

13

MEASURING YOUR STUDY VARIABLES

Allison conducts a support group for caregivers of elderly people in an agency where she is a social work intern. The objective of this intervention is to reduce caregiver burnout, because this phenomenon reduces the effectiveness of caregivers in the care of elderly clients. If the caregiving service does not work for an individual, a less appealing alternative will often be arranged. Allison noticed that the agency had not used any method of evaluating the outcome of this service, so she took on this task. She realized that she needed a good definition of the concept being measured before she sought a tool for this task. Burnout was defined as a state of physical, emotional, and mental exhaustion that may be accompanied by a change in attitude from positive and caring to negative and unconcerned. Allison sought a tool for measuring caregiver burden and found the Zarit Burden Interview Screen. After reviewing this tool, Allison found it to be consistent with her definition of caregiver burnout. She also discovered that this tool had passed several tasks of quality through the examination of reliability and validity. With this information, she concluded that she had the tool she needed for measuring outcomes for the caregiver support group. She suggested that this tool be given to the clients of the caregiver support group prior to beginning the support group experience and again after 6 months of attending the group.

INTRODUCTION

In a nutshell, the task facing Allison is the task that is addressed in this chapter: the measurement of study variables. Allison was conducting research on the outcomes of an intervention with a group of clients. This means that she was engaged in an evaluation study, rather than a descriptive, explanatory, or exploratory study. You might engage in an evaluative study where you are giving a group of clients a service, or you might evaluate the services for a single client. This chapter presents information on measurement that can be used both for groups and for single clients. However, there are some special challenges when you are dealing with a single client. Asking a single client to respond to an evaluation tool at the beginning of each counseling session is a different challenge from asking a group of clients to complete the questionnaire once before treatment begins and once again at the end of treatment. The tool for the single client should be less intrusive into the service experience, so it should be simple and not very time-consuming.

Allison's study was an evaluative study. Your study could be evaluative, but it could also be descriptive or explanatory. You might want to be able to describe the clients of your agency or examine the causes of client no-shows. In all types of studies, measurement is a necessity.

Measurement is one of the key elements of the study of methodology, two of the others being sampling and research design. Two major alternatives for measuring study variables will be addressed in this chapter: (1) finding a published measurement tool and (2) designing your own measurement tool. In this chapter, you will learn how to do the following:

1. Explain the nature of measurement, with emphasis on accuracy in measuring the concepts as they are defined
2. Describe the mechanisms for achieving measurement accuracy, with a focus on reducing measurement error by certain tests of the reliability and validity of measurement tools
3. Demonstrate how to find a measurement tool, with emphasis on tools for measuring client outcome in evaluative research
4. Construct a measurement tool when there is no published tool that will work for the measurement of the variable that has been defined

This knowledge will support your ability to apply what you have learned to the task of finding, or developing, a measurement tool for a study of interest to yourself. At the end of this chapter, there will be a practice exercise where you can demonstrate your skill.

WHY FRET OVER MEASUREMENT?

This chapter will help you determine how to measure each variable in your study. To accomplish this objective, the chapter will include a great number of concepts and checklists to ensure that your measurement is optimal. It will also focus on

objectivity as a core principle, and it will encourage you to heighten the accuracy of the method you employ to measure variables.

For example, if you wanted to understand the level of the effectiveness of your alcoholism treatment program, you could ask the social workers working in this program to give their opinions about whether it is effective in reducing the problems associated with alcoholism. The social workers could either respond to a questionnaire in which they answer “Yes” or “No” when asked if the program is effective, or they could rate the program on a scale of 1 to 4, where 1 = *highly effective*, 2 = *mostly effective*, 3 = *somewhat effective*, and 4 = *not effective*.

However, this method would be highly vulnerable to subjectivity. The social workers who serve in this program would be expected to have a positive bias, as they do not want to engage in behaviors likely to lead to the elimination of a program that provides them their jobs.

Instead of asking the social workers, why not ask the clients to answer a very broad question, like the above example? This would be a step in the right direction because the clients are not protecting their jobs by being positive. However, this broad characterization of effectiveness would be vulnerable to the **social desirability bias**, whereby people are encouraged to give a socially desirable response; most of us have been encouraged throughout our lives to be positive.

If clients are asked about effectiveness, the nature of their responses should be broken down into categories of life where they are more likely to have differentiated responses and, thereby, be more objective and truthful. For example, clients could be asked about their interpersonal lives, their work lives, their home lives, and so forth. More specifically, they could be asked a series of questions designed to measure the extent to which they are depressed or suffering from the symptoms of posttraumatic stress disorder.

We could ask clients to give a statement about how well the treatment program has helped them. This could be very useful, especially in the improvement of services. However, a set of statements about client satisfaction would not be as useful as concrete data on, for example, the clients’ scores on depression. The latter would be less vulnerable to the social desirability bias.

So effective measurement helps us get an objective picture of our research theme and avoid measuring a person’s biases. Also, it helps us achieve credibility with others in the reporting of our results. If we reveal objective data, we are more likely to be heard.

THE NATURE OF MEASUREMENT IN SOCIAL WORK RESEARCH

Sometimes the social worker will have an easy measure of the study variable. This could include grades in school, the number of times a client fails to show up for an appointment, or whether a homeless person found a home. These researchers are lucky—they have a measure of outcome that gives little reason for concern.

But what if you need to measure something less concrete, like self-esteem, caregiver burden, or perceptions of social support? Now you have a challenge. You will need to define your variable carefully and seek a published tool for its measurement. If you cannot find one, you will need to construct one yourself.

In the previous chapters, you have witnessed the measurement of quite a few variables that were defined before they were measured. These variables included (a) the traits that describe the good work manager, (b) sex role stereotype about the characteristics of the good manager, (c) stress, (d) social support, and (e) depression. Therefore, you have had some experience with measurement. Now it is time for you seek more depth of understanding about this aspect of research.

In measurement, your critical issue is accuracy. Readers of your study results about marital satisfaction might have problems with your findings if your measure of marital satisfaction contained no items on extended-family relationships. They may believe that your measurement tool failed to be adequately comprehensive. Inaccurate measurement could also be a result of a tool that is difficult to understand; if your subject does not know what a question means, he or she is not going to answer the question accurately. Measurement accuracy will be discussed in this chapter by an examination of both reliability and validity, two means of reducing measurement error.

You have already seen the definitions of reliability (consistency) and validity (accuracy) in the previous chapters. Your foray into this subject started with the word *credibility*, which is used in everyday language and represents a general category into which both reliability and validity fall. Are your tools credible to the readers of your research report? If not, they will lack confidence in your results.

In a previous chapter, you saw descriptions of three levels of measurement: (1) nominal, (2) ordinal, and (3) interval. You need to know the level of measurement for your variables in order to select an appropriate statistic for finding the answer to your research question. For example, the various forms of the *t* test require that variables be measured at the interval level.

The lowest level of measurement is the nominal level. If your measure of a variable places study subjects into categories that have no order, you will have a nominal variable. An example of a nominal variable would be gender—males are neither higher nor lower than females. Another example would be political party affiliation. The second level of measurement is the ordinal level. If your measure of a variable puts people into categories that have an order, you have a variable measured at the ordinal level. An example of an ordinal-level variable would be an opinion statement with the response categories of *strongly agree*, *agree*, *disagree*, and *strongly disagree*. In this situation, there is an order among the possible responses, which go from lowest to highest. The highest level of measurement is the interval level. If you have measured your variable at the interval level, you have given it a score on a scale or some other measure that is numerical in nature, such as age measured in years (not categories of years).

These levels of measurement form a hierarchy. Nominal measurement is the lowest level, followed by ordinal, which is followed by interval. There is a fourth level of measurement, which is ratio. This level has not been given attention in this book because it has lesser importance for the social work researcher than the other levels of measurement.

