completely agree	Strongly agree	Completely true
Generally agree	Agree	Somewhat true
Generally disagree	Disagree	Somewhat untrue
Completely disagree	Strongly disagree	Completely untrue

There is a discussion in Chapter 3 about the conceptual complexity of combining “strength” of feelings and agreement in a single scale. Although the middle form of the scale above may be the most commonly used, there is good reason to avoid that particular approach. Either of the other two scales, which focus on the cognitive aspects of the task, are more likely to prove satisfactory measures.

Whether one prefers the “true” formulation of the response task or the “agreement” formulation of the response task depends on whether it is an evaluative task or a cognitive task. If the question is how close preferences come to matching a statement, then asking about the agreement makes more sense. If the question is how close perceptions come to a particular statement, then the correct dimension is perception of truthfulness.

We should reiterate two cautions expressed in the parallel section in Chapter 3. In theory, there is a middle point, where a person is perfectly ambivalent about whether to agree or disagree. Many researchers offer a middle category (undecided, neither agree nor disagree). A category like that will appeal to some respondents, but it probably does not hurt to make respondents commit themselves. Moreover, that middle category becomes a haven for those people who lack the information needed to have an opinion about a question. Those people are different from ambivalent informed people. They should be identified separately, with a screening question, and they should not be placed in a middle category between those who agree and those who disagree.

Measuring Priorities: Rank Order Versus Ratings

Researchers often want to measure priorities, and one obvious approach is to ask people to rank order some alternatives:

Example A.5: Here is a list of problems some people are concerned about. Which of these do you consider most important?

- a. Reducing crime in the streets.
- b. Reducing air pollution.
- c. Providing jobs for those who need them.
- d. Providing health care to those who need it.
- e. Lowering federal income taxes.

Respondents could be asked sequentially for the most important, next most important, etc. If the survey is done in person or with a self-administered form, they could rank order the items from 1 to 5.

Example A.6: Consider a scale from 0 to 10, where 8, 9, and 10 are the highest priority; 0, 1, and 2 are low priority; and 3 through 7 are in between. What number would you give to the priority of (EACH).

The second kind of task is easier for respondents, particularly on the telephone. By definition, the rank order task does not give us information about how much people care about any item. In theory, a respondent could consider even the top-rated item as unimportant or the lowest rated item as very important. The data from the rating task can produce ties; all the items could be rated 10. However, the average across groups of respondents usually will produce an ordering, if that is important, and in general, the multiple-rating task provides more information with less respondent effort. For most surveys, it probably represents a better approach.

Dealing With People Who “Don’t Know”

There are many questions asked in surveys for which a “don’t know” answer is not a plausible answer. Often, when respondents say “don’t know,” it is a response style that means that they have not thought very much about a topic, or they are not certain, among several response possibilities, which answer fits best. In such situations, researchers would probably prefer that the respondent work a little harder and come up with an answer.

When respondents are asked for their feelings, opinions, or evaluations of things outside themselves—public policies, political figures, institutions—it becomes quite plausible to think that they lack information in sufficient detail to provide a rating or have an opinion. When it is reasonable that a person might not actually know enough to answer a question, researchers would be best served if they developed an explicit strategy for screening out such people.

There are two ways to do this. First, there can be a screening question asked in advance of the opinion question.

Example A.7: The next question is about the Agricultural Trade Act of 1978. Do you feel familiar enough with that act to have an opinion about it or not?

Alternative Example A.7a: Thinking back to the terms of the Agricultural Trade Act of 1978, would you say that in general you were in favor of that act, opposed to that act, or are you not familiar enough with the act to have an opinion on that?

Either one of those approaches to screening respondents has two real advantages over ignoring the “don’t know” possibility. First, it explicitly condones the notion that some people may not have enough information to answer a

question, making it easier for respondents to admit to ignorance. More important, it defines a specific process for having respondents themselves decide whether or not they know enough to answer the question. In the absence of an explicit effort to identify people who are not familiar enough with a topic to answer a question, it turns out that interviewers and respondents are inconsistent in the way that they handle the uncertainty about whether a respondent should try to answer a question (e.g., Groves, 1989). Providing an explicit process improves the standardization of the measurement process.

Example A.8: The next question is going to ask you about your opinions about the schools here in your town. Do you feel you have enough information about the schools to have an opinion about them?

