Join the Allyn & Bacon First Editions Club!

To start your membership in our First Editions Club, simply read the following chapter and complete a brief questionnaire at: www.ablongman.com/swsurvey.

As a Thank You, we will send you a Penguin-Putnam title of your choice!

We look forward to hearing from you!

Research Methods for Social Work

James R. Dudley

©2005

ISBN-10: 0-205-36529-9

CHAPTER 5:
Defining and Measuring Concepts
The previous chapter focused on the topics researchers investigate. These topics are often referred to as research problems to be solved. That chapter describes a wide range of factors that have an influence on the selection of the research problem and explores the widely different ways in which researchers are likely to understand a particular problem. In this chapter, we will explore how the research problem is more fully understood and described through definitions and measurement.

Concepts, Variables, Values, and Constants

First of all, a research problem is understood by the concepts that give it form. Concepts are ideas or thoughts about which we have a mental image. Concepts are usually in an abstract form and are not readily observable. Because there is often no widely agreed-on measure of a particular concept, it can mean different things to different people. An example is the concept of crime, which can refer to any or all of a long list of illegal acts, such as robbery, burglary, assault, rape, murder, use of narcotics, fraud, and so on. It can refer to either misdemeanors or felonies or both. Crime could also be viewed by some as any type of infraction or transgression by one person against another, regardless of whether it is illegal. The important point here is that different people have different definitions of crime. Therefore, if you were to conduct a study about crime, you would initially be faced with the task of deciding what you mean by crime. This involves creating a definition and measure of crime so that other people will be able to understand it the way you do.

Consider the example below on how another concept, research-related anxiety, is defined in a study conducted by two social work researchers.
What are your reactions to the concept of research-related anxiety described above? Do you think this concept is evident in the real world? If you have research-related anxiety, do you think your anxiety is more of a fear or an eagerness to do well or maybe a little of both? Or would you describe it in a different way? How would you want your research instructor to respond to any research-related anxiety you have?

Let's look at another concept that is relevant to social work practice—substance abuse. Consider the varied ways this concept can be described.

Another factor important in defining and measuring research problems are variables. Variables are a central aspect of any research study. For example, if you don’t understand what the most important variables in a study are, you will probably not be able to understand other aspects of the study. A variable is defined as a concept that has two additional properties. First, it varies, or changes. Second, it is measurable. A variable adds something important to a concept: it provides a means of measuring it.

Poverty, for example, may be a concept of interest in a study, but it cannot be measured unless it is developed into a variable. Let's say that the researcher decides to measure poverty as a gross annual income of $14,000 or less for a family of four. By arbitrarily setting this income limit, a variable has been created for...
poverty. Now poverty is not only measurable but also meets the other property of a variable—it varies. It varies from a value of no income at all to various incomes below $14,000 annually. Any annual income of $14,000 or less (for a family of four) would be considered poverty. Any family income above $14,000 would not be considered poverty. Actually, many alternative measures of poverty could be used as well.

A value is an important aspect of a variable. Don’t confuse this term with social work values; they are altogether different. A value is one specific measure of a variable. For example, the variable, gender, has two values—female and male. Each participant is assigned to one value or the other. Another example is the variable, level of education. If the highest level of education completed by a person is a high school degree, this would be this person’s value of this variable.

A constant, unlike a variable, is a concept that doesn’t change. Since constants do not change or vary, there is little reason to give attention to them in research studies. An example of a constant is “people” in the statement “People communicate with language.” In other words, all people, not just some, communicate in some way using language. “Language,” however, can be turned into a variable, which then could take values such as English, Spanish, Russian, and so on. As another example, in many studies we focus on an identified population of people, such as homeless people, teen parents, infants, or Haitians. In these instances, everyone in the population is the same in terms of why they were selected. For instance, in a study of homeless people, everyone is homeless. Thus, homelessness is a constant, as it is assumed for all participants.

**Constructing and Measuring Variables**

Variables usually need to be constructed and measured to fit the needs of each study. Variable construction is a process of shaping an abstract concept into a measure. This process is also sometimes referred to as operationalizing a concept or providing an operational definition of a concept. Constructing variables usually involves several sequential steps. These steps include first identifying the key concept and then defining, describing, and developing a measure of the variable.

1. **Identifying the key concept.** A concept that is important to a study needs to be initially identified. The concept is then constructed into a variable that provides a measure of it. As was explained in the last chapter, understanding and defining the research problem is the first step in the overall research process. Defining the research problem also helps in identifying the key concepts to be studied. These concepts are usually embedded in the researchers’ definition of the problem.

   In a study about poverty, for example, the researchers begin their study by articulating such things as their view of poverty and its causes, who are the poor with whom they are especially concerned, and how poverty can be overcome. Once these views are articulated, the important concepts usually can be identified, and variable construction can be addressed. Seccombe’s study offers an example of how key concepts were initially identified.
2. Defining the variable. The next step involves defining the variable. A good way to approach this step is to develop a straightforward but simple definition of the variable. Writing definition statements may take some practice. Researchers often write the definitions of their variables over and over until they are concise and accurate. For example, in a study focusing on the assertiveness of social workers, assertiveness would need to be defined. In this case, being assertive could be defined as “being able to express your feelings, stand up for your rights, and those of others, and state your opinions without abusing or taking advantage of others” (Sundel & Sundel, 1980, p. 10).

3. Describing the variable. The third step in variable construction is to elaborate on the definition, or give it more description or detail. This step may involve setting the outer limits for the variable and differentiating it from other variables. Describing the variable, assertiveness, offers an example. Assertiveness can be clarified beyond the above definition by stating what it is not. Sundel and Sundel (1980) suggest that assertiveness needs to be distinguished from being underassertive and overassertive. Underassertive could be considered passive, submissive, and meek. Underassertive people frequently make excuses for themselves, are apologetic for what they say, and often blame themselves when things go wrong. Overassertive people could be described as aggressive, hostile, and arrogant. They often attempt to control or manipulate, blame, dominate, and intimidate others and sometimes use verbal or physical violence.

