

# 4

## Survey Question Construction

### INTRODUCTION

In order to test a hypothesis or answer a research question, survey researchers must measure the concepts they reference as precisely as possible, striving always for the least amount of error. In testing hypotheses, researchers use variables to represent specific concepts. Survey questions tie the **variables** (traits or characteristics about the population that vary from person to person) to the theoretical concepts of interest. This involves paying careful attention to the design of each questionnaire item and response option, so the survey instrument is as valid and reliable as it can be for the job it must do. This chapter elaborates some practical guidelines to aid in this important task. The chapter begins with a discussion of the types of concepts you may wish to measure. We then turn the focus to question development, including several topics and examples that will help you generate precise and unbiased survey items. Following this, we address issues in the crafting of appropriate response categories for multiple choice questions, and when to use open-ended responses instead. Finally, we conclude the chapter with some closing remarks on sources of error in survey research and a summary checklist of major points to aid in the design of your own survey research projects.

The types of measures developed for the survey instrument should ideally produce relevant, unbiased, error-free data. Therefore, the first step is to set concrete research goals at the outset of the design process. Make sure you can clearly answer the key questions: What are you trying to accomplish? What information needs to be collected in order to do so? How are you going to collect this information? And how are you going to analyze the information collected? Next, make sure you can translate these needs into unambiguous questions and concepts that communicate clearly and directly with the wide range of respondents on

whom you will rely for your data (Bradburn, Sudman, and Wansink 2004). It is absolutely necessary that all respondents share a common interpretation of the meaning of each question, and furthermore, that this interpretation mirrors your intent as a researcher.

The importance of the above questions cannot be overstated. Each element—research goals, data collection, and methods of analysis—shapes and constrains the others. For example, the design of questions and response options depends largely on the mode of delivery of the questionnaire—telephone interview, online survey, pen-and-paper mailer, or face-to-face interview. Furthermore, the format of individual questions—such as open- versus closed-ended—informs the type of analyses you can later perform. For example, an open-ended response to a measure about one's attitudes toward abortion would not lend itself easily to the tight quantification required for a statistical analysis. The researcher must thus always be mindful of the optimal coordination of all of these elements for the most successful survey design.

## CONCEPT MEASUREMENT: TRAITS, ASSESSMENTS, AND SENTIMENTS

The primary goal of survey research in the social sciences is to gather data on individuals' demographics, behaviors, personal details, attitudes, beliefs, and opinions. This might seem like an easy task—akin, perhaps, to asking people questions in everyday, real life—but there are actually a number of measurement issues that must be considered in order to develop a high-quality questionnaire.

The first issue relates to the complexity of the concepts about which you would like to ask. Many measures are well established by previous survey research—such as asking one's age and gender—and relate to relatively *simple* (i.e., easily measured) concepts. Other concepts—perhaps those that are more novel, complex, abstract, or unobservable—require a more cautious and carefully strategized approach in order to avoid bias while eliciting accurate responses. Developing questions that are phrased correctly will help lead to more reliable and valid data collection efforts.

### Simple Concepts

As mentioned above, some questions are more easily answered than others, such as asking a respondent's age and gender. While referencing a slightly more complicated concept than these, **single construct** measures such as, "What is your religious preference?" similarly tap into a single, well-understood concept and are likewise typically unproblematic when encountered in a questionnaire. Such straightforward and well-established measures and questions tap into **unidimensional** or **simple concepts** that reflect a single idea, attitude, or behavior.

## Complex Concepts

### *Unobservable and Multidimensional Concepts*

A survey question is a measurement tool, describing your respondent by reference to a singular concept (like age or gender), in something of the same way that a tape measure is a tool that singularly describes an object by its height. However, at times the concept you wish to measure may be something that is not so easily observable, something that lacks an established, singular metric to describe it. Depending on your research goals, a more complex and multidimensional measure may be necessary to describe this obscure construct, just as some purposes might suggest the combination of a tape measure *and* a scale to take a multidimensional measurement of an object's height and weight.

A **multidimensional concept** is one that combines multiple singular constructs or attributes of an object in order to compose some new, abstract attribute that cannot be directly observed or measured. It is only by combining multiple questions that measure related singular concepts that this composite, multidimensional concept can be indirectly described. However, this is not to say that merely asking multiple questions about a single topic automatically produces a multidimensional concept. A group of related questions that yield similar answers aggregated into a single score are still considered to be unidimensional if they ultimately reflect the same underlying concept. On the other hand, multidimensional constructs require researchers to ask questions that span *multiple* concepts, from which combination a more complex and novel concept actually *emerges*.

For example, a researcher might ask various questions all related to particulars of respondents' religious practices: what church they attend, what religious texts they use, the particular rites and practices they observe, et cetera. Yet each of these questions, and even the aggregate score resulting from their combination, aims at describing the same concept: their religious affiliation.

But perhaps rather than a taking simple measure (or measures) identifying an individual's religious affiliation, a survey researcher might want to assess an individual's religious spirituality, behavior, *and* attitudes for a more comprehensive assessment of an individual's *religiosity* (i.e., a person's unobservable, overall level of religiousness). See Figure 4.1 for an example of how a researcher might take measurements across the various facets composing this multidimensional concept.

Here, *religiosity* is a multidimensional concept, because it measures *not only* spirituality *but also* religious behaviors *and* religious attitudes—all similar but separate components of an individual's level of religiousness. Of course, religious attitudes, religious behaviors, and spirituality are probably closely associated with one another, so the responses on each question in Figure 4.1 may be similar. However, the responses related to *religious behaviors* would likely be more closely associated with one another than with responses related to *spirituality* or attitudes—evidence of the essential independence of the composite dimensions.

Figure 4.1 Multidimensional Measures for Religiosity

	<i>Strongly Agree</i>	<i>Agree</i>	<i>Neutral</i>	<i>Disagree</i>	<i>Strongly Disagree</i>
Religion is important in my life. ( <i>attitudinal dimension</i> )	1	2	3	4	5
I regularly attend religious services. ( <i>behavioral dimension</i> )	1	2	3	4	5
I feel connected to a higher power. ( <i>spiritual dimension</i> )	1	2	3	4	5
Religion is the one true path to eternal life. ( <i>attitudinal dimension</i> )	1	2	3	4	5
I frequently read religious literature ... ( <i>behavioral dimension</i> )	1	2	3	4	5
I believe that religion is sacred. ( <i>spiritual dimension</i> )	1	2	3	4	5

## QUESTION DEVELOPMENT

Certain types of questions produce more accurate data than others (Schwarz 1999). As mentioned above, questions about the unidimensional concepts of age and gender tend to produce highly accurate results. But other questions—especially those measuring behaviors, opinions, and attitudes—may yield less precise responses. This is why survey research questions require careful calibration. Questions about behaviors, opinions, and attitudes must be carefully worded, so researchers can obtain responses that might be difficult for a respondent to communicate and express.

### Guidelines for Developing Questions

#### *Aim for Simplicity*

**Be concise; use simple, clear language; and avoid vague, abstract terms.**

Recall the lessons proffered in chapter three about respondent burden and fatigue. When a question is too long or too complex, or it contains abstract language (Converse and Presser 1986), respondents may be unwilling or unable to complete the survey, may

stop paying attention to the question, and/or become annoyed—possibly *more* annoyed—with the entire process. Limited attention or annoyance may lead respondents to provide rushed and potentially incorrect responses or to drop out altogether, further affecting the quality of the data you are able to collect. Thus, the best way to collect high-quality data is to keep survey items short, simple, and clear. Use as few words as possible to ask questions that everyone will understand *in the exact same way*. The examples below will lead you through this (and subsequently other) issues in question design by comparing some improperly worded sample questions with their improved, more efficient counterparts.