A disadvantage of this explicit kind of screening question, as compared with building the "no opinion" alternative into the question, is that the respondent does not know what sorts of questions are going to be asked. It is easy to think of a respondent who would have opinions about some aspects of schools, but who would answer "no" to the above screening question. (Schwartz & Hippler, 1991). Although the general screening question about the Agricultural Trade Act may in fact have been a reasonable way to screen respondents, for many purposes the inclusion of a "no opinion" alternative among a set of response options may permit respondents to make a more informed judgment about whether or not they are able to answer questions. The downside of explicitly giving a "don't know" option is that the number of such answers will be increased. That is the reason why prior analysis of the nature of the response task is so important. If respondents are being asked about their firsthand experiences or feelings, minimizing unanswered questions, "don't knows" and others, is reasonable. On the other hand, when respondents are asked for the opinions or perceptions of things beyond their direct experience, a "don't know" response is a potentially meaningful answer, not missing data, and it is best obtained in an explicit, standardized way.

CONCLUSION

It always makes sense for people designing survey instruments to look at the way that other experienced researchers have designed survey questions. Two of the major resources in the United States are the General Social Survey, conducted by the National Opinion Research Center, and the surveys that are maintained by the Political Consortium at the University of Michigan.

In addition, regardless of how many times a question has been used, or by whom, survey questions need to be evaluated. They need to be evaluated carefully before a survey is launched, and the results also need to be evaluated from the perspective of the validity of the measurement that resulted. (See Chapters 5 and 6.)

Appendix B: Measures of Common Covariates

In most surveys, researchers collect descriptive information about respondents against which to tabulate answers. Some of the most common such variables are:

1. age
2. gender
3. marital status
4. employment status
5. socioeconomic status (education, income, occupation)
6. religious preference or affiliation
7. race and ethnic background

In this appendix, we discuss alternative ways of asking questions about these respondent characteristics. Once again, the goal is not to prescribe the right questions for all surveys, but rather to raise some issues about how best to measure these that researchers may want to consider.

Age

The most obvious way to ask this question, probably, is:

Example B.1: How old are you?

Example B.1a: What is your age?

Comment: One possible ambiguity about these questions has to do with rounding. Some people may report how old they will be on the nearest birthday. An alternative wording that is common is:

Example B.1b: How old were you on your last birthday?

People do a fair amount of rounding when asked for their age. There is heaping at round numbers, such as numbers ending in 5 and 10. This suggests that some people prefer to give an approximate figure for their age, rather than

giving an exact number. This may lead researchers to prefer asking a different question:

Example B.1c: In what year were you born?

When that question is asked, obviously the researcher has to do the calculation of age. "What was your date of birth?" is the question needed to calculate an exact age.

Finally, there clearly are respondents who are sensitive about their age and who would prefer to answer a question in general categories. If a researcher is only interested in analyzing respondents by large categories, they should be asked the question in that form.

Example B.1d: In which of these groups is your age: under 30, 30 to 44, 45 to 64, 65 or older?

Gender

This might be one of the most straightforward questions in survey research. The categories are clear and mutually exclusive. Almost all respondents have a clear notion of the answer.

Example B.2: Are you male or female?

This particular form of the question avoids interactions with the age of the respondent that man, woman, boy, and girl entail. Just about the only problem with this question is that interviewers are reluctant to ask it when they think it should be obvious, but it is not. In face-to-face interviews, this information usually is collected by interviewer observation, rather than by asking the question. On the telephone, however, although most often interviewers think they know the gender of their respondent, occasionally they are not certain. The wording above is as good as any.

Marital Status

Example B.3: Are you married, divorced, widowed, separated, or have you never been married?

This is one of several questions we will discuss that seems quite straightforward but in fact poses a number of potentially important problems.

One problem is the variation in the way that people go about dissolving or ending their marriages. Some people get a divorce, some people get a legal separation, and some people simply split up for a while. Alternatively, some

married couples, people who consider themselves successfully married, live apart temporarily but would not consider themselves to be "separated."

The term "never married" is a better one than "single," which occasionally is seen in these questions. To many respondents, "single" simply means unmarried, which obviously can apply to several of the above categories.

Certainly the most troublesome status for some researchers is designated as "living as married." Some survey researchers have responded to the fact that some people live together, behaving in various ways similar to those who are married, but yet are not formally married, by adding this response alternative. Adding such a category to the list of acceptable "marital states" solves a problem about how to classify such people, and some respondents like having a category that seems to capture their situation so well. On the other hand, that category raises problems of definition. Whether researchers have actually gained useful information about respondent status by having that category is debatable.