4. Developing a measure of the variable. Finally, the variable would be developed into a specific measure. Measurements can be developed easily for some variables but may be more difficult for others. First, let’s consider a couple of examples of variables in which a measure is easy to develop. Age is a variable, and it could be measured by asking participants to fill in their age in a blank space following the question “What is your age?” Another fairly easy variable to measure is satisfaction with various aspects of your job. A question could be asked, such as “How satisfied are you with your current job responsibilities?” with these measures: very satisfied, satisfied, dissatisfied, very dissatisfied.
Let’s look at an example of a more complicated variable like assertiveness. Because this variable has multiple dimensions and is reflected in varied types of behaviors, it will likely need to be measured by asking several questions, not one or two. A standardized scale by Sundel and Sundel (1980) offers one way to measure assertiveness (see Table 5.1). In this case, their scale measures assertiveness within the context of social work activities. Each question identifies a particular behavior that can be important to social workers related to whether they are assertive or not.

If you review the assertiveness scale in Table 5.1, you will see that it measures several dimensions of assertiveness based on different types of relationships. These authors decided that a comprehensive measure of assertiveness needs to take into account five different types of relationships on an equal basis. The first subset of five items (1–5) of this scale covers relationships with clients; the second subset (6–10) covers relationships with coworkers; the third subset (11–15), with subordinates; the fourth subset (16–20), with superiors; and the fifth subset (21–25), with professionals of other disciplines.

Also note that some items in this scale are stated in a way that is favorable to assertiveness, and other questions are stated in the opposite way, as unfavorable. You will see that the unfavorable statements can be easily identified, because they end in an asterisk. This mix of positive and negative statements is helpful in discouraging the research participant from getting into a pattern of answering all of the questions in the same way (e.g., answering all of the statements as being always or almost always true of me). This problem is referred to as a response pattern and does not accurately reflect the participants’ views. During the data analysis step, the scores for the unfavorable statements must be reversed so that these scores have the same meaning as the favorable statements. Then all of the items are stated in the same way, favorable to assertiveness, and a total score can be calculated for each participant by adding up the score for each item. This total score is a summary measure of each respondent’s assertiveness. A summary score that takes into consideration the responses to several statements on the same overall topic can also be referred to as an index of each respondent’s assertiveness.

You may want to fill out the assertiveness scale in Table 5.1 to determine how assertive you are. Do this by answering all of the items on the scale. Then, calculate your total assertive score in two steps. The first step is to change the values of reversed items that end with an asterisk (*). For example, change your responses on these items from 1 to 5; 2 to 4; 5 to 1; and so on. A value of 3 would not change. The second step is to add your numerical responses to get your assertiveness score. As you can see from this assertiveness scale, questions 11 to 15 are questions for those who supervise. A perfect assertiveness score is 100 for nonsupervisors (who do not respond to questions 11–15). A perfect assertiveness score for supervisors is 125. The lowest possible score for nonsupervisors is 20, and 25 for supervisors. You can decide for yourself whether your score reflects enough assertiveness. For example, a score of 75 or over for nonsupervisors could be viewed as very assertive. A score of 50 or less for nonsupervisors may suggest a need to give serious attention to improving one’s assertiveness skills. Incidentally, social workers should either have or work on developing strong assertiveness skills to be effective in their practice.
TABLE 5.1 How Assertive Are You?

Indicate the extent to which you would behave in the manner described by the following statements. Next to each statement, write the number that best corresponds to your behavior, according to the following code:

1 = never or almost never true of me
2 = rarely true of me
3 = sometimes true of me
4 = usually true of me
5 = always or almost always true of me

1. When a client arrives more than five minutes late for his/her appointment, I end the session on time.
2. When a client asks me how I handle my personal problems and I don’t want to divulge this, I refocus the client on his/her difficulty.
3. When a client tells me that s/he didn’t have time to do an assignment that s/he had agreed to do, I accept that and go on to the next issue.*
4. If a client failed to pay his/her fee for services, I would discuss this matter with him/her at the first available opportunity.
5. If a client calls to cancel an appointment, I say something to try to make him/her feel bad.*
6. In a staff meeting, I will voice my opinions when I think I should, even if it disagrees with my peers.
7. When a coworker asks me personal questions, I answer them because I’m too uncomfortable to refuse.*
8. When a colleague asks me to serve on a committee, I agree to even if I don’t want to serve.*
9. If a coworker borrowed money from me, I would ask him/her to pay it back.
10. When a coworker repeatedly asks me to cover for him/her so that s/he can conduct personal business, I refuse.
11. When a worker that I am supervising arrives late for our supervisory session, I end the session on time.
12. When a worker is late in handing in reports, it irritates me and I lose my temper.*
13. When a supervisee uses agency time to conduct personal business that interferes with his/her responsibilities, I become anxious and do not discuss this with him/her.*
14. When I observe a supervisee being rude or providing incorrect information to a client, I call him/her aside at the first available opportunity to discuss the situation.

(continued)
In summary, concepts and variables are a central focus of most research studies. Concepts are measured by first developing them into variables. Variables are then measured by the values assigned to them.

Understanding Measurement

Since measuring the concepts of interest to a study is central to the research process, the topic of measurement is introduced next. Measurement is defined as the process of determining the values of a variable for the people being studied. For example, the measurement of the educational level of the participants in a study would amount to finding out the highest educational level obtained by each person. The values could be the following: some high school, high school diploma, some college, undergraduate degree, and so on.
Measurement is a concept of central importance to research, as it offers valuable tools for conducting the steps in the research process that follows. Social workers encounter measurement issues in many daily activities. We frequently ask questions that beg for a measured response. For example, we may ask about the potential achievement level of an underachieving child with whom we work in a public school. Or we may question how likely a mother or father can parent adequately if child neglect is evident. We may inquire about the potential earning power of a welfare recipient who has neither a high school diploma nor any work history. Or we may find ourselves wondering about the intellectual ability of an adult with severe retardation when she seems to understand most of our nonverbal cues even though she has difficulty dressing herself. We may question whether adult abuse exists for a ninety-year-old woman who sleeps on the floor and turns on her stove to heat her apartment. And we may question whether we can help someone with schizophrenia if he is unwilling to take his medication regularly. These questions and countless others require answers that we need in our jobs. If we could come up with accurate answers, we might be able to make a difference in the lives of our clients.