1a. When did you move from Philadelphia to Los Angeles?

This is an unclear question that does not indicate exactly what type of answer is required. For example, an individual could answer “after high school,” “when I was 17,” “in 1999,” or in another way. In contrast, a more clearly articulated question would read,

1b. In what year did you move from Philadelphia to Los Angeles?

2a. Do you favor or oppose the systemic reform of immigration policies that will assist lawmakers with adequately addressing delays in visa processing and the enforcement of contemporary immigration laws?

This question is just too long and too complex. (In fact, it was difficult to even *write* it without losing interest!) If a question requires a second, third, or fourth read to be completely understood, this is a surefire sign that it is too long or too complex to include in a survey. The following question is much more concise and effective:

2b. Do you favor or oppose immigration reform policies in the United States?

3a. How important are family values to you?

The problem with this question is that *family values* could mean a number of different things to different people. Could family values be related to spending time with one’s family? In that case, exactly how much time spent with one’s family would constitute “family values?”

For some socially conservative individuals, family values may mean opposition to premarital sex, same-sex marriage, and reproductive rights. For many socially liberal individuals, family values might imply accepting same-sex partnership adoptions, embracing nontraditional family forms, and providing financial assistance for underprivileged families. Thus in many respects, this question is not adequately measuring a

single construct that is *interpreted the same way by all respondents*. A better idea is to remove the guesswork by phrasing the question to clarify exactly what you mean:

- 3b. How much do you enjoy spending quality time with your immediate family?
- 4a. Do you favor or oppose the use of assisted reproductive technology?

Researchers should avoid using technical language and jargon in their surveys. *Assisted reproductive technology*, a term commonly used by fertility specialists, may not be easily understood by all respondents. Although the practice of using technical language may appeal as a way to sound more professional, it can easily make questions appear unclear and confusing to respondents unfamiliar with the concept or phrasing. In other words, although the survey may sound less formal, using a conversational tone can sometimes yield higher quality data:

- 4b. Do you favor or oppose the use of fertility treatments in order to conceive children when an individual is unable to do so through sexual intercourse?

### *Be Specific*

- 5a. Are you a “social drinker?”

This question is vague, because not all respondents will interpret the term *social drinker* to mean the same thing. For some people, social drinking could simply refer to enjoying a few drinks with friends once or twice a week. For others, casual drinking with friends might entail multiple drinks across several hours, possibly extended across several nights per week. In the latter case, you might have classed respondents as *heavy drinkers* rather than agreeing with their own interpretation and identification as social drinkers. A more direct question might ask exactly how many drinks an individual consumes in a given time period:

- 5b. In the past 30 days, how many alcoholic beverages have you consumed?

### *Avoid Double-Barreled Questions*

There are times when you must ask multiple questions about a topic to obtain the information you desire; this requires the separation of these questions into completely distinct survey items. Failing to do this creates **double-barreled questions**, “two-in-one” questions that are problematic because a respondent might be willing or able to answer only a single part of the question:

- 6a. How important are family get-togethers to you and your family?

This two-part question is really asking two questions: How important are family get-togethers for you, *and* how important are family get-togethers for your family? The researcher's assumption here that these two answers are always in agreement may confuse, discourage, or mislead a respondent and lead to lower quality data. In addition to this, the latter half of the question is asking the respondent to report on *someone else's* feelings toward something—a task that is virtually impossible to do! In most instances, this type of question can be easily broken out into two separate questions. For this particular example, however, it is better just to eliminate the second, highly speculative component altogether:

- 6b. How important do you think family get-togethers are?
- 7a. Do you favor or oppose the use of fertility treatments in order to conceive children when an individual or couple is unable or unwilling to do so through sexual intercourse?
- Favor     Oppose

This question introduces two additional confounding variables (individual vs. couple, unable vs. unwilling), each of which is problematic on its own. What if a respondent favors treatments for couples but not individuals? Or for people unable to conceive, but not those who are merely unwilling? In combination, this creates the potential for multiple different answers (technically, 16!) that are hardly served by the two response options provided. The researcher needs to decide whether all of these variables are truly of interest and then separate the question into multiple, more precise items:

- 7b. Do you favor or oppose the use of fertility treatments in order to conceive children when a couple is unable to do so through sexual intercourse?
- 7c. Do you favor or oppose the use of fertility treatments in order to conceive children when an individual is unable to do so through sexual intercourse?
- 8a. Do you think that capital punishment is an archaic form of discipline and that the federal government should abolish it?

This is another example of a double-barreled question—it assumes the respondent is going to respond to both parts of the question in the same way, either affirmative or negative across the board. However, if respondents do *not* feel that capital punishment is an outdated form of punishment, but *does* nevertheless have some other reason to oppose it (or perhaps even vice-versa, approve of archaic discipline), they may be conflicted about how to answer, adding burden and frustration. A more appropriate design for a question of this sort is to separate elements, or use a skip pattern if you are truly only interested in respondents who agree that capital punishment is archaic:

- 8b. Do you think that capital punishment is an archaic form of discipline? (And then, depending on response):
- 8c. Do you think that the federal government should abolish capital punishment?

*Misleading Single-Barreled Questions.* Certain phrases may appear to be double-barreled (because the word *and* appears) when they are simply using established terminology that refers to a single construct. For example, “Do you own your home free and clear?” is a question that may seem to be asking two questions in one. However, “free and clear” is a term in property law that indicates that property is owned outright (without an outstanding mortgage or lien). It is important to understand and be mindful of these terms when crafting survey questions with the simplest and clearest possible terminology.

### *Avoid Biased and Leading Questions*

Designing questions with accurate, unbiased, and simple phrasing, especially when measuring complex, multidimensional, and intangible concepts, is one of the most difficult parts of designing a survey. A significant problem related to question phrasing in survey research design is that researchers may bias their results by “leading” respondents to answer questions in a specific way. **Leading questions** often use strong, biased language and/or contain unclear messages that can manipulate or mislead a respondent to answer in a specific way. Even the most ethical researchers can still do this unintentionally, if they are not attentive enough to this common pitfall. Consider the following questions:

- 9a. In the wake of the largest economic downturn since the Great Depression, did you support our Republican-led Congress’s 2013 decision to shut down the government?

The use of the phrases “largest economic downturn since the great depression” and “Republican-led congress” makes it difficult to disentangle whether the responses you receive are reactions to the strong language or even the unnecessary detail about our current congressional profile. Instead, consider the following:

- 9b. Did you support Congress’s 2013 decision to shut down the government?
- 10a. You wouldn’t say that you support affirmative action in California, would you?

This question is judgment laden, suggesting strong disapproval of affirmative action that encourages a respondent to answer “no.” Questions like this make too clear the intentions of the researcher and color the resulting data to the point that it is useless. A less biased question might ask this:



- 10b. Do you agree or disagree with affirmative action in the state of California?
- 11a. Did our president, Barack Obama, make a mistake when he enacted the Patient Protection and Affordable Care Act, which forces individuals into universal health care?

This question uses emotionally suggestive language such as “make a mistake” and “forces individuals” that may bias an individual toward taking a specific stance. There is an even more subtle bias in this example: With the word *our* before *president*, the question may appeal to nationalism and pressure respondents to answer favorably. (To be fair, the word *our* in this example is unlikely to create much confusion; nevertheless, it is certainly possible.) Such a subtle bias shows how survey respondents are influenced in many indirect and complex ways. It is, therefore, important to be vigilant and consistently use the most balanced, inoffensive, and unbiased language possible. Consider this less biased (and much simpler) alternative to our previous example:

- 11b. Tell us how much you agree or disagree with the following statement:  
Enacting the Patient Protection and Affordable Care Act benefited Americans.
- 12a. Do you favor or oppose the use of fertility treatments in order to conceive children when an individual is unable to do so through more traditional means?