The complexity comes from the fact that being married has various implications, depending on the analysis. For studies related to economic well-being, being married means that incomes and expenses are likely to be shared. For studies related to children, married parents mean potentially sharing child care and also having long-term economic responsibilities for support of children. For studies of social well-being, being married implies having a friend. For studies of risk of AIDS, being married suggests monogamy. It is not at all clear which of these, if any, is implied by choosing the category, "living as married."

The simple question with which this section started is carrying a great deal of weight, because it has embedded in it at least half a dozen questions. For many research purposes, four questions supply all the information needed, and in a better and more reliable form than the single summary question.

Example B.4: Are you married or unmarried now?

Example B.4a: (IF MARRIED) Are you and your (SPOUSE) living together now or living separately?

Comment: For many purposes, that is all the information that is needed. A bit of refinement might be gathered by adding a third question:

Example B.4b: (IF SEPARATELY) Do you consider the arrangement of living separately to be a temporary or a permanent arrangement?

In some cases, it also is useful to find out more about those who are not currently married. There are two main questions that one would like to ask in order to replicate the information that presumably we were getting in the initial question.

Example B.5: (IF NOT NOW MARRIED) Have you ever been married?

Example B.5a: (IF EVER MARRIED) Are you widowed, legally divorced, legally separated, or just living apart from your (SPOUSE)?

Comment: Notice that the people who said they were married and living separately would be classified with those who said they were not married, but had not legally separated.

Depending on the goals of the research, researchers might want to ask additional questions about the relationship between the respondent and previous spouses. In addition, if there are concerns such as sources of financial support or the availability of companionship for the respondent, those questions should be asked more directly, not inferred from the answer to a summary question about marriage.

Employment Status

The situation with employment status is very much the same as marital status.

Example B.6: Are you employed, on layoff from a job, unemployed, keeping house, a student, or retired?

In the 1950s, this may have been a reasonable question. Today, it is a very poor question. Among its problems, it violates two of the most important standards for a good survey question: the categories are neither well defined nor mutually exclusive.

Let us begin with the problems of overlapping categories. It is very common for people enrolled in educational programs also to be employed; the majority of students work to some extent, and a good number of people with full-time jobs are taking courses and seeking degrees.

The status of retired also is not clear. For many jobs—jobs in public safety and the armed forces stand out—it is possible to retire at a comparatively early age, far before the traditional 65. Such people often go on and take additional jobs. Are they retired, because they are on a retirement income from a long-time career, even though they are currently employed?

It also is becoming increasingly ambiguous as to what constitutes employment. A large number of people now work for pay but are not considered real employees; they do not receive benefits and the work could stop with a minimum of notice. If the goal is to find out if a person has any income as a result of work, such jobs certainly constitute employment. However, if the goal is to identify people who have a stable relationship with some kind of organization, the coding is more complex.

Once again, the basic problem is that about five or six questions are being asked at once. The question is asking respondents to do a complex coding task, rather than providing respondents with a set of questions that can be directly answered that will provide information the researcher needs to do the coding. The researchers' task, in this case, is to break down and identify exactly what

kind of information is needed from an analytic perspective, then ask respondents to provide it. One possible series might look like the following.

The usual first goal is to find out whether or not a person is working for pay. The question might be as simple as this:

Example B.7: Last week, did you receive any pay from a job or for work that you did?

Comment: Even this question is probably imperfect. For example, there are people who perform services who may not exactly consider it to be work. However, those issues are problems for a small segment of the population. People who have trouble with that distinction are certainly going to have more trouble with our initial summary question.

There undoubtedly are many other things researchers might want to know about the working situation. Does the person have one job or two jobs? Does the person work full-time or part-time? Is the person paid a salary or by the hour? Does the job have benefits, such as paid vacations and health insurance? How long has the person worked at the job? How secure does the person feel that the job can continue indefinitely?

These are all reasonable questions. Which of these are important depends on the research objectives. The answers to none of these questions can be directly inferred from the fact that a person chose "employed" as compared with some alternatives. If the researcher wants to know about these aspects of employment, questions must be asked.