Social workers are known for getting involved in “messy” situations that do not seem to have easy solutions. Many social work educators say that if you can’t stand ambiguity, you probably won’t survive as a social worker. These messy situations often involve complicated social circumstances that are described by ambiguous concepts. Many of these concepts have not had widespread agreement about their definition or measurement. Examples include poverty, happiness, well-being, racism, a woman’s right to choose and the right to life, love, gender equality, spirituality, employability, optimum levels of immigration, and so on. We use many of these concepts to describe the social problems our clients are experiencing or the solutions for their problems, and we plead for public policy responses that will effectively serve the greatest number or help those with the greatest need.

Let’s consider for a moment why the concepts just mentioned are so difficult to measure. In part it is because they are so abstract. The concept of happiness, for example, can vary, based on numerous factors. Many of these concepts are also rather nebulous and easily misunderstood by many people. Gender equality is an example of a misunderstood concept. Does it mean, for example, that males and females should be treated equally on everything? If this is so, how do we take into account factors like childbirth, breast feeding, and differing levels of physical strength that are gender related?

Some of these concepts are also politically laden, because they provide advantages to some people and disadvantages to others once we give them measures. Poverty is an example. A definition of poverty can determine how many people are perceived to experience it (based on this definition) and what can be done to help them. A definition of poverty is used by federal and state governments to determine who is eligible to receive welfare benefits, food stamps, Medicaid, subsidized housing, and other essential resources. We need to remember that when politics become evident, ethical dilemmas are also present. So a lot can be at stake in measurement.

Hopefully, you are now aware that a lot of the needs and problems of our clients are manifested in abstract concepts. Actually, much of our world is evident
in abstractions. Unless we understand this abstract world, we won’t be able to understand how it affects our clients and the problems that confront them. This is another reason why research studies and measurement are so important and often so challenging. If we are to understand and address our clients’ needs and problems, we must first attempt to measure these abstract concepts. We must also investigate the relationships among these concepts.

How Knowledge Is Generated

An unusual term to consider at this point is epistemology. Epistemology refers to the science of how knowledge is generated. Epistemologists provide standards for knowing how to generate knowledge in our field, particularly when we must move from the abstract world of ideas to the observable world where measurement can occur.

Consider the example of human intelligence as our abstract concept. Intelligence is a central dimension of our humanity and an important concept for social workers to understand. While intelligence consists of a complex array of cognitive functions, it is essentially beyond our observation and not directly measurable. As Figure 5.1 indicates, we must make a leap of “calculated” faith to develop a measure of it. Thus, we devise either an intelligence test or an adaptive behavior scale that is intended to reflect human intelligence in the observable world.

This leap of faith is how knowledge is generated. Although an abstract concept like intelligence cannot be measured directly, other standards are used in developing a measure of it that is accurate and consistent. These standards are referred to as validity and reliability and will be discussed later in this chapter.

Quantitative and Qualitative Measurements

Two overall approaches for conducting social research, the deductive and inductive, were introduced in Chapter 2. It was pointed out that the deductive approach
uses quantitative methods and the inductive approach uses qualitative methods. Deductive research begins by crafting a theoretical explanation for why something is happening in the real world (e.g., a hypothesis proposes that women tend to be Democrats and men, Republicans) followed by a second step of attempting to find support for it in the real world (e.g., a study could be conducted to determine the political affiliation of a group of women and men). In contrast, inductive research begins without any theoretical explanations. The first step is to immerse yourself in the real world and study something (e.g., find out why so many people are homeless by talking with homeless people on the street) and generate possible theoretical explanations from what you learn (e.g., possibly coming up with several explanations, such as some homeless people have a mental illness and are not being treated by the mental health system, while others are currently employed in a low-skilled job but cannot afford to pay for decent housing, and still others are suffering from substance abuse, and so on).

Deductive research usually involves measurements in the form of numbers and inductive research involves measurement in word form. Let’s first discuss how measurement can be created using quantitative indicators. This leads us to a familiar term, *numbers*.

**Numbers**

When we think of measurement, we usually think of numbers or numerical scores. For example, weight is measured by a numerical score of pounds, and height is measured by feet and inches. We use numbers to measure many things that we encounter. One advantage of such a measure is that it can be standardized and the scores can be comparable for several people who are studied. We often refer to numerical scores or responses as “quantitative data.” Quantitative data also have an advantage over qualitative data; they can be analyzed using statistics. Many researchers have a strong tendency to develop measures that are in numerical form because statistical analysis becomes possible.

One potential disadvantage, however, is that quantitative measures can be weakened by a problem called *reductionism* (discussed in Chapter 2). Reductionism results whenever an effort to create a numerical measure of a concept results in losing some of the properties of the concept. An example would be success in school being determined solely by a grade-point average. Other properties of success in school that would be overlooked in this measure would include development of interpersonal relationship skills, establishing a socially oriented values system, and developing confidence and skills for employment. Overall, measures using numerical scores have many advantages but also have limitations.

