

Questionnaire Development

CHAPTER

9

What You Should Know!

Data provides the basis of research. Chapters 6 through 8 describe different types of research designs, which provides the structure of the research study and is directly related to data collection. One of the most popular means of gathering data is through a questionnaire. This chapter specifies steps in developing questionnaires for collecting data. After completing this chapter, the reader should be able to:

1. Define the term *questionnaire* and explain its association with the term *survey*.
2. Explain how a questionnaire could be used in both quantitative and qualitative research.
3. Describe what is involved in listing the items one is interested in knowing about the group, concept, or phenomenon.
4. Explain how to establish validity and reliability.
5. Discuss why the wording in the questionnaire must be appropriate for the target audience.
6. Explain why it is necessary to clearly identify who should answer the questions.
7. Discuss why one should avoid asking any questions that are biased, leading, or double-barreled in nature.
8. Explain why before construction, a decision must be made whether to use open- or close-ended questions or a combination of both.

9. Describe the possible impact of the respondents not having the general information needed to complete the questionnaire.
10. Describe the benefits of pretesting a questionnaire before it is used.
11. Discuss why one should set up questions so that the responses are easily recognizable whether the questionnaire is self-administered or conducted in an interview.
12. Explain why the questionnaire should be organized in a concise manner that keeps the interest of the respondent, encouraging him or her to complete the entire questionnaire.
13. Define the term *scales* and explain their purpose.
14. Compare and contrast the Thurstone Scales and Likert Scales.

Surveys and Questionnaires

Surveys are one of the common methods for collecting data in criminal justice and criminology. The word *survey* often creates confusion in research methods courses because it refers to both a research design and data collection device. The term *questionnaire* tends to be used interchangeably with the term *survey*. For the sake of clarity, *questionnaire* will be used within this textbook to refer to the survey data collection device. Questionnaires as a form of quantitative research design rely on the numerical assessment of data. This assessment consists of a series of questions with specific responses provided from which the subjects choose the one that is the most appropriate. In a survey study regarding the occupational stress of correctional officers, a question may be "How stressful is your job?" and the possible responses could be "not stressful," "neutral," or "stressful." Each of the possible responses would be assigned a number: 1 for "not stressful," 2 for "neutral," and 3 for "stressful." Typically, the lowest assigned number indicates the least (e.g., least amount of stress) and the highest assigned number indicates the most (e.g., most amount of stress). Setting up the questions in this logical manner will simplify the analysis and interpretation of the information collected. Questionnaires consisting of open-ended questions are also used within qualitative research, including structured and semi-structured interviews.

As discussed in Chapter 8, types of surveys include personal interviews, mail questionnaires, and telephone surveys. The data collection for each type of survey varies in form, but each is based on the use of a questionnaire. Within the personal interview, the interviewer and interviewee meet face-to-face, and the interviewer reads the questions aloud and records the responses. The interviewer may also record the observations about the behavior or demeanor of the subject, such as ranking the level of the subject's nervousness on a scale of 1 (least nervous) to 10 (most nervous). In the mail questionnaire, a hardcopy or electronic copy of the questionnaire is provided to the subject, who completes

FROM THE REAL WORLD

Dantzker (2010) created a study-specific questionnaire to examine whether there were any differences regarding what they used to conduct the evaluations and the reason for their choice between police and general psychologists who performed pre-employment psychological examinations of police recruits. Because of the information sought, no existing questionnaire could accomplish this goal and a new instrument was necessary. The questionnaire began with closed-ended questions answered through telephone interviews and was then developed into a self-administered questionnaire (Dantzker & McCoy, 2006), which eventually was revised into the final questionnaire used for the final study (Dantzer, 2010).

and returns it. The telephone survey is similar to the personal interview in that the interviewer reads the questions aloud and records the responses; however, the interviewer and interviewee are not face-to-face.