STANDARDIZED TOOLS AND INDIVIDUALIZED TOOLS

Some measurement tools are standardized, and some are individualized. A standardized scale has a designated set of sentences and words that are used in the same way for each person who completes them. Examples include the Beck Depression Inventory and the Hare Self-Esteem Scale. You are more likely to see standardized tools used in published studies than individualized tools. You will see this form in examples of scales that are used in this chapter. The words on these scales are the same for each person who completes it.

However, there are times when you cannot find a standardized scale that measures what you want. This is when you seek, or develop, an individualized scale. You will then need to tailor the individualized scale for your individual client, so that the client will have his or her own unique scale. Some parts of the structure of your individualized scale may be common for more than one client, but the individual items for measurement are unique for each of them. For example, you will see a discussion of the Your Evaluation of Service (YES) scale in a later section of this chapter. When you employ this scale, you will ask the clients to describe, in their own words, the outcome objectives they want to achieve and to rate how well things have been going in the past week regarding each objective. One client may have the objective of “not shouting at my husband when he comes home late,” while another might want to “stay calm when I discuss my son’s school grades.”

While each client will have unique objectives, the general form of the YES scale is the same for all clients. Each scale will have a 10-point rating system with 1 = *miserably* and 10 = *extremely well*. You will ask your client to rate each objective on this 10-point scale each time you seek an evaluation. For the single-client study, this may be done each week when you see the client for a treatment session.

MEASUREMENT ERROR

When you attempt to use human subjects to measure concepts such as anxiety, depression, and feelings of support, you should understand the concept of **measurement error**, which is the distance between the data you have and the truth about the concept you are trying to measure. Some tools for measuring depression will have less measurement error than others because they have been more carefully constructed and subjected to tests for reliability and validity. Measurement error tells you that your measurement tool is not perfect, but it does not have to be perfect; it simply has to be a credible way to measure your variable for your purposes. If you wish to publish your study results, you will pay careful attention to issues such as reliability and validity. However, if your study is designed to help you with decisions about client service, you will still be attentive to whether your measurement is credible but you may not feel the need to delve as deeply into the nuances of reliability and validity in the pursuit of perfection. Perhaps the pursuit of perfection is not cost-effective in this situation.

You have probably taken an exam in a class and been frustrated with your grade. On review of the items you answered incorrectly, you likely saw a question that you misread, leading you to make an error in your test even though you knew the concept that was

being measured by that question. We can call this situation an example of measurement error. You knew more than your grade for this item on the test indicated. On the other hand, if you were very lucky with guessing the correct answer to questions you did not understand, this would also be a measurement error. In this situation, you would be the beneficiary of the error rather than the victim. Both are measurement errors because the grade for the test was not a perfect illustration of exactly what you knew about the subject of the exam.

Keep in mind that you will have some measurement error whenever you attempt to measure human variables. The question is not whether you have measurement error but whether it is an acceptable level of error. If the amount of measurement error is not acceptable, you have an inadequate tool for measuring your concept. If there were many examples of error in the research test you had in college, we might call this a bad test; it does not do a good job of measuring knowledge about the subject under study. There are various ways the professor can test for measurement error with regard to exams.

Sources of Measurement Error

A source of measurement error is the social desirability bias, which is a tendency to answer questions in accordance with what we believe is the socially desirable way rather than what we really believe. To avoid this bias, you should refrain from starting a questionnaire item with the words “Don’t you believe that . . .” This phrase is suggesting what the subject should believe.

The social desirability bias is an example of systematic error in measurement. **Systematic error** occurs when the tool we use reflects inaccurate information in a consistent fashion. If a tool has a high potential for eliciting a social desirability bias, you will find a consistent pattern of socially desirable responses from research subjects. The wording of questions for a survey can encourage a biased response, as in the example above. Information given later in this chapter on constructing a measurement tool will help you avoid this mistake.

Another type of measurement error is known as **random error**. This type of error has no consistent pattern as in the case of systematic error. With systematic error, you might predict that your study’s subjects will answer a given question in a given way because it is worded in a way that elicits a socially desirable response. With random error, however, you cannot predict the direction of the error. If, for example, you administered a questionnaire designed for adults to a group of third-grade students, you would likely get random errors; they would not understand the meaning of the questions, so they would just mark the tool in a random fashion.

Preventing Measurement Error

Among the methods for preventing measurement error are (a) preparing instruments that have items that are appropriately worded for the designated population and (b) using more than one method for measuring the same thing. The example of a tool for adults given to children illustrates the first method. You can employ the second error prevention method by using personal interviews of study subjects to measure your variable and comparing this result with that of a measurement tool designed to measure the same concept. If the two results are consistent, you have evidence of validity.

Two mechanisms for preventing measurement error are tests of reliability and validity. **Reliability** refers to the consistency of a measurement device, whereas **validity** refers to the accuracy of a measurement tool. A tool must be consistent to be accurate, but a consistent tool might not necessarily be accurate. In other words, a tool can be consistently inaccurate.

Reliability as One Means of Reducing Measurement Error

One method of testing for the reliability of a scale is known as **test–retest reliability**. Suppose the members of your research class took the Beck Depression Inventory today. Suppose further that the same students had taken this scale a week ago. Would you expect to see a positive correlation between their depression scores from a week ago and the scores of today? A positive correlation would mean that a person who scored higher than someone else on the first administration of the scale would tend to score higher than the other person on the second administration of it as well. A positive correlation between the two administrations of the scale would be evidence that this scale is consistent. We would then assume that someone who is clinically depressed today would have likely been clinically depressed a week ago and that someone who is not depressed today would likely not have been depressed a week ago. In that case, the clinically depressed person would score higher on the first administration of the scale, and this pattern would be repeated with a second administration of this scale, illustrating a positive correlation. So the answer to the above question is yes—you would expect to find a positive correlation between the two administrations of this scale. If you fail to get this result, you would conclude that you do not have evidence of the reliability of this scale. If a tool fails to be consistent, it is not reliable.

Another method of testing for reliability is testing the **internal consistency reliability** of the tool. This test examines whether the different items on the scale seem to be measuring the same thing. If the various items on the scale are measuring the same thing, different parts of the scale would have a positive correlation with one another. To test for internal consistency, you could compose a variable that contained only the even-numbered items on your scale and a second variable that contained only the odd-numbered items on the scale. You would expect a rather high correlation—.70 or higher—because each half of the scale is supposed to be measuring the same thing. If you failed to find a strong positive correlation, you would have reason to doubt the reliability of your tool and would need to reexamine whether the different items on this scale are measuring the same thing. A common tool for testing for internal consistency is Cronbach's alpha, the subject of Exhibit 13.1.

Validity as a Mechanism for Reducing Measurement Error

Validity refers to accuracy. Does the measurement tool accurately measure the concept as you defined it? To answer this question, you will need to revisit your definition of the variable when you engage in the examination of validity.

The weakest form of validity is **face validity**. This refers to whether the tool seems to be valid on the face of it, or on a surface level. In other words, if you give a tool to a group of knowledgeable people and ask them what it measures, they would consistently

EXHIBIT 13.1

USING CRONBACH'S ALPHA COEFFICIENT TO TEST FOR INTERNAL CONSISTENCY OF A SCALE

People who design published scales for the measurement of social variables often subject their scales to a test of internal consistency using a statistical measure known as **Cronbach's alpha**. This coefficient represents the analysis of all correlations of scale items with all the other scale items. The alpha coefficient combines them in a way that shows the strength of these combinations. In other words, Cronbach's alpha shows how well the various items on the scale seem to be measuring the same thing.