If the person did not receive money in the past week, we might want to find out why. Several of the categories in the initial question have to do with classifying nonworking status. One critical concept is whether or not the status of "not working" is voluntary or involuntary. The question that the Bureau of Labor Statistics has relied on for many years to define involuntary unemployment is, "Have you been looking for work in the last four weeks?" Others have pointed out that this question is fraught with ambiguity, the key ambiguity being what constitutes "looking for work." It has been noted that people who have given up looking for work, and give a "no" answer to that question, become reclassified as voluntarily unemployed people. As usual, one bit of advice is to ask a more direct question. If the goal is to find out whether or not unemployment is voluntary, why not ask something like:

Example B.8: Would you say you are not working now mainly because you do not want to work right now, or mainly because you cannot find a job?

Asking questions about whether people are students should be a discrete line of inquiry.

Example B.9: During the past six months, have you taken any courses for credit toward a degree or diploma?

The advantage of this question is that it does not make assumptions about the pattern by which people acquire their educations. This particular question rules out adult education courses and technical training that is not part of a degree program. Obviously, one could ask other questions aimed at such activities. In addition, one could ask follow-up questions to find out what sort of degree the person was pursuing, whether enrollment was full-time or part-time, and so forth.

Finally, with respect to retirement, the main significance of that classification may be that it identifies a person who has decided not to do any more work for pay. For many purposes, simply the fact that the person is voluntarily out of the labor force is all that is needed. If a researcher is interested in future work plans, then a specific question is needed to identify those people who, at the moment, have no future plans to re-enter the labor force. The same applies to people who consider themselves temporarily or permanently out of the labor force because of disability.

Socioeconomic Status

Measures that researchers use as indicators of socioeconomic status include:

1. income
2. educational attainment
3. occupation

There generally are three different kinds of variables that people are actually trying to measure:

1. information or knowledge, along with some possibly related attitudes, that some people think go along with higher socioeconomic attainment
2. resources
3. social status or prestige

Occupation (or more generally, occupational status), income, and education all tend to be somewhat intercorrelated in a positive way. However, each has limitations, often very severe limitations, for capturing what researchers are truly after.

Educational attainment is probably the most generally useful and interpretable measure of the three. Simply knowing how many years of schooling a person has is only a rough indicator of knowledge, values, and training. Moreover, there have been trends over the years toward higher average levels of formal education, creating an interaction between age and educational attainment. Nonetheless, there is a reasonably straightforward basis for postulating a

direct relationship between education and information. Moreover, educational attainment is not dependent on the circumstances of the respondent in a way that distorts the meaning of education, as is the case with income.

Income would seem likely to be the best, most direct measure of resources or financial well-being. However, that statement must be tempered by three complexities:

1. A person's resources may include not only current personal income but also the combined incomes of other family members.
2. Well-being is also affected by the availability of assets. That complexity particularly confuses comparing the well-being of young people, who typically have few assets other than income, with retirees, who are likely to have more assets.
3. Well-being is dependent not only on income but also on financial obligations. Given the same income, a single, unmarried person or retired person will be better off than someone responsible for providing for a family of four.

Obviously, if one is studying financial well-being in detail, one would want to collect information about all of these topics: income of other family members, assets, and financial responsibilities. However, the point is that knowing an individual's current income alone, without factoring in some of these other considerations, probably is quite an imperfect measure of financial well-being or resources.

Occupation, that is, the kind of job that a person occupies, is an even more complicated and probably less useful way of learning something about socioeconomic status. The notion of using occupation, usually of the "head of the family," as a measure of the social standing of that family has its roots in sociological theory, probably fairly traced to Karl Marx, who thought that the position of a person with respect to work defined his or her position in society. In survey research, a common strategy for making sense of the hundreds of different occupational classifications currently in use is to collapse them into eight or ten categories: professional, managerial, sales, clerical, skilled worker, semiskilled worker, service worker, and unskilled labor (plus farm owners and farm workers). These groupings were thought to roughly define an ordered set of occupational categories. The imperfection of the classification, however, is reflected in the fact that the "professional" category includes grade-school teachers, putting them in the top category, whereas "service workers" includes domestic helpers, people who work at McDonald's, and police officers and firefighters. The last two are often paid more than grade school teachers.

The meaning of occupation as a measure of social standing is also confused by the fact that in most two-adult households, there are two workers with two occupations.

A final barrier to routine use of occupation as a generally useful characteristic of respondents is the difficulty of the classification process itself. In order to code occupation, at least three questions are needed:

Example B.10: What kind of work do you do on your job?

Example B.10a: In what kind of business or organization do you work?

Example B.10b: Do you work for yourself or someone else?