When conducting survey research, examine questionnaires that have been developed and tested previously to see if one of the questionnaires would be appropriate for your research. The primary reason for including a previously used questionnaire is that it reduces the likelihood of validity and reliability issues, two major concerns of questionnaire development. However, an instrument may not exist for a particular research question, or if it does exist, it may not meet the researcher's specific needs. Therefore, the researcher must resort to creating a research-specific questionnaire. Researchers may also use an existing instrument for the basis of their questionnaire by making adjustments where needed. In a study of the impact of violent victimization on post-traumatic stress, researchers may use an existing victimization survey as the basis for questions related to violent victimization.

In creating a new survey instrument there are several things to consider, including reliability, validity, and the levels of measurement to use. Every textbook offers a different way to approach this task, but in the end, the basic elements of questionnaire development are similar. To simplify this task, a set of guidelines are offered in the manner of rules.

Rules for Questionnaire Construction

Many people prefer not to follow rules, or at least to bend them to meet their satisfaction. The general rules presented here do not need to be followed in a strict manner but do provide a useful path to avoid a failed questionnaire. For many students, as well as researchers, getting started is the hardest part.

In Dantzker and Waters's (1999) study of students' perceptions of policing, the data for their sample was gathered using a questionnaire and included the information shown in **Figure 9-1**. As you can see in this figure, basic demographic information is collected, including gender, age, and race, and then questions about the student's year and major. In this case the major is used as a qualifying question. If subjects answered that their major is criminal justice, then they answered the questions regarding perceptions of the police. Subjects with other majors did not answer the questions.

Ford and Williams (1999) used a short-answer question approach in their survey instrument to collection information regarding the impact of taking a human diversity course (**Figure 9-2**). This short-answer survey can be compared to the previous From the Real World example. Both questionnaires provide an example of the mixed methods approach, discussed in Chapter 8, with the inclusion of open-ended questions.

1. When did you take the human diversity course? Year _____
2. How would you rate your understanding of other cultures before taking the course?
3. How would you rate your understanding of other cultures after taking the course?
4. Has there been a difference in your on-the-job behaviors (beliefs, the way you think about others, or your actions) since you took the course? Can you give an example? [Use the reverse side if you need more room.]
5. Has there been a difference in your off-duty behavior or your personal life since you took the course? Can you give an example? [Use the reverse side if you need more room.]
6. What do you believe is the single most important thing you learned from the course? [Use the reverse side if you need more room.]
7. What, if anything, was the worst thing about the course? [Use the reverse side if you need more room.]
8. Would you recommend/have you recommended the course to others?

Figure 9-2 Short-Answer Question Approach.

© Jones & Bartlett Learning.

Regardless of which approach is used, the key is to ask all the questions or create all the statements believed necessary to obtain the information desired. Listing the information that you want to collect from a sample seems easy, but it can be more difficult than it first appears. Often researchers run into problems because they fail to list everything they know or want to know about the subject. This failure could cause a shortage of important data because once

the information has been collected, researchers can rarely follow up with the subjects. The key is to decide what is of interest and what is needed and then base the questions on this information. One can always eliminate data, but it is difficult to go back and get what was missed. In sum, Rule One suggests making a list of information desired before creating any questions or statements.

Rule One: Start With a List of All the Items One Is Interested in Knowing About the Group, Concept, or Phenomenon

How many times have you gone grocery shopping without a list? When that happens, it is common to end up with many things not wanted and missing things that are really needed. Questionnaire development should be approached with a grocery list mentality. List all the things you want to know, things you would like to know, and things that would be interesting to know about each item about which information is being sought. Keep in mind that you are not committing to the items on this list, and some of the items will be excluded from the final version of the questionnaire. Think of this list as the "everything" (including the kitchen sink) list. Once you have the everything list, it will be much easier to remove and modify items rather than having to add items.

For example, the primary interest of Dantzker's (2010) study was to examine whether there were any differences between psychologists as to the measures they use to conduct pre-employment psychological screening as well as the reasons for their choices. To collect the relevant information, a questionnaire had to be developed that asked the respondents to provide certain identifying characteristics (the independent variables) along with their responses to questions about pre-employment screening (dependent variables). What was asked depended on the type of comparisons or analysis that was to be conducted. Because two groups were being compared, minimally required questions included collected information about the respondent and about the evaluation. Regardless of the nature of the sample, similar questions in various formats should be asked.