There are many questions that elicit a numerical response. Some of the questions raised earlier in the chapter could be answered with a numerical measure. For example, the question “What is the potential achievement level of an underachieving child?” can be answered with a numerical score if a school achievement test is administered. A school achievement test may produce a numerical score, such as
the upper 25 percentile of all who have taken the test, or score 355 out of total of 425. Similarly, the question asking about the “potential earning power of a welfare recipient who has neither a high school diploma nor any work history” could be measured by the salary that she earns at different points in time. A salary is in dollars, which is a numerical expression. Neither of the above numerical measures (test score or salary) offers a totally satisfactory answer to these questions, but at least they provide a partial one that is standardized and comparative. Other examples of pertinent concerns that can be answered numerically are as follows:

- The size of a social worker’s workload is measured by the number of cases assigned.
- The clients’ eligibility for services is determined by their salary and other assets.
- The severity of aggressive behavior exhibited by an emotionally disturbed child can be determined by the frequency with which he fights or bullies others.
- The success of a student’s course work (as you well know!) is measured as a numerical grade or a letter grade that is usually based on a numerical range of scores from 0 to 100 percent.

We have many quantitative measures at our disposal in social work. We can ask people numerous questions, such as their age, salary, number of children in their family, their educational level, and miles of travel from home to work or school. All of the answers are in some kind of numerical form. Note that most of these questions refer to demographic or background characteristics that may be of interest to a researcher or a social worker during an intake process.

Researchers could also ask other questions of a quantitative nature that may be important indicators of client progress or improvement. Examples include the number of times that someone

- is absent from school
- completes homework on time
- takes their medications
- verbalizes anger (rather than physically fights) when a conflict arises
- compliments or praises a spouse for any contribution to the family
- talks in a class discussion
- invites other residents living at a retirement village to lunch

Words

We can also use words to measure many things we encounter. We could ask, for example, several demographic questions that elicit a word response. Examples include What is your gender? (female or male) Race? (European American, African American, Latino, etc.) Nationality of birth? (Ireland, Poland, Kenya, Haiti, United States) Religious affiliation? (Protestant, Catholic, Jewish, Muslim) Neighborhood of residence? (in Philadelphia—West Philadelphia, South Philadelphia, Center City, and so on).
We can also ask other questions eliciting word responses that can be important indicators of a client’s attitudes, feelings, needs, or interests. Examples of such questions, with possible answers in parentheses, include:

- Do you like your courses? (yes, no)
- How do you feel today? (happy, OK, sad)
- Who is your best friend? (name)
- What is your favorite sport? (football, baseball, volley ball, soccer, tennis, track)
- How serious is this problem? (very serious, serious, somewhat serious, not serious)

We could also ask questions eliciting a word response that could be important indicators of a client progress, improvement, or increased knowledge. For example:

- How much of what you have learned from this training session can you use in your job? (a lot, some, very little, none)
- How satisfied are you with the supervision you are receiving? (very satisfied, satisfied, somewhat satisfied, not satisfied)
- How often are you aware of how you are feeling? (frequently, often, occasionally, almost never)
- Is having multiple sex partners an example of unsafe sex? (yes, no)
- Is a pattern of drinking alcohol in the morning one characteristic of alcoholism? (true, false)

Note that some of the above examples use a Likert scale. A Likert scale is a set of response categories that proceed in order from one extreme to the opposite extreme, such as strongly agree, agree, not sure, disagree, strongly disagree. Likert scales are discussed later in the chapter. It is also important to point out that all of the word-response questions discussed so far have a common feature: they elicit a forced response. Forced-response questions restrict the respondent’s answer to a predetermined list of response categories. Forced-response questions are also sometimes referred to as closed-ended questions in social work.

**Longer Word Forms**

The questions that have been discussed so far have been forced-response questions that can be answered with either a number or word. In addition, measurement can be evident in longer word forms, such as phrases, sentences, or more extended comments. When such open-ended questions are asked, they are likely to elicit responses in a longer word form.

Open-ended questions do not provide a restricted set of response categories from which respondents must choose their answers. Instead, respondents answer these questions in their own words. While open-ended question can elicit single-word responses, they are more likely to involve several words, a sentence or two,
or lengthier comments. As an example, if we ask a group of adolescents in a residential program if they like their residence, we could ask them to respond by selecting one of the following responses: yes, maybe, or no. This is a forced-response question; respondents are forced to choose one and only one response from the list. However, let’s say we want to find out more information about why these adolescents like or dislike their residence. We could ask a follow-up question that is open ended, such as “What do you like about your residence?” Also, to elicit any negative responses, we could ask “What do you dislike about your residence?” If we recorded their responses to these two questions in the adolescents’ own words, we could easily have as many different responses as we have people who are interviewed. Also, some responses may be short and limited to a word or two, and other comments may be much longer.

In one study, a sample of people with developmental disabilities were asked an open-ended question: “How do the staff members assist you in getting together with your best friend?” (Dudley, 2001). Some of their responses follow:

Drive me in car.
Helps me to find my way there.
They tell me to go and talk to (my best friend) when I’m upset.
They talk to me about who I need to get together with.
Remind me.
Staff lets friends come here.
They arrange visit.
I live with her (best friend).
They take me to see him.
I do it on my own.
Will help with phone calls to friends if needed.
Helping me get to church. Tell me about Jesus.
By doing stuff when you ask them.
Give me cigarettes.

As you can see, these responses reveal a wide variety of ways that staff assist clients in making contact with their best friend. Each response is expressed in an individualized way. Some of the responses indicate that transportation helps. Others suggest helpful directions for getting to a friend’s home. Discussions, encouraging visits, and other types of support are also mentioned. Also, one comment—“Give me cigarettes”—seems, at least on the surface, to be unrelated to the question.

Some researchers do not believe that longer word forms offer a measurement of something. These reservations are largely based on the view that the results are not standardized. In contrast, the viewpoint taken in this book is that measurement can be manifested in many different forms, including numbers, one word, or longer word forms. Measurement should not be limited to findings resulting from forced-response questions just because they are quantitative and can be analyzed by statistics.
Levels of Measurement

Next, let’s explore levels of measurement. Levels of measurement can be defined as the degree to which the values of a variable can be quantified. Levels of measurement are important to understand because they provide guidelines for constructing measurements of variables and determine how a variable can be analyzed. As Table 5.2 below indicates, there are four levels of measurement.