One of the scales for measuring social support has items such as the following: (a) "There are several people who I trust to help solve my problems" and (b) "If I needed help fixing an appliance or repairing my car, there is someone who would help me." The respondents are asked to select one of the following answers for each item on the scale: *definitely true* (3 points), *probably true* (2 points), *probably false* (1 point), or *definitely false* (0 points). Scores for social support are computed by summing the scores for each item on the scale. If you subjected this scale to a test of internal consistency, you would expect that scores for Item 1 on the scale would have a positive correlation with scores for Item 2 on the scale, Item 3 on the scale, and so forth. This pattern would suggest that the various items on the scale were measuring the same thing. If you failed to find this pattern, you would have reason to believe that the various items on the scale were not measuring the same thing,

so you would conclude that internal consistency was lacking. In other words, your analysis failed to show that the scale was reliable, which means that you lack evidence of its validity.

In the computation of the alpha coefficient, the correlation of Item 1 on the scale with Item 2 is computed. Then, the correlation of Item 1 with Item 3 is computed, and each possible correlation with each of the other items is computed. All these correlations are combined using a formula that determines the overall correlation level for the scale.

Coefficient alpha is one of the many options available in SPSS (Statistical Package for the Social Sciences), a statistical software. You enter the data for each item on the scale, treating each item on the scale as a separate variable. Then, you identify the individual variables to be included in the analysis of coefficient alpha, and SPSS will give you the value of alpha. Alpha values, just like correlation coefficients, can range from a low of 0 to a high of 1.0. A value of .70 or higher is considered sufficient as a determinant of internal consistency for a scale. If your coefficient is lower than .70, you can consider your tool to be lower than the normal standard, but this is a matter of opinion. A value of .50 also shows internal consistency, but it does not meet the accepted standard set forth by statisticians. A negative correlation, of course, would indicate that there is something seriously wrong with this scale.

refer to the concepts that you are attempting to measure. To test for face validity, you need a group of knowledgeable people and a procedure for asking their opinions on what the tool measures. It is best that you not tell them the name of the concept you are measuring, so you can avoid the social desirability bias—if they are your friends, they will want to please you. Instead, ask an open question about what key concept is measured by this tool. Report your findings, and the reader can decide if you have a credible tool.

A stronger type of validity is **riterion validity**. This refers to a test of whether the tool in question operates in the same way as another method of measuring the same concept. Suppose the members of your research class had been interviewed by a clinical social worker to determine if anyone in the class should be guided to treatment for depression. The entire class volunteered to be interviewed. The clinical social worker rated each person on a scale from 1 to 4 regarding the level of depression that was displayed in the opinion of the social worker. Level 1 indicated no depression at all, while Level 2 indicated some minor level of depression but not one that required treatment. Levels 3 and 4 were levels of depression that indicated that treatment would help, with Level 4 being the highest level of depression.

Now let us get to the issue of validity. A way to test the validity of the depression scale administered to this class would be to compute a correlation between the score for depression given by the scale and the rating of depression given by the social worker.

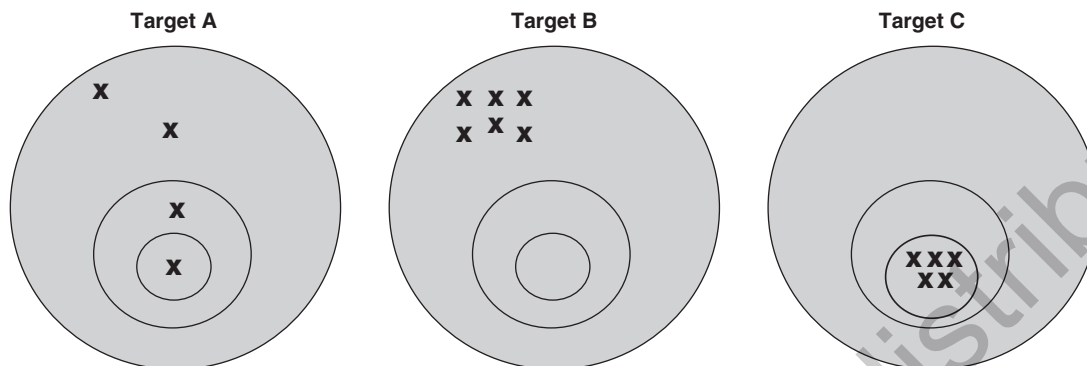
What should we find? Like the previous example, we would expect a positive correlation between these two variables: (1) the depression scale score and (2) the depression rating by the clinical social worker. This means that if John scored higher than Jane on the depression scale, he would be expected to also have been rated higher for depression by the clinical social worker. If this pattern among the participants holds up, you have reason to believe that your measurement tool has criterion validity.

Content validity is the final form of validity we will discuss. **Content validity** refers to the extent to which a measure includes all the dimensions of the variable being measured. For example, what are the dimensions of marital satisfaction? Would it include communication? How about agreement about child care practices? Would it include sexual satisfaction? What about satisfaction with finances or the amount of time the couple spends together? If you believe that marital satisfaction includes all of these dimensions, you would seek a marital satisfaction scale that included them all. If you found one that did not include finances or another key variable, you would continue your search.

There are other forms of validity and reliability that will likely be employed by those who design measurement tools. However, the task of this book is not to prepare you for the job of designing measurement tools for publication but, instead, to prepare you to construct your own tool and possibly employ some of the simple means of testing it for reliability and validity.

Can You Have Validity If You Do Not Have Reliability?

Can your measurement tool be reliable if it is not valid? Can it be valid if it is not reliable? Think about what each of these concepts means before you turn your attention to Figure 13.1. Reliability means consistency. Validity means accuracy. Can a tool be consistently inaccurate? Can it be accurate if it is not consistent? After considering these questions, take a look at Figure 13.1. This figure shows three targets on which you are testing three guns to see if they are reliable and valid. Target A has gunshots all over the target. Is this gun reliable? Is it valid? It appears to be neither. What about Target B? Is it reliable? All the shots landed in a very similar spot, so we would conclude that this gun is reliable. Is it valid? Is it failing to hit the bullseye? It has consistently failed to hit the bullseye, so you would conclude that it is not valid.

FIGURE 13.1 ■ Reliability and Validity in Measurement

Now, take a look at Target C. All the shots are in the center circle, where we were aiming our gun. We would conclude that this gun is not only reliable but also valid. These three figures illustrate that we can have reliability without validity, but we cannot have validity without reliability.

RELIABILITY AND VALIDITY IN QUALITATIVE MEASUREMENT

The previous sections were focused on quantitative measurement, where you measure variables by giving people a score or placing them into categories. The issues of reliability and validity are also relevant to qualitative measurement, where your data are words. When you conduct qualitative research, you will still be concerned with the consistency and accuracy of your measurement.

In qualitative research, reliability is the extent to which a set of meanings derived from several interpreters are sufficiently congruent. It refers to the degree to which different researchers performing similar observations in the field would generate similar interpretations (Franklin, Cody, & Ballan, 2010).

There are several methods for increasing reliability in qualitative research (Franklin et al., 2010). One method is the examination of the equivalence of responses to various forms of the same question. Perhaps you have two forms of a similar question that have different wording. You would examine whether the responses of the study subject were consistent for each version of the question. Another method for increasing reliability is the establishment of clear procedures for recording field notes. This prevents inconsistent recordings of observations on a theme. A third method for increasing reliability is cross-checking, a procedure whereby researchers use multiple team members to confirm their observations in the field. For example, one researcher might cross-check by comparing his or her enumerated codes with those of a fellow researcher.

Another method for improving reliability in qualitative research is being sure to stay close to the actual data (e.g., the actual words of the study subjects). For example, in

first-level coding, the researcher should use the words given by the study subject rather than replacing them with similar words or phrases thought to be more pleasing.

The above methods focus on finding sufficient regularity in the methods used to examine qualitative data to ensure that results are reliable. As with quantitative data, we cannot have validity without reliability when you are measuring your variables qualitatively.

Validity in qualitative measurement focuses on the extent to which researchers see what they think they see (Franklin et al., 2010). If a researcher records that a study subject is expressing anger, is this true? Maybe this is the normal way in which the study subject expresses concern, not anger.