Though it is commonly used, the word *traditional* can potentially be a very loaded term, implying that something is deep-rooted and established. Respondents may be influenced to respond based on their general attitudes toward tradition and change, rather than on careful consideration of the specific issue at hand. For example, people who are resistant to or unnerved by change might be more inclined to oppose fertility treatment simply because it bears the marker of something unconventional or conceptually foreign. A better question avoids introducing this potential bias while also adding greater specificity:

- 12b. Do you favor or oppose the use of fertility treatments in order to conceive children when an individual is unable to do so through sexual intercourse?
- 13a. The tragedy in Newtown, Connecticut, has motivated Americans to enter the debate on gun control. What do you believe is the root cause of gun violence in America?

This question is appealing to respondents’ emotions by referencing the “tragedy in Newtown” before asking a question about gun control. A less biased question will instead get to the point without sensationalism:

- 13b. What do you believe are the primary causes for gun violence in America?  
(Please mark all that apply.)

*Weighing Bias Against Straightforwardness.* Offensive and biased terms are occasionally difficult to avoid, such as when such terms are part of widely recognized colloquialisms to which no clear alternative is available. For example, a question probing whether or not an individual supports “partial birth abortion” is likely easier to answer than a question asking about “abortion in the instance of intact dilation and extraction.” Although “partial birth abortion” may seem like a biased (and possibly even offensive) term, it may still elicit a more valid response than “intact dilation and extraction,” a phrase likely to be unfamiliar to the vast majority of respondents. However, even in such cases, the inclusion of such potentially loaded terms and questions in your work will surely be identified and criticized by readers and peers. Be aware of this potentiality, and carefully weigh all phrasing options so that you can defend your choices later.

### *Avoid Making Assumptions*

Premising a question on a controversial assumption can be just as problematic as the use of such biased language as described above. Consider the question below:

- 14a. In your opinion, does the increase in work hours among employed mothers have an influence on the lack of respect children now have for their families?

This question makes two assumptions that might introduce error into the responses. First, the question proclaims that there has been an increase in work hours among employed mothers. The respondent may or may not even agree with this assertion. Second, the question assumes that the respondent agrees that youth have lost respect for their families. Therefore, even if respondents have no opinion on this issue and accept that there has been an increase in work hours among employed mothers, they are still unable to answer accurately, because they do not agree that children have a lack of respect for their family. As is typically the case, simplicity and brevity can reduce the burden on respondents and improve the precision and quality of their responses:

- 14b. Has there been an increase in work hours among employed mothers recently?

### *Carefully Ask Personal and Sensitive Questions*

Often, researchers need to ask for personal and sensitive information that respondents may be hesitant to provide out of concerns for their privacy. **Item nonresponse** (i.e., skipping questions) is normally higher for these questions (Tourangeau and Yan 2007). Try to reduce the potential sensitivity regarding such personal questions with careful phrasing and question/response structures that make items quicker and easier to answer. Some common examples of sensitive questions include those relating to income, voting behavior, and the more intimate details of peoples' private lives.

Sensitive questions may be more successful when possible responses are presented as **bracketed categories**, in which each response choice encompasses a range of numbers or categories defined in relevance to the question (Tourangeau and Yan 2007). For example, asking respondents to report their number of sexual partners may elicit a greater number of responses when followed by a range of choices like the following: 0, 1, 3–5, 6–10, and 11 or more. Identifying a category may feel less like a revelation of specific, personal information, and may thus encourage a higher response rate at the expense of a little bit of precision. This approach is helpful in another way, since some respondents might not remember their *exact* number of sexual partners—and the longer they think about it, the more burdensome the question becomes.

Above all, do not force respondents to answer any questions. Respondents who feel they are being coerced into answering a personal question may skip the question at best, or abandon the survey altogether at worst. Simply adding a response option of “decline to state” can make the difference between a skipped item and an abandoned survey, and preserve respondents’ trust in the sensitivity of the researcher.

*Social Desirability.* **Social desirability** refers to the inclination for respondents to overreport socially acceptable behaviors and underreport socially undesirable behaviors (Krumpal 2013). For example, it is not socially desirable to express homophobic or racist attitudes, so respondents may be apprehensive about admitting these behaviors. This tendency is especially marked in face-to-face interviews, where the presence of an interviewer can magnify respondents’ perceptions of being judged. In the same vein, respondents will underreport socially undesirable behavior, such as illegal substance abuse; again, this is even more likely in a face-to-face setting. In addition, earlier advice about avoiding bias is especially relevant here, as desirability effects can easily be triggered by loaded words like *illegal* or *abuse* (when referencing substance use; other topics will entail different linguistic sensibilities).

Privacy and social desirability are important concerns that should be taken seriously so that respondents feel comfortable answering fully and truthfully.

*Emphasize Anonymity and Confidentiality.* Again, it is important to emphasize that survey responses will remain anonymous or confidential at the outset and, if necessary, reiterate this assurance when touching on sensitive subjects.

*Choose Delivery Mode Carefully.* There is more response bias and social desirability in face-to-face interview surveys.

*Organize Sensitive Questions Strategically.* Do not start a survey with personal or sensitive items or questions prone to social desirability. When respondents begin a survey, their level of commitment is usually quite low, and personal questions early on may

deter them from continuing. Also, placing these questions at the end of the survey is ill-advised, because this might lead to respondents feeling apprehensive, unpleasant, or offended, as though they are being manipulated. This may even introduce selection bias into future results when follow-up surveys are necessary and respondents are unwilling to participate. It is best to include personal and sensitive questions in the middle of the survey, so those respondents who have committed time and energy to completing the survey will still want to complete it.

### *Additional Guidelines for Question Development*

In addition to the general concerns detailed above, there are a number of minor details that, left unaddressed, may add to overall respondent burden and discourage participation and attention to quality answers. Many of these details are presented here, with positive and negative examples to illustrate each one:

- Avoid abbreviations; spell out the entire phrase instead. If the abbreviation is absolutely necessary (which is rare), define it for the respondent.

**Incorrect:** How do you feel about the GOP?

**Correct:** How do you feel about the Republican Party?

- Avoid slang and contractions.

**Incorrect:** How many kids live in your household?

**Correct:** How many children live in your household?

- Avoid ambiguous phrases, even those that are common in everyday talk.

**Incorrect:** Do you agree that abortion should be illegal most of the time?

**Correct:** Do you agree or disagree that abortion should be illegal under X circumstance? Do you agree or disagree that abortion should be illegal under Y circumstance?

- Avoid negatively phrased questions.

**Incorrect:** How frequently do you not attend church?

**Correct:** How many times in the last month did you attend religious services?

- Avoid double negative questions.

**Incorrect:** Should the Supreme Court not have opposed the right of same-sex couples to marry?

**Correct:** Do you favor or oppose the Supreme Court's ruling on the Defense of Marriage Act?

- Use a realistic time frame when asking about attitudes and behaviors.  
**Incorrect:** How many cigarettes have you smoked in your entire life?  
**Correct:** In the past month, how many cigarettes have you smoked?
- Make sure all questions are absolutely necessary.

One of the most important ways to elicit high-quality responses to survey questions is to have an organized questionnaire with questions that are clearly and impartially articulated. Verification that questions are relevant and do not contain (potentially inaccurate) assumptions will also safeguard against measurement error.

Once you are satisfied that you have included all the relevant questions you need for your research purposes, excluded irrelevant and unnecessary items, and arrived at question wording that is clear, concise, and unbiased, the next step is to focus on the response options you provide for the respondent to these questions. It is to this area of survey design that we turn our attention in the sections that follow.