Let’s look more closely at each of these levels of measurement. Nominal variables have values that are categories. These categories must meet three properties: being distinct, mutually exclusive, and exhaustive. Being distinct means that the label assigned to a category clearly identifies what it means. For example, “Catholic” is a distinct category for the variable, religious affiliation, while “doubter” may not be. Second, nominal variables have values that are mutually exclusive or do not overlap. For example, Catholic, Protestant, Jewish, Muslim, and other would be categories of religious affiliation that don’t overlap. In other words, no one is likely to select more than one of these categories. Third, nominal variables have an exhaustive set of categories, which means that there is one category that fits the circumstances of each of the respondents. The earlier set of categories for religious affiliation (Catholic, Protestant, Jewish, Muslim, and other) is an example. If a respondent was a Buddhist or agnostic, for example, she could still select “other.” The “other” category is a miscellaneous category that can be included to ensure a choice for every respondent.

<table>
<thead>
<tr>
<th>Level of Measurement</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Nominal</td>
<td>Values are in two or more categories that are distinct, mutually exclusive, and exhaustive of all possibilities.</td>
<td>Race, gender</td>
</tr>
<tr>
<td>2. Ordinal</td>
<td>Values are in two or more categories, have all of the properties of nominal variables, and have a sequential order to them.</td>
<td>Social class, Likert-scale questions on attitudes</td>
</tr>
<tr>
<td>3. Interval</td>
<td>Values have a sequential order to them, and there is an equal distance between the different values. There is no true zero value.</td>
<td>Temperature</td>
</tr>
<tr>
<td>4. Ratio</td>
<td>Values have a sequential order to them, and there is an equal distance between the different values. Also, zero is a true zero value that represents an absence of the variable.</td>
<td>Income, age, weight</td>
</tr>
</tbody>
</table>
Ordinal variables have all three properties required of nominal variables. An additional property for ordinal variables is that these categories can be listed in sequential order, from higher to lower or greater to lesser. Likert scales are ordinal variables because they have an order to them. A set of Likert categories (e.g., strongly satisfied, satisfied, dissatisfied, and strongly dissatisfied) are listed in degree of satisfaction. Examples of other variables that have values in a sequential order are social class (upper class, middle class, lower class) and severity of a behavior disorder (severe, moderate, minimal, none). Note that it doesn’t make sense to talk about the distance between categories of ordinal variables.

The values of interval variables do not have response categories. They have individual numerical scores based on a standard unit of measure, which has numerical values along a continuum. With interval variables, we can compare distances between points meaningfully. Temperature in degrees Fahrenheit is an interval variable, because the values are numbers along a continuum. If it is 90 degrees on a particular day, it is 2 degrees warmer than 88 degrees; if it is 70 degrees, it is also 2 degrees warmer than 68. In other words, the distance between 90 and 88 is equal to the distance between 70 and 68. Other examples of interval variables are IQ (intelligence) scores and GRE (Graduate Record Examination) scores. With interval variables, we not only know that one value is higher or lower than another, but we also know how much numerical distance there is between values.

Ratio variables are similar to interval variables, except that zero is a true zero value. In other words, zero represents a true absence of the variable. The true zero makes a ratio variable meaningful in a special way. You can say one value is twice as great as another, for example. Income is a ratio variable. An income of 0 is a true zero. And it makes sense to say that someone who earns $30,000 earns three times as much as someone who earns $10,000. Temperature, on the other hand, is at the interval, not ratio, level, because 0 degrees is not a true absence of temperature. It doesn’t make sense to say that if it’s 100 degrees outside, it is twice as warm as if it’s 50 degrees. Most numerical variables used in research are ratio variables—for example, income, education, age, weight, height, number of children, and so on. These variables have a standard unit of measure, such as dollars and cents for income; years of school completed for education; years, months, and days for age; pounds for weight; feet and inches for...
height; and so on. In addition, zero for all of these variables represents a true absence of the variable.

Another way to distinguish the levels of measurement is to say that nominal and ordinal variables are categorical in nature; their values consist of discrete categories that are discontinuous from one another. For example, the values of gender—male and female—are discrete or distinctly different. In contrast, the values of interval and ratio variables are not categories. They are numerical scores deriving from a standard unit of measure. The values of interval and ratio variables are continuous, without interruption along a continuum. An example is age, which increases with each increment of time.

Because interval and ratio variables have properties that are very similar and use the same statistical tests, they will be combined together and referred to as interval/ratio variables. Interval/ratio variables have an advantage over the other levels of measurement because they can be analyzed using the most powerful statistical tests. More will be discussed on this advantage in Chapter 14.

Finally, we have prenominal variables, which refer to variables in qualitative studies that have not yet fulfilled the definition of nominal-level variables. Prenominal variables could have values that overlap. For example, we may ask participants of a study to tell us what activities they like to do. Their responses could include swimming, tennis, baseball, reading, hiking, skiing, watching TV, and so on. If we allow them to identify more than one activity, some participants will have more than one response. Therefore, a participant’s values may fit into two or more categories rather than just one.

Often, variables are initially developed or constructed in a prenominal stage, which suggests how important this stage can be. In this case, if we were to measure something, it could be in longer word form. We could, for example, be interested in finding out the most important challenges of divorced parents. By interviewing several divorced parents, we could record narratives that describe their problems, pain, hardships, hopes, and so on. After analyzing these narratives, we may conclude that three or four types of responses were predominant in the interviews. For example, some of the types of responses could be complaints about a shortage of money for covering daily expenses, a strained relationship with their former spouse, feeling overwhelmed by assuming most of the child-raising responsibility, and a desire to start a new life. Each of these types of responses could be formulated into a new set of questions that could be asked in another study. Also, these responses could be left in narrative form and presented as examples of the different types of responses that divorced parents are experiencing. More attention will be given to analyzing prenominal variables in Chapter 15.