Given the answers to these questions, quite a complex coding process must be applied, one that requires good training and supervision in order to carry it out in a reliable way.

When one is doing studies focused on workforce participation, having good information about occupations and the kind of work people do is essential. However, as a general descriptive characteristic of people, to be used as a covariate in analyzing their other answers, measuring current or past occupations is probably a poor choice for most surveys.

For most surveys, the most generally useful question to ask is:

Example B.11: What is the highest grade or year of school you have completed?

In all probability, asking income is done much more frequently than it needs to be. Moreover, for most of the analytic purposes for which it will be used, getting people's incomes in very broad categories is probably all that is required. The most usual, legitimate objective in measuring income is assessing something about the resources available to people. For that purpose, probably total family income for the preceding year is the best single measure.

Example B.12: We would like a general estimate of the total family income for you and for all family members living with you during (LAST YEAR). Considering income from all sources—from jobs, interest, rents, and so forth—for you and all family members living with you, what would you say was your total family income in 19__ before taxes—less than \$20,000, \$20,000 to \$39,000, \$40,000 to \$59,000, or \$60,000 or more?

It is quite obvious that this question, even in this simplified form, is very complicated and does not meet the standards articulated in this book very well. The reporting of income, if a good number is really required, will be improved if incomes from the various possible sources are asked about individually before the respondent is asked to give a total figure. Ask specific questions about pay for the respondent, pay for the respondent's family members, rents, welfare payments (if appropriate), Social Security, pensions, and interest. Then ask for a total. The number and complexity of categories can be elaborated, and for some purposes that is an important and useful thing to do.

Religion

Example B.13: Do you consider yourself to be Protestant, Catholic, Jewish, or some other religion, or do you have no preference?

A person's religion can be measured in one of three main ways:

- a. the religious culture in which the person was raised;
- b. the person's current self-perception; or
- c. the person's affiliation with organized religion.

The three are interrelated: Adults are more likely than not to think of themselves in terms of the religions in which they were raised, and they are more likely than not to belong to religious groups that reflect their self-perceptions. However, there are some differences that result from the choice of one of these approaches.

The most important limitation of asking for membership as a way of classifying people religiously is that many people do not belong to organized religious groups. Membership varies by life cycle; young adults before marriage are less likely to have formal religious ties than people who are married and have children. In addition, formal religious affiliation varies by religious group; it is particularly significant that some groups require members to pay dues or membership fees in order to be formal members, whereas that is not true of others. Most important, because many people who have religious identities and beliefs do not belong to congregations, if affiliation is the measure of religious orientation, many people will not be classified at all.

Whether people should be classified religiously by how they were raised or by their current preference may depend on the purposes for which data are being collected. There are two advantages of asking people the religion in which they were raised. Some people can be put in a religious category on that basis who currently express no religious preference. Also, some people would argue that the important elements of religious orientation are developed during childhood. The disadvantage is that such a classification misses the changes from childhood to adulthood, including conversions, withdrawal from religious convictions that occurs in adulthood, or the development of new religious convictions and commitments that were not apparent during childhood.

Finally, the basic question at the beginning of this section used the term "Protestant," which is not a universally understood term. Some people who are Baptists or Methodists identify themselves that way, but not as Protestants. Moreover, there is not general agreement on exactly who constitutes a Protestant. For example, some people think any Christians who are not Roman Catholics are Protestants, whereas others have more refined definitions. Although the initial, basic question is imperfect, it probably is the most generally useful question to ask as a starter. As a follow-up, because the views and ideologies of Protestant sects vary greatly in the United States, it usually would

be a good idea to ask a follow-up question: Is there any particular branch or denomination that you prefer?

Race and Ethnic Background

Race and ethnic background are, of course, not the same, though they are related at some level. The racial categories commonly used in the United States are:

Black

White

Asian (or Pacific Islander)

Native American Indian (Eskimo or Aleut)

Ethnic background sometimes refers to country of origin, such as Irish or Italian. However, culture and political systems do not always coincide. There are many countries that are considered Hispanic. Armenians and Jews are both ethnic and religious groups, and they "originate" from places other than Armenia (which was part of the U.S.S.R. until its demise) and Israel. Moreover, countries change. The political units in Eastern Europe and Africa were far different in 1993 from what they were in 1970. Thus identifying a person's ethnic or cultural heritage by country of origin can be elusive.

In the United States, the race "Asian" tends to be treated synonymously with being from an Asian country. However, of course, black and African American are not always the same.