Rule Two: Be Prepared to Establish Validity and Reliability

If an existing questionnaire is not available for use within the study, then establishing validity and reliability for the instrument is critical. Validity and reliability were briefly discussed in Chapter 4. If the questionnaire is not considered to be both valid and reliable, then other researchers may not place much value in the results.

Validity refers to whether the questionnaire is in fact measuring what it claims to measure. It is imperative that the questions or statements be a true measure of the topic under study, which requires that the researcher be able to determine and establish the questionnaire's validity. There are different types of validity, including: face, content, construct, and criterion validity. Each has a specific purpose in establishing validity.

Face

The simplest means of establishing validity is face validity. Essentially, it is a judgment call by the researcher that the questionnaire is measuring what is being attempted to be measured. However, this type of validity lacks empirical support and requires the researcher to demonstrate why he or she believes it measures what is expected. Face validity decisions are often based on the researcher's knowledge, background, and observational experiences. Although this process is an acceptable means of validation, it is the least acceptable because it is empirically weak.

Content

The second form of validity, content validity, also suffers from being judgmental and usually non-empirical. Unlike face validity, where the belief in validity focuses on the questionnaire as a whole, content validity emphasizes each individual item's ability to measure the concept in question. Instead of simply supporting the complete questionnaire's ability to measure what is expected, the researcher must be able to explain why each item measures what is expected. Again, the responsibility falls to the researcher to support why it is believed that each item measures what is expected.

Construct

Although one is seeking to measure a particular phenomenon, there may be related concepts that are equally important to understanding the phenomenon in question. Construct validity seeks to demonstrate that the questions actually measure what they have been designated to measure by examining the relationship the measure has to other variables. The interest here is in establishing the fit between the theoretical and operational aspects of the item. With respect to the job satisfaction study, it was expected that individuals with a college degree would indicate a higher level of satisfaction with some items than individuals without a college degree. If the responses to those particular questions support this measure, then one has established construct validity. However, if the responses from both groups indicate equal satisfaction, then the construct validity may be challenged. Construct validity can be reinforced through empirical measures, unlike face or content validity.

Criterion

Criterion validity is concerned with the relationship between the questionnaire's results and an external criterion. The assumption is that if the questionnaire is valid, a specified empirical relationship should exist between the data collected and other existing measures of the phenomenon. There are two types of criterion validity: concurrent and predictive. The key requirement for concurrent criterion validity, however, is that a reliable and valid measure must already exist to make the comparison. To apply criterion validity to a job satisfaction measure, one

development

the job satisfaction measure, which is then compared to a validated job stress measure (because it has previously been established that stress is linked to job satisfaction). With the two sets of data, a correlation coefficient can be computed to provide what is called predictive validity, the more common form of criterion validity is predictive validity. The questionnaire's ability accurately to predict future behavior is what is being tested. Upon exiting prison, offenders may have to complete a questionnaire to determine the likelihood they will recidivate. After a period of time, the researcher may compare the results of the assessment and the actual recidivism. If offenders who were assessed as more likely to recidivate had a higher occurrence of recidivism, then predictive validity has been established.

There are several types over the best type of validity. Failing to establish any type of validity is the data. The more ways one can establish validity, the more confident one is not enough. It must be accompanied by reliability. A car battery that starts the vehicle day in and day out. A car battery purchased because of its reliability to do just that. When a car battery is no longer considered reliable. A questionnaire is consistent over time and yields similar results. To establish reliability further, one must establish consistency.

WORLD

Dantzker (2010) created a questionnaire specifically for his research. The questionnaire, face validity was addressed through the research subject, comparison to previous research efforts, and support with questionnaire construction experience and criteria for content validity. The questionnaire also met the criteria for content validity by relevant protocols identified by relevant researchers.

group of officers is given the job satisfaction measure, which is then compared to the results of a reliable and validated job stress measure (because it has previously been established that job stress is linked to job satisfaction). With the two sets of results, a correlation coefficient can be computed to provide what is called the *validity coefficient*. The more common form of criterion validity is predictive validity, which rests on the questionnaire's ability accurately to predict future conditions or responses. Upon exiting prison, offenders may have to complete an assessment designed to determine the likelihood they will recidivate. After three years, the prison may compare the results of the assessment and the actual occurrence of recidivism. If offenders who were assessed as more likely to recidivate actually had a higher occurrence of recidivism, then predictive criterion validity has been established.