Validity and Reliability

Before leaving the topic of measurement, we must consider an important question: How “good” are the measures we are using? We are asking how much we can depend on a measurement instrument to be satisfactory in measuring something of
interest to us. Actually, we are asking two important issues: Does the measure have both validity and reliability? Validity refers to whether the measure of the intended concept is an accurate reflection of the concept. Reliability refers to whether the measure of the intended concept consistently measures the same thing. Referring back to Figure 5.1, we are reminded that a leap of calculated faith is needed to develop a measure of a concept. This leap of faith involves using the standards of validity and reliability.

Validity

Validity is a standard that we use to determine whether an instrument measures what it is supposed to measure and whether it measures it accurately. Accuracy is the key issue here. For example, do the Federal Bureau of Investigation’s Uniform Crime Reports (2002) measure the incidence of rape? The answer is yes, because the FBI compiles statistics on the incidence of rape reported by local law authorities in each state. Next, we should ask, how accurately do the FBI Uniform Crime Reports measure the incidence of rape? The answer to this question is less certain. We know that many victims of rape do not report it to local authorities for several reasons, including the difficulty involved in prosecuting rapists and the embarrassment incurred in sharing such a horrific experience. So we must conclude that the FBI Uniform Crime Reports on rape are only partially valid because so many incidences go unreported. This point reminds us that determining validity is a matter of degree. There is no such thing as complete or absolute validity. The more validity that a measure has, the better.

Validity can be a problem if the research participant does not know a lot about the topic of inquiry. For example, a traditional father who is employed every weekday away from his home will not be able to give accurate answers to questions about what the children are doing during weekdays, such as who they play with, how much TV they watch, what they eat for lunch, and so on. If the mother in the family is usually home raising the children, she will be a more accurate source.

How Is Validity Measured? How do we know whether a measurement instrument is valid? There are many different ways of determining validity. Face validity is the simplest. Face validity means that the measure appears to be valid “on its face” (Babbie, 2001). In other words, as you look over the measure, it appears to be valid based on what you know about the concept or based on what your mental image is. If you are not sure, an additional way to consider face validity is to ask the same questions of an “expert,” or someone else who knows more about the variable than you. You could also ask such experts what they would suggest adding or changing to make it a more accurate measure of a variable you are constructing.

For example, let’s say you are attempting to develop a measure of active tenant participation in a neighborhood tenant council. Your goal is to achieve active participation in six months, and you need a measure of whether you have accomplished your goal. Initially, you decide to measure active participation as being at least eight tenants attending every monthly council meeting during the six-month
period. Next, you can ask the tenant council for their opinions on this measure. They agree with your measure, except that they believe that active participation can only be achieved if at least one tenant comes from each of the three high-rise apartment buildings in the housing complex. If you add their suggestion, you will not only increase the number and diversity of people who agree with this measure. You may also get the tenants more involved in efforts to achieve the goal.

Face validity can be a commonsense test to use. It involves carefully reviewing the measurement instrument and asking yourself whether it appears to measure accurately what you want it to measure. You can conduct a check of face validity fairly easily. The steps are simple, as outlined here.

**An Easy Way to Determine Face Validity**

You can use the Leadership Comfort Scale (Table 5.3, page 105) or the Sundel assertiveness scale in Table 5.1 (page 87) to practice these steps:

1. Fill out the measurement instrument.
2. Determine or calculate your overall score.
3. Use the author's interpretation of what your score means if such an interpretation is provided. (It is provided with the Sundel assertiveness scale.)
4. Ask yourself if you agree with the results (e.g., Does your score reflect what you perceive to be your comfort level with group leadership? Or does your score reflect your understanding of how assertive you are?).

Three additional tests of validity can help determine a measure’s validity in more rigorous ways. These three tests are content validity, criterion-related validity, and construct validity. Like face validity, these tests determine the degree to which a measure is valid, not whether or not it has absolute validity. Because each of these tests measures a different aspect of validity, it is preferable to use as many of these tests as possible in determining validity.

**Content validity** is the degree to which a measure covers the range of meanings included within the concept (Babbie, 2001). Content validity is particularly useful if you have developed an instrument that is to measure a multidimensional concept. Content validity would be used to determine if the questions you selected reflect all of the dimensions of the variable. An example would be a measure of prejudice. In order for a measure of prejudice to have content validity, it needs to reflect all types of prejudice, such as prejudice against women, older people, racial and ethnic minorities, religious minorities, and so on. If the variable was to be limited to prejudice against women, for example, content validity would be concerned with the measure covering prejudice in all of the important areas, such as in employment, family matters, financial opportunities, and so on.

The assertiveness scale in Table 5.1 is an example of a multidimensional variable that has content validity related to the different types of relationships that social workers encounter. The authors of this scale decided that an overall measure of assertiveness needed to take into account five different types of relationships, so they
constructed a scale that has five items about assertive behaviors for each of five types: clients, coworkers, subordinates, superiors, and professionals of other disciplines.

**Criterion-related validity** is another test of the validity of a measure. Criterion-related validity means that a measure is valid if its scores correlate with the scores of another measure of the same construct (Babbie, 2001). The other measure of the same construct would be the external criterion. Perhaps the external criterion could be a behavioral indicator that correlates with a verbal response to an interview question. For example, you may want to find out if the research participants in a sample are religious. You decide to measure this variable by asking, “How religious are you?” Response categories include very religious, religious, somewhat religious, and not religious. You could also find out how frequently the participants attend religious services. This would be the external criterion, which is a behavioral indicator of how religious they are. You could claim that your measure (“How religious are you?”) has criterion-related validity if the participants who indicate that they are religious also usually attend religious services regularly and those who indicate that they are not religious usually do not attend religious services.

Another example of criterion-related validity would be a correlation between a future behavior, such as actual future arrests, and a scale used to predict future crime recidivism. In another example, a correlation between a depression scale that you develop and a well-established measure like the Beck Depression Scale would be an example of criterion-related validity. Criterion-related validity is also referred to as predictive validity, because the external criterion predicts the new measure, and vice versa.