Finally, language is another aspect of culture and background on which some researchers base questions.

Example B.14: Are you of Spanish or Hispanic origin?

Example B.14a: (IF YES) What group—for example, Mexican, Mexican American, Puerto Rican, Cuban, etc.

NOTE: Some people prefer the term "Native American" to "American Indian." However, in general surveys, people routinely interpret "Native American" to mean they were born in the United States. The term "Indian" is needed to make the meaning clear.

Example B.15: What is your race—American Indian, Asian, Black, White, or something else?

Example B.15a: (IF ASIAN) What group, for example, Chinese, Asian Indian, Hawaiian, Vietnamese, etc.?

This is a slight variation of the series of questions asked in the United States Census and in many government surveys. Culture, race, and country of origin are thoroughly mixed and confused in this series, reflecting a general confusion about what background characteristics matter.

Identifying Hispanics is routine in government surveys. Because what constitutes a Hispanic origin is not well defined, asking for "group," in a question that really is aimed at country of origin, is added as a check for respondents who are from places where Spanish is spoken who do not identify themselves as Hispanic, and to sort Hispanic people into subgroups. Asking about Hispanic origin before asking race is a good idea, because many Hispanics seem to resist racial classification until they have identified themselves as Hispanic (Martin, DeMaio, & Campanelli, 1990).

For many purposes, the single most useful question to ask adults, in combination with race, is "In what country were you born?" The impact of a foreign culture is greatest for those born outside the United States. Adults born in the United States were raised speaking English. For those who were foreign born, country of origin is an easy question to answer and provides considerable information about the respondent. It then permits the researcher to group countries in any way desired, rather than asking the respondents to classify themselves.

If cultural heritage is of great interest, the birthplace of the mother and father might be the next most useful information.

Example B.16: In what country was your (mother/father) born?

Of course, there are cultural impacts that endure more than two generations. There are millions of third- and fourth-generation Americans who identify themselves as Irish, Italian, Greek, and Polish. To get at that, one might ask:

Example B.17: Most people in this country think of themselves as Americans. However, in addition, is there a particular nationality or ethnic group to which you think of yourself as belonging?

Many people in the United States have lost any sense of pre-American identity, so there will be many "no answers." Moreover, it is obvious that such a question opens the identity gate widely, so a great variety of results can emerge. However, if the goal is to find a cultural heritage of which respondents are aware that is not evident from a person's race, country of birth, and parents' country of birth, that question should find it.

Appendix C:

About Open-Ended Questions

Thirty or forty years ago, as many as half the questions in a standard survey instrument would be open-ended; that is, they would be the type that respondents answered questions in their own words. In the 1990s, however, the open-ended question is becoming increasingly rare. In part, this trend is because many of the measurement goals of survey research can be better achieved when respondents are asked to choose from among a set of answers or to put an answer on a rating scale. In part, as data collection becomes increasingly computerized, interviewers record answers by entering them into a computer, and the survey process goes more smoothly if most answers can be recorded by entering a number rather than a narrative answer.

The open-ended question was discussed to some extent in Chapter 3. As noted there, if the task is to produce an ordered set of answers, or to put answers on a rating scale of some sort, open-ended questions will not do. Moreover, Schuman and Presser (1981) found that even for questions such as, "What is the most important problem?," for which either open or fixed-response strategies will work, providing people with a list of answers to choose from provides more reliable, more interpretable, and possibly more valid responses than the open-ended format.

Nonetheless, there is a role for open-ended questions in survey research. In this appendix, it seemed valuable to briefly go over when such questions might be most appropriate.

The first, and most obvious, situation is when the range of possible answers greatly exceeds what reasonably could be provided. For example, if people are asked to choose their favorite entertainer, their favorite food, or their favorite song, the question must be open-ended; the possible answers clearly would be numerous, diverse, and unanticipatable.

Second, there are some questions that should be answered in a narrative form because the answers are virtually impossible to reduce to a few words. For example, in order to reliably code occupation, respondents need to describe "what kind of work they do on their jobs." Answers consisting of only a word or two often are ambiguous and do not provide enough information to permit coding. Having people answer in their own words, so they can describe what they do, is the way to acquire the information needed. Similarly, to learn about the kind of health problem or condition that led a person to go to a doctor, answers should be in narrative form. Some conditions can be labeled with a

single word or two, but in many cases the best answer will require a more descriptive response.