In qualitative research, credibility refers to the truthfulness of study findings (Guba, 1981). Therefore, credibility is essentially the same as validity. A key mechanism for establishing credibility is to test the study subjects' endorsement of the study findings. Do they agree with your observations? Another is to ensure that the amount of time researchers spend engaging with the study subjects is sufficient to seem credible to a reasonable person. A 10-minute interview would be viewed in this context very differently from regular contacts of significant duration over a period of 6 weeks. A third mechanism for establishing credibility is the search for negative information—data that contradict certain basic findings. If there is a meaningful effort in this direction and little negative information was discovered, you have enhanced the credibility of your qualitative findings.

SECURING THE TOOL FOR MEASURING YOUR STUDY VARIABLES

If you are engaged in evaluative research, you will seek (or design) a tool for measuring the target behavior. There are a great number of published tools for this purpose, so your first step would be to seek such a tool. One useful source is Corcoran and Fischer (2013), who have produced a two-volume set of books containing hundreds of scales. Each scale is revealed in its entirety, and information is supplied regarding the validity and reliability of each scale, along with how it should be scored.

If you are unable to find a suitable published tool, you will need to design your own. This is not an easy task. Those who have published scales are typically experts on both the content of the tool and the methods of testing for reliability and validity. If you are not such an expert, you will need the assistance of the second part of this section on securing a tool.

Tips for Finding a Measurement Tool

A key advantage of the published scale is that it was designed by an expert on the theme of the tool. This person has spent a lot of time with the various tasks one must undertake when designing such an instrument. Furthermore, the scale has normally been tested for reliability and validity. So if you can find a published scale to meet your needs, you should use it, trusting that it will provide the reliability and the validity you need. Some scales cost money to use, but there are a great number that do not. Keep looking until you find the right one.

There are several things you need to do before you seek a published measurement tool. First, you must be keenly aware of your definition of the concept that you are measuring. Second, you should examine the nature of the persons who will reply to the tool; for example, well-educated adults and children in the third grade should have different tools to respond to. Third, you should examine the constraints that you face, such as the amount of time you will have for your study participants to complete the questionnaire. A tool that takes 20 minutes to complete may not be practical if you only have 1 hour weekly with this client for therapy. A fourth consideration is what you plan to do with the results. If you plan to submit your data for publication, you will need to abide by a higher set of standards than if you plan to use the results for agency decision making.

Let's examine these considerations in more detail. You have seen from previous chapters that the key to the selection of the measurement tool is your definition of the variable to be measured. Let's suppose that you have defined the concept of stressor as an event in life that can lead to stress, such as divorce, being fired from a job, the death of a close family member, and so forth. Let's suppose further that you have defined stress as a psychological state exemplified by words such as *tense*, *nervous*, and *uptight* and you have defined the opposite of stress as indicated by words such as *relaxed*, *comfortable*, and so forth. When you seek a scale for measuring stress, would you be looking for one that asks for the extent to which the respondent has had experience of events such as divorce, the death of a friend, or being fired from a job? If you did, you would have a measure of the concept of stressor rather than stress.

If you wanted a tool for measuring knowledge, you would seek tests that look a lot like the tests you have seen in school. This might take the form of a multiple-choice test, or a true/false test. If you want to measure opinions, on the other hand, you would examine tools with items that have options such as (a) *strongly agree*, (b) *agree*, (c) *disagree*, and (d) *strongly disagree*. The measurement of mental health conditions such as depression or anxiety would likely take a different form. This type of measure may ask the respondent to indicate the extent to which he or she feels a certain way. Options might include (a) *I do not feel sad*, (b) *I feel sad a lot of the time*, and (c) *I feel sad so often I cannot stand it*.

You will need to find a measurement device that is appropriate for your study subjects. You should look at the tool for wording to see if there may be problems. Characteristics of the study sample such as age and ethnicity are considerations. Will the items on the scale make sense to them? Will the method for their responses be clear? You will likely find information accompanying the description of the scale that identifies the appropriate audience.

Will you face constraints regarding time? Some scales take 20 to 30 minutes to complete. If you are using a tool for a single client where you are measuring the client at the beginning of each weekly therapy session for 6 weeks, you will not likely consider a scale with 50 items to be appropriate because it will take up too much of the weekly therapy session time. Instead, you would want to use a very simple tool in this situation. Another option, the modification of a published scale, is discussed in the next section of this book.

The previously mentioned work by Corcoran and Fischer (2013) is highly useful because it is rather comprehensive and convenient. Entire scales are included, and they do not cost anything to use for research studies. A sample of scales from that work are listed in Exhibit 13.2, but it is not a comprehensive list; many other scales are also found in this work.

EXHIBIT 13.2**A SAMPLE OF THE SCALES IN THE WORK OF CORCORAN AND FISCHER (2013)***Alcoholism:*

Michigan Alcoholism Screening Test (MAST)

Abuse:

Physical Abuse of Partner Scale

Anger:

State-Trait Anger Scale

Anxiety:

Clinical Anxiety Scale

Self-Rating Anxiety Scale

Stressful Situations Questionnaire

Assertiveness:

Assertiveness Scale for Adolescents

Children's Action Tendency Scale

Depression:

Depression Self-Rating Scale (Children)

Geriatric Depression Scale

Center for Epidemiologic Studies

Depression Mood Scale

Eating problems:

Bulimia Test

Compulsive Eating Scale

Eating Attitudes Scale

Family functioning:

Index of Family Relations

Family Awareness Scale

Guilt:

Perceived Guild Index

Health:

Coping Health Inventory for Parents

Hypochondriasis Scale for Institutional

Geriatric Patients

Illness Attitude Scale

Hope:

Hope Index

Loneliness:

Children's Loneliness Questionnaire

Revised UCLA Loneliness Scale

Marital relations:

Index of Marital Satisfaction

Obsessive-compulsive disorder:

Compulsiveness Inventory

Obsessive-Compulsive Scale

Parenting:

Adult-Adolescent Parenting Inventory

Parent-Child Relationship Survey

Peer relations:

Index of Peer Relations

Spouse abuse:

Index of Spouse Abuse

Posttraumatic stress:

Parent Report of Posttraumatic Symptoms

Impact of Event Scale

Problem solving:

Problem Solving Inventory

Satisfaction with life:

Life Satisfaction Index

Self-esteem:

Hare Self-Esteem Scale

Index of Self-Esteem

Social support:

Social Support Index

Stress:

Perceived Stress Scale

Using the Internet to Find a Measurement Tool

The Internet, of course, is a good source for finding a published scale. One option is to enter key words into an Internet search engine in order to find a tool. You will likely run into many results designed to get you to buy something, but if you push through those options, you should find something more in line with what you're looking for, like a book that contains the scale. Keep looking until you are successful.

What If the Published Scale Takes Too Long to Complete?

If an appropriate scale takes too long to complete, you can consider the option of taking a sample of items from this scale as your measurement tool. You should draw the items in a random fashion and report your procedures. In the opinion of this author, you should have a minimum of five items on your scale, and they should represent at least one third of the items on the original scale.

You will likely find that the administration of this modified scale and the administration of the entire scale would generate a rather high correlation, meaning that both versions of the scale are working in a very similar way, so the more extensive version is not necessary. You must, however, explain the nature of your modification in your report of your study. Furthermore, you cannot claim the reliability and validity reports that come with the full scale as the data for your modified scale. You should report both the information that accompanies the full scale and the fact that you have created a modified form of it.

Taking a portion of an existing scale for your own measurement will not work if you plan to publish your results. In that case, you will be expected to use the entire scale, the one with information on reliability and validity. Expectations about other aspects of the study will also be higher, such as the presentation of a knowledge base and the use of more sophisticated statistics. Getting published is a different ball game from doing a study for your agency or your research class.