## RESPONSES TO QUESTIONS

### Types of Responses

There are many options to consider in the way you allow respondents to answer your survey questions. For some purposes you may wish to allow them to speak freely in their own words; other times, it will be more effective and efficient to provide them a range of responses from which to choose. The type of response you solicit will depend on the nature of the concept you wish to measure (and the research question or hypothesis that underlies it) as well as the subsequent analysis you intend to perform with the data collected. The sections below outline the strengths, weaknesses, and special considerations of the various response options, to help you choose the ones best suited to your specific research needs.

### Closed-Ended Questions

**Closed-ended questions** are questions formatted such that the response possibilities are limited (“closed”) to a specific list from which the respondent must choose—they are also referred to as **fixed-choice questions**. There are several different types of closed-ended questions, and depending on the concept being measured, some are more appropriate than others. Your questionnaire will likely involve a combination of different types and may require separate sets of instructions to ensure respondent comprehension. Nevertheless, questions of different types must flow together seamlessly in order for you to collect first-rate data.

## Strengths of Closed-Ended Questions

The strengths of using closed-ended questions are mostly on the back end, in the ease they bring to the compilation and analysis of data by the researcher. Strengths include the following:

- Responses are easy to quantify (e.g., 1 = Strongly Disagree; 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly Agree).
- They are easier to enter into statistical analysis software and analyze.
- Data collection is usually quicker.
- Results are easier to summarize and present in tables, charts, and graphs.
- There is more reliability across responses—especially when only a few collapsed response categories are included.
- There is less interviewer and social desirability bias.
- There is a higher degree of anonymity.

## Types of Closed-Ended Questions

### *Multiple Choice*

Most readers will be familiar with the standard **multiple-choice** format so prevalent in the exams we encounter in our years of schooling. The only difference in survey questionnaires is that, ideally, less guesswork will be involved.

Which of the following is the mode of transportation you most frequently take to work?

- 1) Automobile (self-driven)
- 2) Automobile (carpool)
- 3) Bus
- 4) Train
- 5) Other (please specify) \_\_\_\_\_

As in all aspects of survey design, it is important to check that all response categories are clear and unambiguous, that they do not suggest different meanings to different (types of) respondents. For example, questions asking respondents about their occupations often list “education” among the options. However, “education” does not mean the same thing to everyone. “Education” includes students, teachers (at various levels), and administrators. These are all quite different but would be categorized as the same if the response options are ambiguous.

*Dichotomous*

When there are only two possible response options (e.g., agree/disagree, yes/no, true/false), the question is **dichotomous**. Technically, this is a type of multiple-choice question, but with only two options, which makes it simple to design. However, keep in mind that questions rarely have only two possible answers, and so responses to dichotomous questions are easy to misinterpret. They also make it even easier to respond incorrectly or thoughtlessly for ambivalent respondents already prone to random guessing. Consider whether the item below is well suited for a dichotomous type of response:

My neighbors are an important part of my life.

Yes    No

*Checklist*

A **checklist** is appropriate when you want to allow the respondent to select multiple responses. Note the use of check-boxes (rather than numbers or letters) to identify separate items, to remind respondents that choices are not mutually exclusive. For example:

In the past 30 days, which of the following have caused you a lot of stress? (Check all that apply).

- My friends
- My partner
- My job/finances
- My family
- My children
- None of the above

*Scales*

**Scales** provide a set of response options representing ordered points on a continuum of possible answers. There are several distinct types of scales suited to different types of applications, as outlined below.

**Rating Scale.** **Rating scale** questions are a type of multiple-choice option that uses ordered responses to represent a continuum from which respondents choose the single best answer choice. It is often helpful to include additional explication of the categories in parentheses to ensure all respondents (and researchers) interpret them in the same way. Consider this example:

How often are you late for work?

- 1) Very frequently (almost daily)
- 2) Frequently (twice a week)
- 3) Occasionally (once a week)
- 4) Seldom (twice a month)
- 5) Rarely (once every six months)
- 6) Never

*Rank Order Scale.* A **rank order scale** allows respondents to put answer choices in order themselves, according to some criteria expressed in the prompt. Note here that blanks or brackets should be used to differentiate the choices from a rating-type or checklist-type question. Here is an example:

If you had to choose an alternate mode of transportation to work, what would be your order of preference among the following? (1 is first choice, 2 is second choice, etc.)

- Walk/run
- Bicycle
- Skateboard
- Rollerblade
- Taxi cab

*Likert Scale.* In a **Likert scale**, participants are asked to indicate their agreement or disagreement with a statement (or number of statements) by scoring their response along a range. Responses typically range from “strongly agree” to “strongly disagree,” with each response option on the scale associated with a numerical score. Likert scales are particularly helpful when measuring respondents’ attitudes and opinions about particular topics, people, ideas, or experiences. Likert scales should not be used when the responses are not on a scale or when the items are not interrelated—these would, in effect, not even be called scales.

Typically, Likert scales will have a midpoint for a neutral response between agree and disagree (see below). However, researchers sometimes use an even number of possible responses, to the exclusion of a neutral/undecided option. This is called a **forced choice question**, because ambivalent respondents are forced to form an opinion in one direction or the other. The issue of whether or not to include a midpoint is addressed on page 62.

Overall, I feel that the current U.S. president is . . . :



		<i>Strongly Agree</i>	<i>Agree</i>	<i>Neither</i>	<i>Disagree</i>	<i>Strongly Disagree</i>
1	Trustworthy	[1]	[2]	[3]	[4]	[5]
2	Strong	[1]	[2]	[3]	[4]	[5]
3	Capable	[1]	[2]	[3]	[4]	[5]
4	Intelligent	[1]	[2]	[3]	[4]	[5]

*Semantic Differential.* **Semantic differential scales** provide contradictory adjectives as endpoints on a Likert-type scale where the respondent can assess a person, idea, or object according to a dimension of special interest to the researcher (Osgood, Suci, and Tannenbaum 1957):

Indicate your attitudes regarding the current president of the United States on the scale below:

Trustworthy	[1]	[2]	[3]	[4]	[5]	[6]	[7]	Untrustworthy
Strong	[1]	[2]	[3]	[4]	[5]	[6]	[7]	Weak
Capable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	Incapable
Intelligent	[1]	[2]	[3]	[4]	[5]	[6]	[7]	Unintelligent
Respectful	[1]	[2]	[3]	[4]	[5]	[6]	[7]	Disrespectful

### Important Guidelines for Categorization Scheme Development

When developing response categories for closed-ended/fixed-choice questions, there are three important things to keep in mind: relevance, comprehensiveness, and mutual exclusivity.

#### *Relevant*

Just as appropriate *questions* need to be designed with the population and topic in mind, *response* scales and categories must be designed with the same considerations. Researchers need to be familiar with the most **relevant** or common types of answers given in order to choose relevant response options. Response options can be based on common knowledge or can be identified through research or by asking individuals

involved with or particularly knowledgeable about a specific topic. Appropriate response categories vary depending on culture, time, and even geographic region of the country. The goal is to make sure that all of the most important (and expected) possible response categories are listed, so respondents will not need to struggle to align their own ideas with the options available to them.