There may be debates over the best type of validity. Failing to establish any type of validity devalues the data. The more ways one can establish validity, the better. Still, validity alone is not enough. It must be accompanied by reliability.

Everyone who lives in a climate where there are several days of extremely cold weather wants a car battery that starts the vehicle day in and day out. Perhaps the battery was purchased because of its reliability to do just that. When it does not start as expected, it is no longer considered reliable. A questionnaire has the same expectation—that it reliably does what it is designed to do every time it is used. If the questionnaire is consistent over time and yields similar results each time it is used, it is reliable. To establish reliability further, one must demonstrate stability and consistency.

FROM THE REAL WORLD

As previously noted, Dantzker (2010) created a questionnaire specifically for his research. For this questionnaire, face validity was addressed through the researcher's knowledge of the subject, comparison to previous research efforts, and support from neutral observers with questionnaire construction experience and knowledge of the subject. The questionnaire also met the criteria for content validity. The content of the questionnaire was primarily the identified protocols and reasons for choosing them, both of which have been established by relevant literature. The questionnaire was reviewed by three licensed psychologists who all indicated they found that the questionnaire does reflect the intended objectives. Finally, for criterion and construct validity, the questionnaire had to cover the full range of possibilities within the concept and required a preexisting questionnaire with which to compare or a concept that cannot be directly observed or isolated. Because the concepts measured were based on actual use and established reasons for its use, the questionnaire was found to measure the primary concept.

Stability
the same an
to expect
Quotient)
be stable.
would be
is determin
predictabl
retest (pre

Prete
inconven
distribut
are the s
question
samples

A p
questio
same gr
suppon
naire, t
metho

Al
the tes
statist
of Cr
numl

Rule
for t

In W

to c

qui

Wh

gea

wi

mi

fo

us

ca

m

th

Stability occurs when, under similar conditions, a respondent provides the same answers to the same questions on a second testing. It is reasonable to expect that individuals would have similar results on IQ (Intelligence Quotient) tests two weeks apart. However, not all measures are expected to be stable. The expectation is that questionnaires designed to measure fear would be noticeably different before and after a terrorist attack. Consistency is determined when the set of questions is strongly related and provides predictable results. There are three standard ways to test reliability: (1) test-retest (pretesting); (2) split-half technique; and (3) using multiple forms.

Pretesting is the best method for testing reliability, yet perhaps the most inconvenient in terms of time and money. The test-retest method requires distributing the questionnaire to the same population twice. If the results are the same, then reliability is accepted. Another method is to distribute the questionnaire to similar samples and look for consistent results between the samples.

A popular and widely used method is the split-half technique. Here the questionnaire is divided into sections or halves. Both sections are given to the same group or among similar groups. A similarity of scores between both halves supports reliability. Using several variations or formats of the same questionnaire, the multiple forms method can support reliability. As with the previous methods, if scores on each format are similar, one can assume reliability.

All these methods are acceptable; however, when possible, one should use the test-retest method. In addition, many statistical packages offer methods for statistical comparison by item-to-item and item-to-scale analyses and the use of Cronbach alpha, a commonly used reliability coefficient. Overall, there are a number of ways to establish validity and reliability, and it is important to do so.

Rule Three: Word the Questionnaire Appropriately for the Target Audience

In Western democracies it sometimes feels as if we are constantly being required to complete questionnaires of one type or another. Sometimes the questions are quite clear. However, at other times the questions may befuddle respondents. When developing a questionnaire, the first guideline is to be sure to use language geared toward the target population. One should not use words or phrases with which the respondent is not familiar because it can cause confusion and misunderstanding, which may lead to tainted data. For example, on a questionnaire for college students on criminal behavior, a question about drug use should use the common usage (or slang) rather than scientific: Have you ever smoked cannabis? (wrong); Have you ever smoked marijuana? (right). Also, keep in mind the reading level of your audience. If your audience is the general public, then writing questions at a graduate school level could create problems because some members of the audience would not understand the questions. Pretesting

the questionnaire within a similar audience to your target audience can help identify issues. It is important that the questions or statements be written in a manner that the target audience can understand.