**Construct validity** is a third rigorous test of the validity of a measure. It is especially relevant to use when it is difficult to use criterion-related validity, because we cannot find a behavioral criterion of the measure (Babbie, 2001). In these instances, we can consider how the variable to be measured is logically associated to other variables. A hypothesis could be crafted that proposes that the variable in question is logically associated with another variable. For example, construct validity could be determined for an assertiveness scale by identifying other variables with which it has a logical relationship. In this case, it seems logical that assertiveness has a relationship to such variables as effective communication skills and success in salary negotiations. Construct validity would involve determining if measures of either of these two variables are correlated with scores of the assertiveness scale. Generally, construct validity provides less compelling evidence of validity than criterion-related validity, because the variable to which it is being compared is not another measure of the same variable but one that has a logical association with it.

**Reliability**

Reliability is the second key standard to be considered in determining if a measure is satisfactory. **Reliability** refers to the internal consistency of the measure. Does the instrument measure a concept consistently from one time to another? In addition, does the instrument measure a concept consistently among different people (i.e., is
Several factors impact reliability, as follows:

1. **Ambiguous wording.** Reliability is weak if the questions asked in a measurement instrument use words that are ambiguous. Words are considered ambiguous if they are likely to be interpreted differently by two or more research participants or by the same participant at different times. Let's consider the phrase “social work services” in the question “How satisfied are you with social work services?” One person could associate this term with one image, while another person associates it with something altogether different. “Satisfaction with social work services” could be associated with “liking the social worker” by one participant, “not having delays in receiving services” by a second participant, and “receiving the help that was requested” by a third participant. Since the phrase “social work services” is so general and open to so many interpretations, it would be better to ask about specific aspects of service, such as:
   - How satisfied are you with the availability of the social worker?
   - How satisfied are you that the social worker listens to what you say?
   - How satisfied are you with getting your needs met?

2. **Retrospective questions.** Reliability can also be a problem with retrospective questions. These questions ask for information about something that was experienced in the past. For example, a retrospective question could be “How many cups of coffee did you drink on average last month?” or “How many times did you visit a doctor last year?” Whenever a question asks for information requiring memory of the past, consideration needs to be given to how well participants will be able to recall. While people have widely varying capacities to remember things from the past, usually any question that asks about experiences a month or more in the past should be avoided.

3. **External factors.** External factors beyond the instrument could interfere with reliability as well. The characteristics of the people who administer the measurement instrument could influence reliability. For example, some participants’ responses could be swayed by varying the interviewer’s gender, race, age, accent used, attire, energy level, or interest level. Reliability can also be influenced by other external factors that influence the research participants. For example, the participants’ fatigue, motivation, and depression, all of which may vary, can affect reliability. Although all of these external factors and others need to be carefully thought about before administering an interview, there are limits to what can be done to eliminate all such interference.

**How Is Reliability Measured?** How do we know if a measure is reliable? Two simple methods are important to consider in determining the reliability of a measure: test-retest and interrater reliability. Both tests, like the validity tests, determine the degree to which a measure is reliable, not whether or not it has absolute reliability.

A **test-retest method** is the first means of determining reliability. A test-retest method determines whether a question is consistent by administering it to the same
person at two different points in time. The two different responses to the question are then compared to see if the participant’s responses are the same or similar. The second administration of the measure should be done fairly soon after the first measure to ensure that nothing has changed in the participant’s life related to what is being measured.

A test-retest method can be conducted fairly easily to strengthen a survey instrument. The questions can be administered to the same person(s) at two different times within close proximity of each other. Next, the answers to the same questions are compared to see if they are the same or not. Finally, you can look more closely at the questions that have different answers from the same person to see if they have ambiguous wording, retrospective questioning, or other shortcomings that may have added confusion. An exercise involving test-retest reliability is provided here.

**An Exercise Using the Test-Retest Method**

You can conduct a test-retest of the assertiveness scale in Table 5.1 to practice determining how reliable a measure is. Proceed as follows:

- Complete the scale at two different times within close proximity of each other. Make sure that your first set of responses are concealed when you complete the scale a second time.
- Compare your first and second pairs of responses for each item to determine if they are the same or not.
- When the two responses to an item are different, consider whether the item has ambiguous wording, is retrospective, or has other problems that can be corrected.
- Revise the items with inconsistent responses to improve their reliability.

The second method is called **interrater reliability**. In a test of interrater reliability, two or more people, referred to as raters, measure the same episode independent of each other. Afterward, they compare their results to determine if they are the same or similar. For example, let’s say you want to observe a client group session to determine who talks the most and the least. You want to make sure that the observational rating instrument you are using is reliable. You could have two raters observe the same group session independently of each other through a one-way mirror. They would be recording on the rating instrument how frequently each client talks. You would then compare the results of the two raters to see how much they are in agreement. Following is an example of how a researcher used interrater reliability in a study.

**Example of a Study Using Interrater Reliability**

Cheryl Waites (2000) conducted a study to determine if BSW students who completed their degree program gave better answers to the questions about a case vignette than BSW students enrolled in the introductory course. The case vignette
Two other more complicated tests of reliability can also be used. These are the split-half and the alternate form methods. The split-half method involves splitting the items of an instrument into two random groupings and finding out if there is a strong association between the responses to the two sets of items. The alternate form method involves administering two different forms of an instrument that are very similar and determining if the responses to one instrument are strongly associated with the responses to the other instrument.

Measures of reliability and some of the measures of validity are reported as correlation scores. These correlation scores provide indicators of how much validity or reliability are evident in a measure. Typically, measures of correlation range from $-1$ to $+1$, with a 0 being in the middle. Perfect validity or reliability would be a score of $-1$ or $+1$. No evidence of validity or reliability would be a 0. Also, a score of $-0.3$ or $+0.3$ is usually considered poor validity or reliability, $-0.4$ or $+0.4$ would be moderate validity or reliability, and $-0.7$ or $+0.7$ would be good validity or reliability.