### *Comprehensive*

On a distinct but related note, response options must reflect a **comprehensive** list of the possible options, exhausting all categories a respondent may wish to select. The best way to ensure a comprehensive category set is to include an “other” category at the end of a list, followed by the instructions “please specify” and a blank line for respondents to write on. This is also useful for future survey designs: If a large number of similar open-ended responses are noted, a researcher can create a new category option for it. Unless the researcher is 100% confident that all categories are covered, then an “other” category is necessary. As sociologists, we know that gender is composed of more than the typical two categories (e.g. male/female); therefore including an “other” response option will improve participant responses. A lack of comprehensive options is both common and very frustrating for individuals excluded by the choices (especially with regard to overlooked racial and ethnic categories). The examples below illustrate the difference between a noncomprehensive list and a comprehensive one:

What is your marital status?

- 1) Single
- 2) Married

What is your marital status?

- 1) Single
- 2) Married
- 3) Divorced
- 4) Widowed
- 5) Separated
- 6) Never Married

### *Mutual Exclusivity*

Having **mutually exclusive** categories means that the response options do not conflict or overlap with each other in any way. In other words, respondents must perceive

that only one of the available responses fits the answer they imagine (unless, of course, the item includes a checklist of responses). If not, respondents have the burden of deciding which is closest or most appropriate—and how they choose to do so may vary from survey to survey, lowering the precision and overall quality of your data. Note that this is the opposite of the previously discussed problem, that of a lack of comprehensive categories: In this case, there are *too many* possible choices instead of *not enough*. For example, consider the following question:

How long have you been seeing your primary care physician?

- 1) 1–2 Years
- 2) 3–5 Years
- 3) 6–10 Years
- 4) 11 or more years

How would an individual respond who has been seeing a primary care physician for 2.5 years? Do not assume everyone will round up—people who are rarely sick or do not frequently see a doctor may be inclined to round down. You also should not assume everyone will round down—people who like their doctor and frequent the doctor’s office may round up. It is important to exercise caution when creating numerical categories where an individual might land in between two possibilities. Sometimes this requires instruction to round up or round down. Other times, it may be a better idea to simply add specificity to your presented options.

In recent years, it has become common to use an “infinite cap” on response options based on the last category listed. In order to avoid confusion, do not list categories using a plus sign (e.g., 1–2 times, 3–5 times, 6–10 times, 10+ times). An individual may not realize that “10+” implies “anything more than 10” and may interpret “10 times” as being included in this categorization. A better response option will spell out “11 or more” to avoid this confusion.

### *The Problem of the Neutral Point*

As noted at the beginning of this chapter, the testing of a hypothesis or answering of a research question requires survey researchers to measure concepts as precisely as possible. Therefore, it is necessary to decide whether or not a midpoint or neutral point in a scale is helpful for your specific research purposes. There is no definitive evidence whether or not the midpoint is valuable—the best answer is that it depends on the respondent and the type of question being asked (Kulas and Stachowski 2013; Krosnick 2002).

When researchers argue that we should not include a midpoint, they are basically claiming that respondents should always be forced to take a stance on a given topic.

They also suggest that neutral points are potentially meaningless and offer no insight on an individual's real opinions or attitudes. However, the unavoidable reality is that people sometimes have neutral feelings. There are several scenarios when this is conceivable, such as when a respondent

- Lacks interest in a topic.
- Has limited recall regarding the event(s) in question.
- Is legitimately undecided on an issue.
- Lacks knowledge about a topic.
- Lacks experience relating to a topic.
- Finds the question too personal to answer.

In such cases, respondents may randomly guess, skip the question entirely, or abandon the survey altogether. (Recall from the previous section on sensitive questions that respondents may be hesitant to complete a survey when they feel forced to answer a question.) Therefore, given the topic, question, and types of respondents participating, a researcher should choose carefully whether or not a neutral point is advantageous for the study, whether the value of "forcing their hand" outweighs the risks of introducing error or additional respondent burden to the survey experience.

In addition to these general concerns, there are some common, more specific sources of error, as detailed in the next section.

### *Additional Guidelines for Response Options*

- Use an equal number of positive and negative responses for scale questions. To understand why this is important, consider the question below, where 75% of the answers indicate some form of close relationship, and there is only one possible option for something other than close:

**Incorrect:** How emotionally close would you say you are to your mother?

- 1) Extremely close
- 2) Somewhat close
- 3) Close enough
- 4) Not close at all

**Correct:** How emotionally close would you say you are to your mother?

- 1) Very close
- 2) Somewhat close
- 3) Somewhat withdrawn
- 4) Very withdrawn

- Use a consistent rating scale throughout the entire survey. Do not mix questions with a scale where 1 = strongly *agree* with other questions with a scale where 1 = strongly *disagree*.
- Similarly, do not mix scale ratings within the same survey. Choose the standard number of points in your Likert scales (e.g., three, five, seven) and be consistent to avoid confusion.
- Limit the number of points on scales. Any more than nine points will hinder the respondent's ability to discriminate between points. Scales with five or seven points are common and reliable.
- Neutral is not the same as "no opinion." "Neutral" might actually be an opinion. Individuals may feel "neutral" about the hours they work per week, but this is not the same as having no opinion on the matter.

The above guidelines should sensitize you to some of the advantages, drawbacks, and special considerations surrounding the various types of closed-ended questions. In the next section, we turn to a very different type of survey item, the open-ended question, and elaborate a similar discussion of strengths, weaknesses, and other issues.

## LEVEL OF MEASUREMENT

The end result of collecting survey data is to quantify our concepts into variables and then analyze them statistically to test our hypotheses (Nunnally and Bernstein 1994). The types of statistical analyses that are possible depend entirely on the levels at which variables are measured. To allow for our chosen type of statistical analysis, we must know, and even plan ahead for, the level of measurement of our variables. **Level of measurement** can be defined as the mathematical property of variables. As the precision of measurement increases, more mathematical options or tools become available with which to analyze the variables statistically. This ultimately suggests that the process of measurement is an ongoing iterative process that needs to consider both theoretical and empirical needs of the research (Carmines and Zeller 1979).

An important reason for choosing a particular set of response options for each survey question is that the response options dictate the level of measurement of the variables we will use to answer our research question. We are literally assigning numbers to attributes of the variables. Note that in the examples of survey questions presented above, all the response categories were tied to a number. This is measurement in a nutshell and the end product of all our hard work developing questions and assessing validity and reliability.

We classify variables with four levels of measurement that we will present in order from the lowest level of precision to the highest level of precision: nominal, ordinal, interval, and ratio. Again, precision is the degree of specificity of the numbers assigned to specific response options. Nunnally and Bernstein argue that measurement is really about “how much of the attribute is present in an object” (1994: 3). Therefore, as precision increases, we are better able to gauge how much of the attribute we have observed. Another way of thinking about it is that lower levels of precision mean there are fewer mathematical operations we can use with the variables because it’s harder to estimate the quantity we have, while greater precision increases the mathematical options we have.

**Nominal** measures have the least precision. Nominally measured variables are actually qualitatively measured variables rather than numerically measured variables. That is, we cannot evaluate *how much* of an attribute is present; rather, we can only know whether the attribute *is or is not* present. Nominal variables can be distinguished only by their names. Favorite type of fruit is an example of a nominal variable. Response options include apples, oranges, grapes, bananas, et cetera. We can assign a number to apples (1), oranges (2), and bananas (3), but the number is essentially arbitrary. We cannot say, for example, that there is more favorableness associated with bananas than is associated with apples based on the number assigned. Any number used with a nominal variable is used as a label only (Nunnally 1967). Apples, oranges, and bananas are qualitatively distinct, and all we can do is examine whether or not participants chose apples (in the set of apples) or did not choose apples. A few other examples of nominal variables are gender (male, female, other), race/ethnicity (white, African American, Asian, Native American, Hispanic, multiracial), and participation in food stamp programs (yes, no).