Rule Four: Clearly Identify Who Should Answer the Questions

You have probably received a questionnaire in the mail that has been addressed to Dear Occupant or Dear Current Resident. On opening it, you may have discovered that it is not clear who should be completing this questionnaire. For his questionnaire, Dantzker (2010) included this question and response instructions to clarify who should complete the questionnaire:

Do you personally provide psychological screening services to any type of police/law enforcement agency?

Yes ☐ IF YES, please respond to the remainder of survey.

No ☐ IF NO, you may stop here. Thank you.

All it takes is a simple statement advising who should complete the questionnaire. If the questionnaire requires that multiple subjects complete different questions within the same survey, then the directions should clearly identify who should be completing which questions. For example, the first data collection from the panel study National Longitudinal Survey of Youth (1997) used a questionnaire to acquire information from both youth and their parents. It is important to clearly specify who fills out or responds to which questions.

Rule Five: Avoid Asking Questions That Are Biased, Leading, or Double-Barreled in Nature

Questions must be well worded. Questionnaires should not contain questions designed to elicit a specific response. "Does it not feel great to get high?" and "Do you cheat on exams and if you do, do you know you are only cheating yourself?" are examples of biased and leading questions because they create a push toward a specific response and convey the researcher's bias. Questionnaires should also not include double-barreled questions, which are questions that actually contain more than one questions. For example, "How often do you get high and do you enjoy it?" is a double-barreled question. The simple fix to a double-barreled question is to split it into two different questions.

Questions or statements that confuse respondents can cause ambiguous responses, and questions that seek to guide the respondent can create blatantly false responses. In addition, the structure of the questionnaire may be such that preceding questions influence the responses to later questions. If a question regarding recent incidents of terrorism precedes questions regarding the respondent's fear of terrorism within a questionnaire, then it is possible (or even probable) to assume that presence of the previous terrorism question

would increase fear of terrorism. If any of these issues with question wording, structure, or instructions occurs, then the validity of the findings is decreased. Ultimately, the wording of questions and the questionnaire format can influence responses. Be particularly aware of these issues when responding to surveys conducted by ideological groups or organizations that hold particular positions on controversial issues, such as the opposing views of gun control advocates and the National Rifle Association. Potential biases may also be observed among the questionnaires of scholars and students who are trying to support a favored hypothesis. If those administering a survey have an interest in or might benefit from the outcome of the survey, they may word the questions or structure the format so as to enhance the likelihood of desired responses. Though this issue may be unintentional, be skeptical of study findings by such individuals or groups.

Rule Six: Before Constructing a Questionnaire, Decide Whether to Use Open- or Closed-Ended Questions or a Combination of Both

Because the goal of the questionnaire is to acquire specific information related to the topic for ready analysis, deciding what type of questions should be asked is crucial. Open-ended questions can make data analyses somewhat more difficult but can provide more in-depth responses. On the other hand, well-constructed closed-ended questions can provide sufficient data that are easier to analyze. An increasingly popular method is to combine the types of questions to collect the most pertinent information about the topic, which is an example of mixed methods research discussed on Chapter 8.

The questionnaire in Figure 9-3 was created as a telephone survey to investigate community satisfaction with a police department. Although this questionnaire had been created as part of a proposal to evaluate a police agency, it was never tested or published elsewhere. Observe how the questionnaire is composed of both closed- and open-ended questions. Notice how the open-ended questions are specifically worded so that responses could be more readily coded for statistical analysis. For example, the inclusion of "the main problem", "the best thing", "the worst thing", or "one change" within the question are designed to elicit very specific opinions and do not ask for an explanation for the opinion, which would be much more difficult to code.