EVALUATING MEASUREMENT TOOLS

Here are a few questions to ask yourself as you review possible measurement tools:

1. How well does this tool fit with my definition of the variable I am trying to measure?
 - a. What does the author say it measures?
 - b. What does it look like it measures?
 - c. Is it close enough to my definition of the variable?
2. Will it work with my study sample? Will they understand the words or instructions?
3. Will it take too much time for my situation? If so, can it be shortened so that it will be appropriate?

4. Does it have reasonable reliability and validity?
 - a. Start with face validity. Does it seem to you that it measures what you want to measure?
 - b. Examine the information on reliability and validity. If it has passed the test of any form of reliability or validity (after you have examined it for face validity), you should have confidence that it will work for you, assuming that your study does not need to meet the standards for publication.
5. Do you understand how to administer it and compute the scores?
6. Is the tool sensitive enough to measure the kinds of changes you expect in situations where you are measuring gain?

Some scales are more complicated than others. You will often see scales with items that mean the opposite of what you want to measure, which necessitates the reverse scoring of certain items. For example, on the stress scale used in a previous chapter, you saw items that indicated stress and some that indicated the opposite. For those items that indicated stress, respondents were scored as 3 if they answered *most of the time*, 2 if they answered *often*, 1 if they answered *some of the time*, or 0 if they answered *seldom or never*. But for those items that indicated the opposite of stress, the response *most of the time* was given the lowest score. Other choices followed suit, so that the higher respondents score on this scale, the more they have indicated that they are stressed.

When you conduct evaluative research, you will measure client progress in some way. What level of progress can you expect to achieve given the nature of your intervention? You should select an outcome and a related measurement tool that is sensitive to change in this context. If your intervention is brief, you should be realistic about the level of measured change you can expect. You can expect knowledge to change in short periods of training, but depression is likely to require many treatment sessions for notable change to occur.

TIPS FOR DEVELOPING YOUR OWN MEASUREMENT TOOL

If your search for a published scale is not successful, your task will be to develop your own tool. This process starts with your definition of the variable you wish to measure. Does your definition mostly focus on knowledge, feelings, opinions, behaviors, or attitudes? The format for items on a measurement device should be consistent with the nature of the thing you are measuring. For example, if your objective is to improve the knowledge of teenage mothers on good parenting practices, you would likely use a test that looks a lot like what you have seen in school, with multiple-choice options. However, you would not measure knowledge by giving response options appropriate for opinions, like *strongly agree*, *agree*, *disagree*, and *strongly disagree*. These options are suitable for measuring opinions, not knowledge.

When you are constructing the items for your measurement tool, you should be attentive to a number of suggestions. Here is a set of suggestions from Grinnell and Unrau (2014):

1. Make questions clear.
2. Keep sensitive questions to a minimum.
3. Avoid socially desirable responses.
4. Use only relevant questions.
5. Use simple language.
6. Ask questions the respondents are qualified to answer.
7. Avoid double-barreled questions.
8. Avoid negative questions.
9. Keep questions short.
10. Use items that have options that are mutually exclusive.
11. Use items that have options that are exhaustive.
12. Pretest the instrument.

To achieve clarity, you should think of the different ways in which study subjects might interpret the question. For example, when you ask if the respondent has been feeling a certain way, do you mean presently or at any point in time? Also, be sure to avoid jargon or the use of acronyms. Keep sensitive questions to a minimum, and place them toward the end of the tool rather than at the beginning. This is especially important if you are asking about the respondent's past incidents of crime or unsafe behavior with others. You have recently reviewed the nature of the social desirability bias, so you know to avoid questions that might solicit such a response.

Many of the suggestions on this list are easy to understand, such as keeping questions simple, asking only relevant questions, and keeping questions short. However, you may not know what a **double-barreled question** is. These are questions that pose two separate opinions in one question while asking for an overall agreement or disagreement with the question. For example, you should not have a question like "Do you agree that the Supreme Court has made good decisions and can be trusted?" A person could agree with one part of the question (made good decisions) but disagree with the other (can be trusted). If you need both pieces of information, you should pose two questions here.

Another suggestion is to avoid negative questions. Here is an example of a negative phrase you should not put before your respondents: "The federal government should not give welfare benefits to people who were not born in the United States." Instead, find a way to make it positive, like "The Federal Government should give welfare benefits to all eligible people living here." You may have caught that the first statement actually had a double negative, with two instances of the word *not*, which is worse than a single negative.

Before you employ your tool in a study, you should pretest it by giving it to a sample group of people to see how they respond. You might give it to your fellow students or members of a social club. If possible, you should use it with people who are similar to the ones who will be in your study. Then ask the respondents to give you their evaluations of the tool. Look over the results for anything unexpected, like finding that every single person in your pretest sample answered a given question in exactly the same way, preventing this tool from measuring a variable (if all answers are the same, you do not have a measured variable).

Keep the questionnaire simple and short. Do not put items in the questionnaire that you do not have plans to use. It is easy to start developing a questionnaire with a multitude of questions about a person's characteristics, such as gender, age, race, and political party affiliation. Take time to ask yourself if you will be using this information in your study. Don't yield to the temptation of putting questions on a questionnaire because you are "just curious." The longer the questionnaire, the more likely potential respondents will fail to complete it.

The options for a question should be both mutually exclusive and exhaustive. If they are mutually exclusive, the options do not overlap; if they are exhaustive, they include all possible categories. Consider this question:

What is your age? 0–30 30–40 40–50 50–60?

You may have noticed that someone who is 30 years of age would have two different appropriate options, as would someone who is 40 or 50 years of age. This means that the options on this question are not mutually exclusive. You may have also noticed that there is no place for someone to indicate that they are 61 years of age or older. This means that the question fails to be exhaustive. If it is exhaustive, it will have a category for everyone.

COMPUTING THE SCORE FOR YOUR SCALE

You will find instructions for the computation of the score for your scale in the article that describes the scale. Usually, the instructions are not very difficult to understand. You will take each questionnaire and compute a score for each of the variables in your study. Each study subject will have a score for each variable that you measure at the interval level. Each item of the scale will have points assigned in order of the severity of the response, with higher scores normally representing a higher level of the condition you are measuring. For example, if you have higher self-esteem than John, you will receive a higher score than John on the Hare Self-Esteem Scale, because this scale gives higher scores for higher self-esteem. However, some scales operate in the opposite fashion, in that higher scores represent a lower level of the condition. Therefore, you should pay attention to this fact when you read the description of how to score the scale.

Your biggest challenge will likely be the reverse scoring of items. Items that are to be reverse scored are ones that are stated in a different direction from those items that will not be reversed. You can find an example of reverse scoring in the Hare Self-Esteem Scale. The instructions for this scale tell you to reverse score negatively worded items, where some items are positive and others are negative. An example of a positive statement is "I

have at least as many friends as other people my age.” An example of a negative statement is “I am not as popular as other people my age.”

When an item is positive on this scale, you will assign the following scores:

1. A response of *strongly disagree* gets a score of 1.
2. A response of *disagree* gets a score of 2.
3. A response of *agree* gets a score of 3.
4. A response of *strongly agree* gets a score of 4.

However, when an item is negative, you will reverse the scores as follows:

1. A response of *strongly disagree* gets a score of 4.
2. A response of *disagree* gets a score of 3.
3. A response of *agree* gets a score of 2.
4. A response of *strongly agree* gets a score of 1.

You can see from the above instructions that in both cases the response indicating the highest level of self-esteem gets the highest score.

Always be sure to determine whether a higher score on your scale represents more of the target behavior or less. This is especially important when you are examining a correlation coefficient. If higher scores represent more negative conditions for your target behavior, you would normally expect to find a negative correlation between scores for this variable and other variables where higher scores represent positive behavior.