It is possible to measure nonnominal variables nominally. For instance, let’s say a research team is interested in the food security of Americans who are living below the poverty line. They use income as their primary variable of interest, because it defines the US poverty line. However, the researchers may decide to measure income—a variable that is easily and precisely quantifiable, simply as “yes” for below the poverty line or “no” for not below the poverty line. This limits what the researchers can do with income. They cannot tease out how variations in income levels among those below the poverty line may affect food security, because that information was not gathered; the nominal variable used lacks the precision necessary to conduct that analysis.

**Ordinal** measurement quantifies variables by ordering the response categories from least to most or most to least (Nunnally and Bernstein 1994). The following question is measured ordinally, because the three response categories are rank ordered from least problematic to most problematic.

Upon moving to your new home, please tell me to what extent your commute to work may be a problem.

Not a problem

Somewhat of a problem

A big problem

The numbers assigned to the response categories are rank ordered, with 1 being one unit less than 2, which is one unit less than 3. There is the same distance numerically between each category. Yet, the difference or distance between “not a problem” and “somewhat of a problem” may be bigger or smaller than the distance between “somewhat of a problem” and “a big problem.” In terms of mathematical computation, this is limiting. All we really know is that 1 is less than 2, but we cannot calculate the actual distance between the ordered categories. This means we cannot add or subtract across the scores, nor can we multiply or divide the values of this variable. Most Likert scale variables, such as opinion or attitude scales with five category response options, are ordinal—going from least amount of agreement to most amount of agreement without meaningful numbers to determine the distance between the categories.

An **interval** level of measurement is one that rank orders response categories and, additionally, provides known distances between each response category without providing information about the absolute magnitude of the trait or characteristic (Nunnally and Bernstein 1994). Conceptually, this is very difficult to understand without an example. Temperature is an interval level of measurement. In the United States, temperature is measured on the Fahrenheit scale. On this scale, 32 degrees is considered freezing. In other parts of the world, temperature is measured on the Celsius scale where 0 degrees is considered freezing. Therefore, 32 degrees Fahrenheit and 0 degrees Celsius are the same value on different scales. The zero in Celsius is meaningless, though the “freezing” designation is meaningful. The difference between 0 degrees and 1 degree Celsius is clearly understood as a 1-degree change, as is the difference between 32 degrees and 33 degrees Fahrenheit. But, 33 degrees Fahrenheit is not equivalent to 1 degree Celsius. When the absolute magnitude of the trait is not specified, we do not have a zero score that is consistently anchored to a meaning. Mathematically, then, we can add and subtract values with interval measured variables, but we cannot multiply or divide them. Another example of an interval measured variable is intelligence quotient (IQ), which is measured on a scale created from many questions and is standardized to have a mean of zero. Thus, the zero has no true meaning, and we cannot say that someone with an IQ of 100 is twice as smart as someone with an IQ of 50.

**Ratio** measurement is the most mathematically precise of the levels of measurement. Variables that are measures on the ratio scale have response categories that can be rank ordered, the distances between the response categories are known, and there is a true zero value that is meaningfully anchored (Nunnally and Bernstein 1994). Income is an excellent example of a ratio measured variable. The question “How much did you earn in wages and salary in the previous year?” followed by a line for participants to write in a value will create a ratio measured version of income. A participant who earns \$30,000 per year makes half as much as someone who earns \$60,000 per year. If a person reports \$0, then that person had no earnings or wages in the previous year. The absolute magnitude of the trait is known. Thus, we can add and subtract, multiply and divide, rank order, or create nominal variables out of

income if it is measured on a ratio scale. Other examples of ratio measures are age, height, weight, and years of education.

When devising questions for a survey, think carefully about the best configuration of response categories. Measure them as precisely as possible to allow for more statistical analysis options later in the research process. Look to previous published literature to determine how the field is measuring the variable to see if there are known issues with measurement. For example, how well can people articulate their previous year's earnings and wages? If this is too hard, participants might guess, estimate, or skip the question entirely. If this is what the literature says, use a less precise level of measurement.

## OPEN-ENDED QUESTIONS

**Open-ended questions** are survey items formatted to allow respondents to answer questions or provide feedback in their own words. In contrast to closed-ended questions, there are no limitations to the response possibilities, and respondents are encouraged to provide in-depth answers. A quick caveat might be raised before proceeding: Open-ended questions are not necessarily always questions per se; they may include any type of prompt—question or not—to elicit an original and self-guided response from the respondent. Consider the following example:

**Closed-Ended Question:** How emotionally close do you feel to your mother?  
(Add response options.)

**Open-Ended Question:** Tell me how you feel about your mother.

Open-ended questions are sometimes useful when they follow a closed-ended question. This configuration ensures that researchers can learn about some specific aspect of the issue they find relevant, but it also opens the floor to detailed elaboration, or even novel issues, that the respondent finds interesting or important as well. Consider the following question posed by a researcher studying public opinion of the US Supreme Court:

1. Please rank your support for the current US Supreme Court:
  - 1) Strongly support
  - 2) Somewhat support
  - 3) Somewhat oppose
  - 4) Strongly oppose



If researchers are studying public opinion of the US Supreme Court, they are probably not going to get the information they need from this single closed-ended question—even several more questions might not elicit the fine-grained information necessary. Assuming the researchers find that most people strongly oppose the current Supreme Court, they are faced with a new problem: They are unable to determine *why* there is such strong opposition. Thus, open-ended questions are sometimes necessary in order to produce more nuanced ideas about the topic in question.

## Strengths of Open-Ended Questions

- Respondents are not limited in their responses. They are able to explain, qualify, and clarify their answers, especially when “other (please specify)” is an option.
- Open-ended questions are frequently easier to craft, because they do not require response options (which require a detailed design process).
- Open-ended responses do not force potentially invalid responses.
- Open-ended responses have more nuance, depth, and substance than closed-ended responses.
- Open-ended questions are fairly straightforward (unlike the many types of closed-ended questions and scales).

## Guidelines for Open-Ended Questions

While open-ended questions are by definition less constrained than closed-ended ones, there are still some important guidelines to keep in mind to elicit the highest quality responses in the most efficient manner.

- Avoid dichotomous questions and other questions that elicit one-word replies (agree/disagree, yes/no, true/false).
- Do not supply response options in the question:
  - **Incorrect:** Did you feel happy or sad about having another baby?
  - **Correct:** How did you feel about having another baby?
- Try not to use phrases such as “To what extent were you happy?” or “How happy were you?”
- Be aware that the rich detail and varied responses you may receive can be of great value for some subsequent analyses, but will be ill-suited for others. For example, it can be very difficult to quantify such responses for use in statistical analysis.
- Too many open-ended questions can add to respondent burden and fatigue and therefore should be used sparingly.

## COMPARATIVE CONCEPTS (VIGNETTES)

The next type of question can be either open- or closed-ended, or a combination of both. **Vignettes** are systematic descriptions of hypothetical situations; they are included in a survey to elicit respondents' thoughts and emotions about specific topics that cannot be encapsulated in the standard, concise question format (Finch 1987). They are often intended to elicit opinions, values, and attitudes arising from unique situations and/or social norms that are difficult to define or articulate. In a vignette item, respondents are presented with a hypothetical story and are asked to react to the events it describes. While stories have traditionally been presented as blocks of text, technology-based survey development (e.g., Qualtrics and Survey Monkey) has recently introduced the exciting possibility of including video vignettes in online surveys.

Vignettes can provide an opportunity for discussing more sensitive topics (for example, abortion, bullying, or racial prejudice) in a format that is potentially less aggressive and/or imposing than that of an outright question. The specific contexts vignettes provide can help respondents to feel less put on the spot to issue definitive, universal proclamations, reducing the pressure often associated with the discussion of sensitive issues. Hypothetical stories also offer the respondent specific illustrations for reference that can help them clarify some of their own otherwise ambiguous ideas or feelings. For example, rather than simply answering a question about abortion in specific circumstances (e.g., young and unprepared mother), a vignette allows respondents to read (or see) the context surrounding this situation, perhaps providing them with the cues they need to feel more invested in and committed to their subsequent responses.