Third, as discussed in Chapter 3, asking open-ended questions is among the best ways to measure knowledge. When knowledge is measured in a true/false or multiple-choice format, some correct answers can occur by chance and not reflect knowledge of the subject. Open-ended answers to questions usually are a better way to find out what people know.

Fourth, when the reasoning behind a conclusion, a behavior, or a preference is of interest, the best way to learn about it is to hear the respondent's own words. Narrative answers about why people do things or the basis on which they have preferences are subject to some unreliability because of differences in respondents' verbal skills and styles. On the other hand, narrative answers give researchers a much more direct window into what people are thinking. If a researcher wants to understand why voters prefer Candidate A to Candidate B or what they do or do not like about their schools, there are many reasons to want to hear narrative answers in addition to, if not instead of, the responses to standardized, fixed-response questions.

Finally, sometimes asking an open-ended question can be the simplest way to gather systematic information about a potentially complicated situation. For example, suppose the goal was to find out how a person came to be homeless. There are numerous possibilities, such as having a house burn down, losing a job, or being thrown out of the house by parents or other family members. However, the diversity of the possible answers, and the potential complexity of some of the scenarios, makes a series of fixed response questions artificial, cumbersome, and not very effective communication. The best approach might be to begin by asking respondents to explain in their own words how they came to be homeless. Once the situation has been explained, respondents can be asked some standard, fixed-response questions to clarify details that are appropriate to the particular situation described by respondents. Such an approach will make the interview interaction much more sensible, at the same time providing the researcher with better, more appropriate information about the actual events.

Answers in narrative form produce data that researchers sometimes find hard to work with. In particular, answers must be coded; someone must read the answers and put them into meaningful numerical categories. The diversity and complexity of answers that can result can make analysis less straightforward than when respondents answer in more structured forms. On the other hand, the result can be a better reflection of what respondents have to say.

Traditionally, when respondents answer in narrative form, coding is separate from data collection. When it is not possible to anticipate exactly what kinds of answers people will give, or how they will phrase them, considerable care can be required to set up a set of standardized coding rules that consistently classify answers into meaningful categories. When this is done appropriately, the rules and policies for classification are refined as the coding process progresses, exceptions are identified and resolved, and coding decisions are check-coded to make sure that they are being made in a consistent way.

When respondents answer questions in a fixed-response format, the answers essentially have already been turned into numerical form, ready for processing. When interviewers enter answers into a computer, the answers are ready for processing; when interviewers are using paper-and-pencil forms, the answers are ready for data entry without a further step. In that context, it is very tempting to have interviewers code answers given in open-ended or narrative form into categories during the interview, rather than introducing a separate coding step.

When the range of answers is quite predictable, the number of possible answers is small, and the rules for classifying answers are relatively simple, it may be appropriate to have interviewers do this kind of coding. For example, it is not uncommon to ask respondents to report level of education in an open-ended way, then have interviewers classify the answers into an ordered set of categories. However, if the classification scheme is at all ambiguous, having interviewers do coding is probably a mistake. Interviewers have a lot to do without having to deal with coding. Moreover, the coding scheme cannot be adapted and refined throughout the data collection process, and none of the coding that interviewers do can be check-coded. Hence, it is not possible to know whether or not they are following consistent procedures for coding, nor is it possible to train them to be better at their coding task. Finally, in studies of interviewer behavior, when questions are asked in open-ended form to be coded into specific categories, interviewers are likely to probe directly in order to ease the classification process (Fowler & Mangione, 1990).

If the structure of the answers is quite clear and the classification system easy, having interviewers classify answers may be all right. However, for most of the interesting and useful things that can be done with narrative answers, their use also requires a separate coding operation to make sure the results that emerge are reliable and valid reflections of what respondents actually say.

Finally, it is worth repeating a theme that recurs in this book: It is critical to specify as clearly as possible in the question what constitutes an adequate answer. This admonition applies particularly to open-ended questions for three reasons.

1. Questions that do not clearly define the response task adversely affect interviewing (Fowler & Mangione, 1990).
2. Variability in understanding the response task is a needless source of variation (and measurement error) among respondents.
3. Diversity of answers stemming from differences in perceptions of how to answer the questions increases the heterogeneity of answers and reduces their analytic value.

Properly designed and coded, questions that are answered in narrative form can be important contributors to good surveys. Knowing when and how to use such questions effectively is one key aspect of being skilled at the design of survey questions.