Rule Seven: Keep in Mind That Respondents May Not Have All the General Information Needed to Complete the Questionnaire

Under any circumstances, making assumptions could be problematic, but this is a special risk in questionnaire development. It is a common error to believe that possible respondents have all the information needed to respond to the questionnaire. For example, several questions about drug and alcohol programs

POLICE COMMUNITY TELEPHONE SURVEY

4. In your opinion, what is the BEST thing about the XXXXX Police?
5. In your opinion, what is the WORST thing about the XXXXX Police?
6. In your opinion, what is one change you would make that could improve the XXXXX Police?

For Analysis Purposes Only:Gender ☐ Male ☐ FemaleRace/Ethnicity ☐ Caucasian ☐ African-American ☐ Hispanic ☐ Asian ☐ Native American
☐ OtherAge ☐ 18-29 ☐ 30-49 ☐ 50-65 ☐ Over 65Employment ☐ Professional (i.e., doctor, lawyer, etc.) ☐ Education ☐ Retail Business
☐ Food Service ☐ Clerical ☐ Laborer ☐ OtherEducation ☐ High School ☐ Some College ☐ Two-year Degree ☐ Four-year Degree
☐ Graduate Degree

In what area of the city do you live?

☐ North ☐ South ☐ East ☐ West ☐ Central**THANK YOU FOR YOUR ASSISTANCE.**

Figure 9-3 Questionnaire: Open- and Closed-Ended Questions. (continued)

on campus might be asked in a Student Criminality study simply because it is assumed that all students are familiar with these programs. This assumption might result in few responses because respondents do not have the necessary information and choose not to respond. The responses that are received could be biased because only students who are familiar with these programs may have responded. Or students may have responded based on their own assumptions about the programs rather than firsthand knowledge. To avoid this dilemma, always provide an "escape" response, such as "unknown," "no opinion," or "unable to respond" (refer to Figure 9-3).

Rule Eight: Whenever Possible, Pretest the Questionnaire Before It Is Officially Used

Although this is a time-consuming step, pretesting is undoubtedly important and should be used whenever possible. For example, with the Student Criminality questionnaire, after completing the first draft of the questionnaire, researchers may have students in their classes complete the questionnaire. Pretesting can uncover errors in construction, language, or other problems that may cause the data to

Rule Nine: Set Up Questions So That the Responses Are Easily Recognizable Whether the Questionnaire Is Self-Administered or an Completed in an Interview

The fastest way to jeopardize research is through a questionnaire in which the directions are not clear on how to respond. Be sure to provide adequate, clear instructions and establish recognizable means for responding. Also, try not to make the format too busy, such as the overuse of decorative designs. Avoid using small print because it is hard to read and can be confusing if too many questions are squeezed on a single page.

Rule Ten: Organize the Questionnaire to Keep the Respondents' Interest, Encouraging Them to Complete the Entire Questionnaire

How often have you started a novel only to give up after the first few chapters because you found it boring? If you had continued to read, it might have become more pleasurable, but you lost interest. This same concept is applicable to questionnaire development. If the beginning questions are not interesting and do not hold the respondents' attention, chances are that they may not complete the rest of the questions, which results in missing data. Therefore, it is beneficial to create questions that may pique respondents' interest at the beginning and the end. In addition, if respondents see several pages of questions, they are less likely to begin the survey. Although it is tempting to try to cover everything, a clear and concise survey that consists of a few easy-to-read questions typically receives more responses than a lengthy questionnaire. Although specialized questionnaires that target a specific population (who may have a vested interest in the subject matter) may be longer, it is recommend that most surveys be kept to two pages of questions using normal-size type.

Rules and guidelines are fallible, but the 10 rules offered for questionnaire construction, if followed, improve the chances of obtaining quality, analyzable data. Still, keeping all the rules in mind might not prevent the creation of a poor questionnaire. Beyond these questions, a key aspect in questionnaire construction within a quantitative research design is development of questions that use scales.