### Important Guidelines for Vignettes

- If you are using a written vignette, keep it practical and realistic so that it can be recreated on video at a later time if necessary or desired.
- Try to keep the story interesting. You do not want the respondent to get bored or lose motivation.
- Be realistic and avoid sensationalism.
- Try to keep it brief enough to maintain respondents' attention, while at the same time detailed enough to include all of the points you consider relevant for the study.
- Keep in mind that bias is especially problematic here, as the many details of a narrative—such as the race, age, marital status, et cetera of a woman in a video about abortion—can all introduce sources of bias (from both researcher and

respondent) that may be subtle and largely undetectable. Even seemingly extraneous or inconsequential elements of the narrative may color the general perceptions of respondents in significant ways.

- Depending on the length of the vignette, the number of questions that follow it will vary, but five to seven questions is suitable.

## Sample Vignette

The following is an example of a vignette used to assess individuals' perceptions of later-life decision making. It incorporates both closed-ended and open-ended prompts.

Mr. and Mrs. Market are an elderly couple that were married directly out of high school and have lived in the same home for 40 years. Mr. Market has recently been diagnosed with terminal cancer, and his symptoms have become quite severe. He is unable to perform everyday tasks, can walk only very short distances, and reports that he is in constant pain. Mrs. Market is unable to attend to Mr. Market's many medical needs and is finding it increasingly difficult to mind her own health. Mr. and Mrs. Market do not have any children and cannot afford to move to a retirement home or later-life facility.

1. Of the following situations, which do you think would be best for the Markets?
  - (1) Mr. and Mrs. Market should remain in their home and apply for welfare services.
  - (2) Mr. and Mrs. Market should sell their home and relocate.
  - (3) Mr. and Mrs. Market should appeal to friends for financial assistance.
  - (4) Mr. Market should be hospitalized and Mrs. Market should remain in their home to care for herself.
2. Why do you believe this choice is the best decision for the Markets?

This concludes our discussion of the general classes and specific types of survey items you may find useful in the pursuit of your research goals. In the remainder of the chapter, we turn to data collection issues that can apply to all types of survey items. Attention to these should be part of every researcher's efforts to reduce error in the collection of survey data.

## RETROSPECTIVE AND PROSPECTIVE QUESTIONS

Researchers will sometimes expect a respondent to recall events that happened in the past or to count the number of times some event has occurred. They also often ask about opinions formed, beliefs adopted, and behaviors displayed in the past. These are

**retrospective questions.** These are often necessary in cross-sectional surveys because we only talk to participants once. Because human memory is limited and fallible, it is important for researchers to try to obtain the most accurate retrospective results possible, and they should do their best to improve respondent recall. There are two ways to improve recall:

- (1) Use *very* specific time and object references:

**Unclear:** How many times in the last year have you been to the library?

In this question, a respondent may not know how to answer, as “a year” can be interpreted differently by many people. Many things could be “about a year ago” and rounded into or out of the year timeline. People may also refer to a year as an *annual* year (e.g., an individual in September 2014 may refer back only as far as January 2014—and so they are only considering nine months). Students might interpret a year as being an *academic* year and go back as far as the prior August. Using very specific and standardized time references helps respondents frame the period in their minds to assist recall.

**Clear:** In the last twelve months, how many times have you been to the public library?

Being as specific as possible about reference dates is also important to reduce telescoping error and recall loss. **Telescoping error** is the tendency for respondents to remember things as happening more recently than they actually happened (Bradburn, Huttenlocher, and Hedges 1993). This occurs most often with recent events. **Recall loss** occurs when respondents forget that an event occurred. Recall loss occurs most often with events that occurred in the distant past (Sudman and Bradburn 1973).

- (2) Limit retrospective questions and “time frame overload”:

Questions that require recall may demand a great amount of mental energy from the respondent—and this energy is of finite supply. Therefore, it is important not to include too many retrospective items. When recall items are necessary, avoid alternating the length of time referenced (“How many times in the last week . . . ?” followed by “How many times in the last month . . . ?” followed by “How many times in the last week . . . ?” followed by “How many times in the last year?”). The mental gymnastics required to follow such questioning can easily lead to respondent burden and fatigue.

Another type of time-based question asks respondents to look forward and anticipate something happening, for instance, “At what age do you expect to be married?” “When do you plan on moving next?” and “How much more money will you be earning when you get your next raise?” When asking these types of questions, it is important to keep in mind that responses are often unreliable and inaccurate (because respondents cannot tell the future).

In a longitudinal study we can ask **prospective questions**, meaning questions about current time events, behaviors, and attitudes. If we ask these same questions multiple times, we can assess how past events affect current events or attitudes and know that recall bias has been minimized. Furthermore, we can follow up on the accuracy of future expectations or aspirations with prospective questions in a later round of data collection.

## UPDATING TIME-SPECIFIC SURVEYS AND MULTILANGUAGE SURVEYS

Often, a researcher will use measures that have been used in past research (“tried and true” measures)—this is perfectly acceptable (and even recommended). When using or building on existing instruments, it is important to retain the essence of the question but also to appropriate it according to the specifics of your own, unique study. For instance, elements may need to be changed to reflect important societal changes that have occurred since the previous study, so your survey can maintain the impression of being timely and relevant to the outside world. By the same token, some words or entire topics may need to be updated or dropped to avoid referencing a past event that has since lost its meaning for your current respondents, who may no longer remember or care much about it. For example, it would not be very helpful to include a vignette about children’s reactions to Cold War tensions that was developed in 1984, or to recycle a question about Mad Cow Disease for a questionnaire administered in 2014.

Adopting an existing measure as is does allow for valuable comparisons of newer results with those obtained by the past research. However, just because a survey or a scale or a single question has worked in the past, there is no assurance that it will work today, or that it will work with your particular types of respondents (based on cultural differences, etc.).

In addition to updating the content, it is important to update the format of your survey according to up-to-date technological advancements. For instance, a text vignette may work better as a video, now that video vignettes can be introduced into web-based survey instruments like Qualtrics and Survey Monkey. Even certain questions that were once posed on pen-and-paper surveys may need to be updated for online survey administration.

## MULTILINGUAL AND CROSS-COMPARATIVE SURVEY PROJECTS

Given the diversity of most research in the social sciences, multilingual survey projects have become more common. The most important part of conducting multilingual studies is to plan survey translation as part of the study, rather than addressing it

merely as an afterthought. When multilingual issues are taken into consideration at the outset, researchers can reduce the time and money associated with translation later (McKay et al. 1996). Furthermore, when considered early, researchers can develop survey instruments in two or more languages simultaneously to avoid a potential translation bias toward one language.

There is a reason that translation is a profession that requires rigorous understanding of different words, phrases, and meanings. A common mistake is to use web-based translation services to translate words, phrases, and entire survey instruments. For example, in the Spanish language, double negatives are often used to produce a negative, whereas in English, double negatives are confusing and usually imply a positive response. Using translation services that are anything less than professional invites more risk than is justified by whatever cost savings it generates.

It is advisable to seek the assistance of a pair of bilingual translators who are familiar with the study content and colloquialisms in both languages to **back translate** the survey. Back translation occurs when one person translates the survey into another language, and a second person (without seeing the original survey) translates this version back into the original language (Bernard 1988). The researcher is then able to check for inconsistencies in translation based on awkward and complex language in the back translation.