DESCRIBING YOUR MEASUREMENT TOOL

You will need to describe your measurement tool when you give your report about your research study. In this description, you need to clarify what higher scores represent. You also need to describe the nature of the items on the scale and define the variable that you are measuring. It is also helpful to include one or two items on the scale that give the reader a clearer idea of your measurement.

Typically, you will describe the phenomenon you are measuring along with your description of the scale. For example, you will see the following information on the Hare Self-Esteem Scale in the book by Corcoran and Fischer (2013).

- The Hare Self-Esteem Scale measures self-esteem in schoolchildren who are 10 years of age or older.
- The Hare Self-Esteem Scale is a 30-item instrument that consists of three subscales: (1) peer, (2) school, and (3) home. These dimensions of self-esteem represent the three main areas of interaction whereby a child develops a sense of self-worth.

- The Hare Self-Esteem Scale was tested on 248 students in the fifth and eighth grades, both boys and girls, and including both African Americans and Caucasians.
- The scoring entails the reverse scoring of negatively worded items and summing the scores using the following scores: a = 1, b = 2, c = 3, d = 4. Thus, higher scores represent higher self-esteem. Because there are 30 items on this scale, each of which can generate a score from 1 to 4, the total score for the scale can range from 30 to 120.
- This scale was tested for reliability using the test–retest method, where the correlation of the two scores was found to be .74. It was tested for validity through a comparison with two other scales designed to measure self-esteem, with the results indicating excellent validity.
- The primary reference for this scale is Hare (1985).

In some scales, you will see that the items are organized into categories that represent the dimensions of the scale. For the Hare Self-Esteem Scale, these dimensions include (a) peer self-esteem, (b) home self-esteem, and (c) school self-esteem. As you can probably guess, each of these categories represents self-esteem as viewed from each of these perspectives. It is possible that a child has higher self-esteem about peers than about the home or about school. If this is important to your study, you could calculate different scores for each of these dimensions of self-esteem.

Here is an example of one way to report on the Hare Self-Esteem Scale:

The Hare Self-Esteem Scale (Corcoran & Fischer, 2013) was employed to measure self-esteem for our study subjects. This scale measures self-esteem in children 10 years of age or above. It measures three dimensions of self-esteem: (1) peer self-esteem, (2) home self-esteem, and (3) school self-esteem. The items on this scale are consistent with the definition of self-esteem used in this study. For example, one of the peer self-esteem items is as follows: “I have at least as many friends as other people my age.” An item on measuring home self-esteem is “My parents are proud of the kind of person I am.” This scale has 30 items, each of which is scored on a 4-point scale with higher scores representing higher self-esteem. Scores can possibly range from 30 to 120. This scale has been subjected to tests of reliability and validity with positive results. For example, test–retest reliability for the scale was found to have a correlation of .74, while scores on this scale were found to have a correlation of .83 with scores on another scale of self-esteem.

DETERMINING PRACTICAL SIGNIFICANCE WITH YOUR MEASUREMENT METHOD

When you examine your data, you will address two issues: (1) statistical significance and (2) practical significance. As you have seen, statistical significance refers to the likelihood that your data can be explained by chance. If it can be explained by chance, you cannot conclude that your hypothesis was supported. The other issue is **practical significance**,

which refers to whether the data had sufficient magnitude to be meaningful. You determine magnitude by examining the extent of the client gain, the extent of the difference between the groups being compared, or the strength of the correlation. Each of these items answers the question of *how much*. If your data are of practical significance, you would say that the difference was good enough to meet your expectations. However, practical significance is not relevant to the question of whether the data supported the study hypothesis—that is an issue addressed only by statistical significance and whether your data went in the hypothesized direction.

Once you have a measurement device for your study variable, you should examine it for indicators of magnitude. Is there a percentage that can guide this decision? Is a 30% gain for a group of clients good enough? Is there a threshold that can be useful? You might, for example, have a depression scale where the authors have decided that a certain range of scores represents a certain level of depression. You could consider each of these levels to be a different threshold of functioning. Perhaps you will decide that if your treated clients have moved from a low threshold to a higher one, you have achieved practical significance.

Another method of examining the tool for practical significance is to examine what the scores for the items mean. Suppose, for example, that you are measuring marital stability and you ask the client to check each of the following things they have experienced in the past 3 months: (a) being separated and getting back together, (b) having a fight that was physical, (c) having a fight that led to police action, (d) being evicted from the home, and (e) losing a job because of marital trouble. You can see that each item on this scale is noteworthy. The presence or absence of each of these things is quite important. You might therefore conclude that a change in score of 1 or 2 points is of practical significance.

SPECIAL CHALLENGES OF THE SINGLE-SUBJECT RESEARCH STUDY

With group research, you normally measure a group of clients once before service begins and again at the end of treatment. Another group option is to measure progress for your group of clients and use the same measure for a comparison group of people who did not receive the service. You would then compare the gains of each group to see if the treatment group did better than the comparison group. When you are conducting single-subject research, on the other hand, you will measure a single client repeatedly throughout the treatment period. You cannot measure a single client just once before treatment begins and again at the end of treatment and subject these two scores to statistical analysis, because you need several recordings of data for proper statistical analysis.

There are special challenges with the single-subject research study. The most prominent one is that you will likely not find it feasible to administer a complex published scale at the beginning of each treatment session if it takes 20 minutes to complete the scale. In fact, you might find that even a 10-minute time period is not feasible because it still takes too much of the time available for treatment.

So what do you do? The solution is to select a simple tool. If you cannot find a simple published scale, you can design your own tool. Another option is to select a small, **random sample** of items on a published scale to use as your tool—a scale that would normally take 10 minutes to complete might only take 3 minutes in its modified form.

Another challenge with the single-subject study is that you may find that your single client expresses an array of target behaviors, not just one. If depression is only one of five target behaviors that the client exhibits, measuring depression will give you only a small amount of outcome data.

INDIVIDUALIZED SCALES AS AN OPTION

Sometimes the published scales do not address the unique outcomes that a client needs to achieve. There are two options for meeting this challenge that you will review here: (1) the use of the Outcome Rating Scale (ORS) and (2) the use of the YES scale. Each of these is simple, and you can use them with any client. Neither option identifies a specific behavior to be evaluated. Instead, both give the client much flexibility in expressing their perceptions on how well things have gone in their lives. There is also the Session Rating Scale (SRS), which shows the client's perceptions of the conduct of the treatment sessions without reference to outcome.

The Outcome Rating Scale and the Session Rating Scale

The ORS measures the client's perceptions of how his or her life has been going individually (personal well-being), interpersonally (family, close relationships), socially (work, school, friendships), and overall (general sense of well-being) (Duncan & Miller, 2017). You ask the clients to look over the past week and rate how well they have been doing with regard to these four dimensions of life. You ask them to mark where they are on a line. The line has no anchor points showing what each place on the line represents; instead, it has a negative end and a positive end. Marks to the left represent negative feelings, while marks to the right represent positive ones. You compute the client's score based on the length of the line from the negative end to the place that the client has marked, using a ruler to compute the score for each measurement. You can submit your data to the authors of the SRS or the ORS (Miller, 2017). Otherwise, you can obtain a copy of the ORS or the SRS by simply entering the names of the scales into an Internet search engine—one of the results will provide you with a copy of the scale.

The SRS gives the client the opportunity to show how he or she feels about the treatment sessions. Clients are asked to mark a point on the same type of scale as the ORS, except that in this case each end of the line has an anchor point description. There are four categories on this scale:

1. Being heard, understood, and respected
2. Working on the things the client wanted to work on
3. The approach the therapist was taking
4. How the session went overall

Each end of the line for the client's rating mark is a statement related to one of the above, with a negative statement at one end (e.g., We did not work on or talk about what I wanted to work on or talk about) and a positive statement at the other end (e.g., We worked on and talked about what I wanted to work on and talk about).