Scales

A common element of quantitative survey research is the use of scales. A scale can either be a measurement device for responding to a question or statement or a compilation of statements or questions used to represent the concept studied. For example, in the questionnaire in Figure 9-3, responses to several statements regarding police performance are on a scale that ranges from 1 (wholly unsatisfactory) to 4 (completely satisfactory). Each statement represents a separate variable of the respondent's perception. Putting the statements from

each perceptual question together (attaining a numerical result for responses to all statements) gives the researcher what might be referred to as the "Community Members' Perceptions of Their Local Police Agency Scale." Scales as compilations are particularly important and a relevant part of research for three primary reasons (Creswell, 2008; Dunn, 2009; Frankfort-Nachmias & Leon-Guerrero, 2008): (1) to allow the collapsing of several variables into a single variable producing a representative value and reducing data complexity; (2) to offer measures that are quantifiable and more open to precision and statistical manipulation; and (3) to increase the measurement's reliability.

To accomplish these ends, a scale must fit the Principle of Unidimensionality (Frankfort-Nachmias & Leon-Guerrero, 2008). According to this principle, the items making up the scale need to represent one dimension befitting a continuum that is supposed to be reflective of only that specific concept. For example, if one is measuring job satisfaction, the scale should not be capable of also measuring job stress.

Scaling Procedures

Ultimately, to conduct research one looks to complete a measurement. But what is the actual purpose of this measurement? Measurement is used in research as a means of connecting phenomena with numbers for analytical purposes. Scaling is identified as a means of assisting in making these necessary and proper connections. Despite the existence of numerous scales, there are times when researchers must create their own. The key is to keep in mind that the goal is to explain a phenomenon. Thus, the scale must meet this goal. There are two primary types: arbitrary scales and attitudinal scales.

Arbitrary Scales

An arbitrary scale is designed to measure what the researcher believes it is designed to measure and is based on face validity and the professional judgment of the researcher. Although this process allows for the creation of many different scales, it is easily criticized for its lack of substantive support. Still, this type of scale does provide a viable starting point for exploratory research, even though it is the less recommended method of scaling.

Attitudinal Scales

More often employed in criminal justice and criminological research than arbitrary scales are two common types of attitudinal scales: Thurstone and Likert.

Thurstone Scales

The construction of a Thurstone scale relies on the use of other individuals (sometimes referred to as "judges") to indicate what items they think best fit

the concept. There are two methods for completing this task. The first method is *paired comparisons*. Here, the judges are provided several pairs of questions or statements and asked to choose which most favorably fit the concept under study. The questions or statements picked most often by the judges become part of or comprise the complete questionnaire. For example, the judges are asked which of the following questions might best fit the concept of job satisfaction: "I enjoy going to work every day." "Sometimes I am very tired when I get home from work."

The second and more frequently used procedure is referred to as *equal-appearing intervals*. For this method, the researcher submits a list of questions or statements to the judges who are then asked to give each a number, indicating the strength of the question or statement to the concept. The researcher then keeps those items on which judges were in the strongest agreement and eliminates those with the weakest indicator scores. For example, the researcher decides to design a questionnaire to examine criminality among college students and calls it the "College Students' Criminality Questionnaire." A decision is made to have the questions form a 15-point item criminality scale. Fifty questions are submitted to judges, who are instructed to score each question from 1 (strongest indicator) to 15 (weakest indicator). The top 15 questions become the scale.

Thurstone scaling is not very popular because of the time it takes for the judges to complete their tasks. Furthermore, because the judges must be experts in the area of study, finding an adequate pool of qualified judges could be problematic for the researcher.

Likert Scales

Probably the most commonly used method in attitudinal research is the Likert scale. This method generally makes use of a bipolar, five-point response range (i.e., strongly agree to strongly disagree). Questions where all respondents provide similar responses are usually eliminated during the data analysis portion of study, which will be discussed in Chapter 11. The remaining questions are used to comprise the questionnaire. Figure 9-5 provides an example of Likert-type questions. One of the benefits of the Likert scale is that it provides a neutral response, such as the "Doesn't Really Matter" option located in the middle of the scale. Another benefit is the balance it provides between collecting detailed information and not being so intensive that respondents do not finish answering the questions.

There are various other types of scaling procedures. However, because so few are used in criminal justice and criminological research, they are not discussed here. Furthermore, advanced statistical techniques, such as factor analysis and Chronbach alpha, are much faster and simpler to use to determine a scale's composition. For example, the Student's Perceptions of Police scale