It is also important to account for cultural differences in cross-comparative research, especially when translating instruments from one language to another (as opposed to developing them simultaneously) (Bernard 1988). Even simple differences between “disagree” and “somewhat disagree” may bias the results of one survey. Examples provided with certain questions may differ based on culture. Even when translating US English to British English, there are culturally different terms. For example, a question about “football” will be interpreted differently depending on where and to whom the survey is administered.

Thus, cross-comparative survey conversion requires skilled translation and separate pretesting (see Chapter 6) to achieve linguistic, cultural, and conceptual equivalence—meaning that words and phrases as well as constructs and concepts are to be culturally and linguistically similar. It is also very important to document the translation process and to include the translation methods (and concerns about any potential problems with the translation) in the study report.

## MEASUREMENT ERROR

Before closing our discussion of question construction, it is worth noting that throughout this chapter, the word *error* has been mentioned several times. This is because error is an inescapable aspect of all surveys and indeed all research in general. In the social sciences, we strive to find the best way to estimate a given concept, but we also acknowledge that

to some degree, our research will always be prone to measurement error. **Measurement error** occurs when the question chosen to gather data on a particular concept does not reflect that concept accurately. It is not a valid or reliable question to some degree, and this is the topic of Chapter 5. This does not mean we ignore this propensity, however. Rather, we constantly work to reduce error by studying and defining it. To this end, social scientists have defined distinct types of error; these include nonresponse bias, selection bias, poorly designed survey questions, and data-processing errors. Delineating and addressing each of these types in its own terms allows us to make headway in the constant struggle to reduce error.

Not all measurement error is the same, and the different types of error are dealt with in different ways. The two main causes of error in survey research are *systematic error* and *random error*. **Systematic error**, also known as bias, occurs when the instrument is skewed toward a certain type of measurement or a specific (incomplete) representation of the population under study. Such bias is “systematic” because it affects all of the data collected by the skewed instrument equally.

For example, if the US Census were conducted only in metropolitan areas, there would be a systematic bias in the sampling frame. Any results would be biased in favor of urban residents and against residents of suburban and rural areas. No matter what section of the census you looked at, from population to employment, the error would be present. In another example, if students taking an exam were interrupted by a fire alarm, there would be a systematic downward bias across all of their scores that would make the entire population appear to have been less prepared for the exam than they actually were.

In contrast, **random error** is error that affects results for any single individual (in any direction), but it is expected to balance itself out in large samples. With large sample sizes, random errors average to zero, because some respondents are randomly overestimating and some are randomly underestimating.

## Primary Types of Systematic Survey Error

**Nonresponse bias** occurs when individuals who do not respond, refuse to answer, or are unable to answer specific survey items differ from respondents who *are* able or willing to answer—especially with regard to the attributes being measured. When measuring religiosity, it may be that individuals who refuse to answer certain questions are similar in certain characteristics (less religious, perhaps)—and different from those who do answer (perhaps those who are more religious). The resulting data would underrepresent this important group, and thus would not truly speak to the population it claimed to reflect. In other, more severe instances, entire surveys may go uncompleted and unreturned by members of some significant group, whose data are then never collected. To

expand on the previous example, if highly religious individuals were less inclined to return the survey, complete the survey, or answer certain questions, then the data and results would be biased, because the responses of highly religious people would be less likely to be included, and again the results would claim to speak for a group that was not fairly represented in reality.

**Questionnaire bias** is error created by the questionnaire design. This is usually associated with a confusing layout of the survey and question order, poor wording, or survey content that otherwise confuses or misleads respondents. Questionnaire bias is more likely when the questionnaire is too long or cluttered, but it can occur in any project in which researchers do not take the time to plan, assess, and test their instrument in the various manners described in this text.

**Interviewer bias** occurs when characteristics of the interviewer (or just the interviewer's presence) can influence the way a respondent answers questions. This bias is also related to poorly trained interviewers (and their ability to conduct an interview, prompt for answers, and record open-ended responses).

## Strategies to Minimize Error

*Improving Nonresponse Bias:* Minimization of this type of error is based on assurance of confidentiality, compensation for participation, calling back or mailing reminders, properly training surveyors and interviewers, and having clearly defined and simple concepts in the survey.

*Questionnaire Design Bias:* Survey questions should be unambiguous, clear, and free of unnecessary technical language and jargon. They should also be short to avoid respondent fatigue.

*Interviewer Bias:* The researcher should ensure that interviewers are properly trained and not overworked, to avoid interviewer fatigue. In many cases, it is important to conduct field testing; to train interviewers in proper interviewing techniques, the topic of the survey, the research procedure and scope, and to manage the workload of interviewers carefully.

*Environmental Bias:* It is important to keep in mind the context in which the survey is administered. Asking respondents to remark on their current health during flu season might lead to a downward bias in responses, since more people are ill during this time.

All in all, measurement error is often difficult to recognize (and measure), but it can be minimized with proper sampling, up-to-date and precise measurement, properly trained interviewers, carefully scheduled survey times and intervals between surveys, and carefully and meticulously planned research design. Refer to the checklist below, which summarizes all of the guidelines presented in this chapter, to help reduce error in your own research project.



**CHECKLIST: SURVEY QUESTION CONSTRUCTION***Early Stages*

- Set concrete research goals.
- Clearly define your population.
- Address your own biases and limitations.
- Review existing survey instruments to inform the questionnaire.
- Decide on survey length.
- Construct simply written and well-structured questions.

*Development Stages**Questions*

- Are the questions short and worded simply?
- Are the questions specific and direct?
- Is only one question posed at a time?
- Is there any way to rephrase sensitive or private questions?
- Are specific time references included for questions that require recall?
  - Longer time periods are fine for important milestones.
  - Shorter time periods are preferred for items of low importance.

*Response Options*

- Given the social context, are these response options relevant?
- Is an "other" category necessary in order to have comprehensive categories?
- Are response options mutually exclusive?
- Are response options weighted appropriately with equal numbers of positive and negative responses?

*Overall Survey Design*

- Are sensitive questions in the middle?
- Are demographic questions at the end?
- Are all questions and response options absolutely necessary?

**KEY TERMS**

Variable 44  
 Single construct 45  
 Unidimensional concept 45

Simple concept 45  
 Multidimensional concept 46  
 Double-barreled question 49

Leading question	51	Vignette	69
Item nonresponse	53	Telescoping error	71
Bracketed categories	54	Recall loss	71
Closed-ended question	56	Prospective question	72
Multiple-choice question	57	Retrospective question	71
Fixed-choice question	56	Back translation	73
Dichotomous question	58	Measurement error	74
Checklist	58	Systematic error	74
Rating scale	58	Random error	74
Rank order scale	59	Nonresponse bias	74
Likert scale	59	Questionnaire bias	75
Forced-choice question	59	Interviewer bias	75
Semantic differential scales	60	Level of measurement	64
Relevant	60	Nominal	65
Comprehensive	61	Ordinal	65
Mutually exclusive	61	Interval	66
Open-ended question	67	Ratio	66

## CRITICAL THINKING QUESTIONS

Below you will find three questions that ask you to think critically about core concepts addressed in this chapter. Be sure you understand each one; if you don't, this is a good time to review the relevant sections of this chapter.

1. In addition to religiosity, what is another multidimensional concept used in the social sciences? Identify the different dimensions of this concept and how you might measure them in a survey.
2. How might a double-barreled question influence the measurement error of a survey?
3. Identify the rationale for and steps involved in back translation. How might this be an effective safeguard against measurement error in multilingual survey designs?
4. Discuss the ways you might measure (a) family size, (b) age, and (c) income at the nominal, ordinal, and interval/ratio level.