You will see an elaboration of the ideas of the ORS and the SRS in Exhibit 13.3. Here, you can see a description of these scales in the words of the scales' authors. If you choose not to sign up for the service provided by these authors, you may find it difficult to score the client. In that case, perhaps you could develop your own version of the scale with numbers

EXHIBIT 13.3

THE SESSION RATING SCALE AS A MECHANISM FOR MONITORING THE THERAPEUTIC RELATIONSHIP

Miller, Duncan, Brown, Sparks, and Claud (2003) developed the SRS as a means for seeing how therapy sessions are progressing in the eyes of the client. The scale is completed by the client at the end of each therapy session. It does not focus on the outcome but on the nature of the session and how the client views it. This exhibit provides information from Duncan and Miller. References to the "alliance" in this work mean the relationship between the client and the therapist.

The following is a quote from this website: <https://www.psychotherapy.net/article/therapy-effectiveness>

Research shows repeatedly that clients' ratings of the alliance are far more predictive of improvement than the type of intervention or the therapist's ratings of the alliance. Recognizing these much-replicated findings, we developed the *Session Rating Scale* (SRS) as a brief clinical alternative to longer research-based alliance measures to encourage routine conversations with clients about the alliance. The SRS also contains four items. First, a relationship scale rates the meeting on a continuum from "I did not feel heard, understood, and respected" to "I felt heard, understood, and respected." Second is a goals and topics scale that rates the conversation on a continuum from "We did not work on or talk about what I wanted to work on or talk about" to "We worked on or talked about what I wanted to work on or talk about." Third is an approach or method scale (an indication of a match with the client's theory of change) requiring the client to rate the meeting on a continuum from "The approach is not a good

fit for me" to "The approach is a good fit for me." Finally, the fourth scale looks at how the client perceives the encounter in total along the continuum: "There was something missing in the session today" to "Overall, today's session was right for me."

The SRS simply translates what is known about the alliance into four visual analog scales, with instructions to place a mark on a line, where negative responses are depicted on the left and positive responses are indicated on the right. The SRS allows alliance feedback in real time so that problems may be addressed. Like the ORS, the instrument takes less than a minute to administer and score. The SRS is scored similarly as the ORS, by adding the total of the client's marks on the four 10-cm lines. The total score falls into three categories:

- SRS score between 0 and 34 reflects a poor alliance
- SRS score between 35 and 38 reflects a fair alliance
- SRS score between 39 and 40 reflects a good alliance

The SRS allows the implementation of the final lesson of the supershrinks: *seek, obtain, and maintain more consumer engagement*. Most clients drop out of therapy for one of two reasons: (1) either the therapy is not helping (hence monitoring outcome) or (2) there is an alliance problem—the clients are not engaged or turned on by the process. The most direct way to improve your effectiveness is simply to keep them engaged in therapy.

from 1 (negative end of each continuum) to 10 (positive end of the continuum.). Another option is to do what the authors do and measure the length of the line in centimeters.

One of the limitations of the SRS and the ORS is that they do not allow clients a great range of self-expression in describing the outcomes they seek. On these scales, each client is asked to rate how things are going with regard to the four dimensions specified on the scale, not in their own words. The YES scale, however, does use the words of the client.

The YES Scale

The YES scale asks the client to describe in his or her own words the outcomes being sought. These outcomes might include “feeling like getting up in the morning and going to work,” “not yelling at my child when we talk about homework,” or “attending all of the AA meetings.” The clients list their desired outcomes and then rate their lives with regard to each outcome on a regular basis using a 10-point scale. The YES scale is presented in Exhibit 13.4.

EXHIBIT 13.4

THE YES SCALE

Your Evaluation of Service

This form is designed to receive your feedback on how well things are going for you with regard to the outcomes you would like to achieve from the service you are receiving. Your first task is to list the outcomes below:

Outcomes You Would Like to Achieve

-
-
- 1
 - 2
 - 3
 - 4
 - 5
-

Ask yourself how well things have gone for you (in the designated time period) with regard to each of the above outcomes. If things have gone miserably, you would circle Number 1. If things have gone only a little better than that, you would circle Number 2, and so forth, all the way up to a score of 10 if things have gone extremely well with regard to the designated outcome.

Objective 1	Miserably 1.....2.....3.....4.....5.....6.....7.....8.....9.....10	Extremely well
Objective 2	Miserably 1.....2.....3.....4.....5.....6.....7.....8.....9.....10	Extremely well
Objective 3	Miserably 1.....2.....3.....4.....5.....6.....7.....8.....9.....10	Extremely well
Objective 4	Miserably 1.....2.....3.....4.....5.....6.....7.....8.....9.....10	Extremely well
Objective 5	Miserably 1.....2.....3.....4.....5.....6.....7.....8.....9.....10	Extremely well

Your first step in the use of the YES scale is to see if the client likes this idea. You will explain how the scale works and show the client a copy of it. Your second step, assuming that the client would like to use the scale, is to get the client to articulate his or her desired outcomes. You could also get the client to think of how to describe each end of the continuum. For example, if the outcome is to be more positive when talking with your son, the most negative end of the continuum might be that the client was very negative all the time, while the positive end might be that the client was never negative.

If the client is oriented to further development of the scale, you could ask for descriptions of the middle points on the scale as well. Middle points might indicate that the client was slightly more positive than negative or slightly more negative than positive.

Determining the time frame of each evaluation would be the third step in the process. If you are meeting with the client once a week, then the past week would be the logical time frame for the client's use of the scale. In other words, the client would reflect on how things have gone specifically in the past week when completing the scale.

Now the client is ready to use the scale at the beginning of each treatment session. He or she will complete the scale, and you may find the review of the scores from past weeks to be a useful point of discussion in a treatment session.

Chapter Practice Exercises

Practice Exercises on Measurement

You have seen that measurement in social work research refers to the selection of a means for measuring each of your study variables. The selection process requires an understanding of a number of concepts and resources, starting with your definition of the variable to be measured. Once you define your variable, you will first seek a published scale. If you cannot find a suitable scale, you will design your own tool for measurement. Next, before you analyze your data with your chosen measurement tool, you need to decide on what change in scores would represent practical significance (if you are using this scale for evaluative research).

In Practice Exercise 1, you will find a tool to measure the concept of alcoholism. You begin by securing a definition of this target behavior and finding a measurement tool that is consistent with this definition. You will then describe the selected

tool in a manner suggested by the content of this chapter. You will also be asked to determine the amount of gain for clients on this tool that would be considered to be of practical significance after eight outpatient therapy sessions.

In Practice Exercise 2, you will select a tool for measuring client progress for a familiar intervention, preferably one with which you have had some experience. You will define the target behavior to be measured, secure a published scale for measuring it, and discuss what level of gain on this scale will constitute practical significance.

Practice Exercise 1: Finding a Tool to Measure Alcoholism

In this exercise, you will find a tool for measuring alcoholism. You will begin this process by developing a definition of this concept. Use the literature (and the Internet) for guidance on these tasks. Prepare a report that answers each of the following questions:

1. What is your definition of alcoholism?
2. Explain where you found guidance in the development of your definition. List the references that were helpful.
3. What measurement tool did you find for measuring alcoholism? How would you describe the structure of this tool? Give the source where you found this tool.
4. Explain where you found guidance in finding your measurement tool.
5. How much of a gain in pretest and posttest scores would you expect to find to conclude that practical significance has been achieved? Assume that the treatment has eight hour-long therapy sessions and four hour-long group therapy sessions. Identify the amount of gain (e.g., 8 points on the scale), along with the rationale for declaring that a certain amount of gain would constitute practical significance.

Practice Exercise 2: Finding a Tool for Measuring Your Clients' Progress

In this practice exercise, you will select a scale for measuring the progress of your clients on the objective being pursued. This exercise is only suitable for examples where you will be measuring some psychosocial variable such as depression, anxiety, or marital satisfaction. It is not appropriate if your target behavior can be measured by concrete agency records, such as school grades, whether the client showed up for an appointment, or whether the patient was readmitted to the hospital. It must be a scale that generates a score where the variable is measured at the interval level.

If your objective is to improve self-esteem, you will select a self-esteem scale. But if it is related to self-confidence, you would select a tool for measuring self-confidence, taking into consideration the differences between self-esteem and self-confidence. Your task, of course, will begin with your definition of the behavior to be measured.

You will prepare a report that answers the following questions:

1. What is the service you are evaluating?
 - a. What is the objective of this service?
 - b. What is the structure of this service?
2. What are the label and definition of the target behavior that is to be measured?
3. How would you describe the scale you will use to measure the target behavior?
 - a. What is the name of the scale?
 - b. How is the scale described in the literature with regard to what it measures and how?
 - c. What are a few of the items on the scale?
 - d. What is the range of scores that a person can get on this scale?
 - e. Is there a set of suggested thresholds showing levels of the condition that is being measured (e.g., severe depression, mild depression, no depression)?
 - f. What information is provided in the literature regarding the testing of the reliability or the validity of this scale?
4. What is the gain in functioning on your selected scale that would constitute practical significance? Explain.

CHAPTER REVIEW

Chapter Key Learnings

1. Effective measurement helps us get an objective picture of our variables and helps us achieve credibility with others in the reporting of our results. If you reveal objective data, you will be more likely to be heard.
2. Effective measurement requires attention to the reduction of measurement error, which is the distance between the data you have and the truth about the concept you are trying to measure. You reduce measurement error by demonstrating the reliability (consistency) and validity (accuracy) of the tools you are using for measurement.
3. You must understand the level of measurement of the variables in your study hypothesis to select an appropriate statistic for testing that hypothesis.
4. The chief mechanism for measurement in evaluation research is the standardized scale, where you give all study subjects the same instrument. Examples include the Beck Depression Inventory and the Hare Self-Esteem Scale.
5. You might use an individualized scale because you failed to find an appropriate standardized scale. Individualized scales are tailored for each client.
6. You evaluate the reliability and the validity of measurement tools in quantitative research by quantitative analysis of data regarding the tools you are using. These issues can be addressed in qualitative research by reviewing consistency among different observers and by having the study subjects review the results.
7. Finding a published scale for measuring your study variables is preferable to developing your own scale because the published scale has been developed by an expert on the concept being measured, who would have already tested the scale for reliability and validity.
8. Effective measurement requires a good definition of the study variable as well as congruence between the definition of the concept being measured and the tool that is measuring it.
9. When you need to develop your own scale, it can be helpful to review existing scales that measure concepts close to the one your scale will measure.
10. When you are developing your own measurement tool, you should be cognizant of the nature of the concepts you are measuring. For example, you will measure knowledge by test questions, you will measure opinions by items that ask about the degree of agreement one has about selected ideas, and you will measure psychological conditions with items that ask for the degree to which one has experienced certain feelings or thoughts.
11. In your description of your measurement tool, you will identify the name of the scale, give a careful definition of the concept being measured, provide information on the reliability and the validity of the tool, describe what the tool looks like, and report on how it is scored. There should be sufficient information from this description for the reader to assess the change in scores required for practical significance.

Chapter Discussion Questions

- Under what circumstances would you likely use a standardized scale for measuring a study variable rather than an individualized one?
- Under what circumstances would you likely use an individualized scale for measuring a study variable rather than a standardized one?
- Describe the situation where the YES scale or the ORS would be most appropriate.
- Explain why a scale can be reliable without being valid but a scale cannot be valid without being reliable.
- Is Cronbach's alpha useful for the testing of reliability or validity? Explain.
- Explain how your data might achieve statistical significant but not practical significance.
- What is one tip for developing your own scale that you found helpful?
- Review the description of the Hare Self-Esteem Scale under the section on describing your measurement tool. Did you find this description to be helpful? Do you have any suggestions about it?

Chapter Test

- Which of the following is/are true?
 - An example of measurement error is the distance between a person's score on a test (e.g., 80% correct) and the level of knowledge possessed by the person on the subject of the test (e.g., 90%)
 - There are at least two types of measurement error—systematic error and random error
 - Both of the above
 - None of the above
- Which of the following is/are true?
 - Reliability refers to the accuracy of an instrument for measuring a study variable
 - Validity refers to the consistency of an instrument for measuring a study variable
 - Both of the above
 - None of the above
- Which of the following is/are true?
 - A measurement tool can be reliable without being valid
 - A measurement tool can be valid without being reliable
 - Both of the above
 - None of the above
- Which of the following is/are true?
 - Content validity refers to the extent to which a measurement tool is internally consistent
 - Internal consistency can be evaluated with the employment of the alpha coefficient
 - Both of the above
 - None of the above
- Which of the following is/are true?
 - Because practical significance is a matter of opinion, the researcher does not need to explain why he or she came to the

- conclusion that practical significance was achieved
- b. Measurement tools should include items that elicit a socially desirable response
 - c. Both of the above
 - d. None of the above
6. What is the level of measurement of the following item on a questionnaire:
What is your gender? __ Male __ Female
- a. Nominal
 - b. Ordinal
 - c. Interval
 - d. Consistent
7. When you compute a correlation between the score on your scale for measuring self-esteem with another scale designed to measure self-esteem, you would expect to find which of the following?
- a. No correlation between the scores on the two scales
 - b. A positive correlation between the scores on the two scales
 - c. A negative correlation between the scores on the two scales
 - d. A curvilinear correlation between the scores on the two scales
8. Which of the following are advantages of using a published scale rather than developing your own scale to measure depression?
- a. If it has been designed by an expert on depression
 - b. If it has typically been tested for reliability and validity
 - c. Both of the above
 - d. None of the above
9. What is the weakness of the following item on a questionnaire?
What is your age? __ 0–30 __ 30–40 __ 40–50
- a. The options are not exhaustive
 - b. The options are not mutually exclusive
 - c. Both of the above
 - d. None of the above
10. Which of the following is *not* a thing you should do for a scale that you have designed yourself?
- a. Keep sensitive questions to a minimum
 - b. Keep questions short
 - c. Pretest the scale
 - d. None of the above—that is, all should be done

ANSWERS: 1 = c; 2 = d; 3 = a; 4 = b; 5 = d; 6 = a; 7 = b; 8 = c; 9 = c; 10 = d

Chapter Glossary

Content validity. The extent to which a measurement tool includes all the dimensions of the defined variable.

Criterion validity. The extent to which a given scale achieves the same results as another method of measuring the same thing. If the ratings for depression among a group of clients by

clinical social workers correlate positively with the scores given to these same people using your depression scale, you have evidence of criterion validity for your depression scale.

Cronbach's alpha. A coefficient that indicates the internal consistency of a scale by examining the correlations of different items with one another.

Double-barreled question. An item on a questionnaire that contains more than one issue but asks for only one answer.

Face validity. The extent to which a measurement tool appears to measure the intended variable in the opinion of knowledgeable people.

Internal consistency reliability. The extent to which a measurement tool contains items that are reasonably correlated with one another.

Measurement error. The distance between the reality of the variable under study and the value measured by an instrument designed to measure it.

Practical significance. The extent to which data results from a scientific study provide sufficient magnitude to be of practical utility. For example, you might ask, "Did the clients gain enough relief from their anxiety to suggest that the treatment was worthwhile, given its costs?"

Random error. Error in measurement that cannot be predicted.

Random sample. A study sample drawn from the study population on a random basis.

Reliability. The extent to which a measurement tool is consistent.

Social desirability bias. The tendency for some people to answer a question in a way that is socially desirable rather than saying what they really believe.

Systematic error. A form of measurement error that can be predicted, such as error based on the social desirability bias.

Test-retest reliability. The extent to which a measurement tool is consistent when applied to the same group of people at different times.

Validity. The extent to which a measurement tool is accurate.

Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.

This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.