Reliability should be considered in small, practical ways as you prepare your measurement instrument. You could, for example, include two identical questions in different places on a survey and determine if participants answered both questions with the same response. If not, you may have to consider the possibility that these and other questions are not reliable. Also, reliability can be checked by asking a forced-response and open-ended question on the same specific topic. If the responses to the open-ended question are consistent with those to the forced-response question for the same respondents, then you can have some confidence that these questions are reliable. For example, if most participants respond to the forced-response question “Do you like where you live?” with a yes but then respond to the follow-up question “What do you like about where you live?” with comments like “the staff don’t like me” or “I wish that I had a different roommate,” you may want to consider why there is an inconsistency between these responses. It could be that the respondents don’t fully understand the questions or that possibly the forced-response question is problematic because it favors a normative response of yes.
The connections between validity and reliability can be summarized as follows: A measure of an intended concept

- can have neither validity nor reliability
- can be reliable but not valid
- if valid, must be reliable
- has both validity and reliability

A measure can be reliable but not necessarily valid. In other words, an instrument can be measuring something consistently but not measuring what it is supposed to measure. For example, the FBI Uniform Crime Reports may be a fairly consistent measure but not a valid measure of mental illness. In contrast, a measure must be reliable if it is valid. If a measure is not consistently measuring the same thing, it is logical that it will not be an accurate measure of it.

Triangulation is a method of ensuring that a measure has validity and reliability. Triangulation is a process of using multiple methods to measure one concept. If researchers determine that the results of one measure of a variable are similar to the results from another measure of the same variable, they are triangulating the findings. This process also adds confidence that these two measures are both reliable and valid, because they have similar results.

Social workers are often involved in triangulating the information they collect from clients to ensure that it is accurate. For example, a social worker may cross-check the self-reporting of a parent accused of abusing her children with her behavior and with other sources, such as family members, the children, and neighbors. Sometimes, the more sources used in triangulation, the better.

Triangulation is often thought of as a way of guarding against any biases of the researcher and research participants (Taylor & Bogdan, 1984). Such biases can present serious threats to the integrity of the data that are collected, because they lessen their authenticity and accuracy. Whenever the researcher’s presence interferes with how people respond to a survey or how they behave when observed, a reactive effect is introduced. This reactive effect is often referred to as reactivity. Reactivity is a condition in which the researcher’s presence interferes with how people behave. Reactivity can be extremely problematic in any study.

**Standardized Scales**

Before leaving the topic of measurement, it is important to give more focused attention to standardized scales. Standardized scales are structured data collection instruments and can be questionnaires, interview schedules, or observation rating scales. These instruments are composed of a set of questions or statements that measure important concepts that are otherwise difficult to measure. These scales have been used extensively in prior research studies and practice and have strong evidence of validity and reliability.
Scales are available that measure numerous concepts relevant for social work practice in assessment, evaluation of treatment progress, and determination of overall program effectiveness. Standardized assessment scales are currently being recommended to social workers as a means of addressing new accountability requirements of regulatory and funding agencies (Dennison, 2002). These scales measure such things as anxiety, depression, hope, loneliness, self-esteem, stress, contentment, abusive and neglectful parenting behaviors, social supports, and well-being. Examples of such scales can be found in Ginsberg (2001) and Royse and colleagues (2001). The *Mental Measurements Yearbook* (2001) is another source for information on standardized tests covering scales on educational skills, personality, vocational aptitude, psychology, and related areas. This source is updated every year. The *Mental Measurements Yearbook* also has an online version you may be able to access through your university library. Sometimes a fee is required to use a scale, based on who has copyright to it.

Standardized scales are also useful in helping social workers assess their knowledge and skills in particular areas of practice. For example, Toseland and Rivas (2001) offer a scale to measure a person’s comfort with group leadership (see Table 5.3). This scale is intended to help people identify their strengths and weaknesses related to leading small groups, based on degree of comfort in ten different situations frequently experienced by group leaders.

### Table 5.3

**Leadership Comfort Scale**

Indicate your feelings when the following situations arise in the group. Circle the appropriate feeling.

<table>
<thead>
<tr>
<th>Situation</th>
<th>comfortable</th>
<th>uncomfortable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Dealing with silence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Dealing with negative feelings from members</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Having little structure in the group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Dealing with ambiguity of purpose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Having to self-disclose your feelings to the group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Experiencing high self-disclosure among members</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Dealing with conflict in the group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Having your leadership authority questioned</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Being evaluated by group members</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Allowing members to take responsibility for the group</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sometimes scales are suitable to use in studies, and other times they are not. One obvious reason for using a scale is that it may be the only way to validly measure a central concept in a study such as depression or self-esteem. However, standardized scales also have their disadvantages. As was mentioned earlier, a fee may be required for using them. The language used in a scale may not be compatible with the language used by the participants, or the scale itself may not quite measure a concept of interest. Furthermore, scales can be viewed by some research participants as too impersonal, and they can be difficult for some participants to fill out if they are not fully literate or have an inadequate command of the English language.

**Likert Scales**

Several specialized scales also exist. A few of them will be briefly introduced because they can be used in developing instruments for social work research. Likert scales are the most familiar. Likert scales give primary attention to the response categories used in a measurement instrument. As mentioned earlier in the chapter, a Likert scale is a set of response categories that proceed in order from one extreme to the opposite extreme, such as strongly agree, agree, not sure, disagree, strongly disagree. Note that Likert scales can be used as a set of response categories for a wide variety of statements, including statements that clarify values and attitudes and descriptions of behaviors. With a Likert scale, the respondent is presented with a statement in a questionnaire and asked to choose a response from one of the response categories given (e.g., “strongly agree,” “agree,” etc.). When identical response categories are used for several items or statements in a questionnaire, they can be scored and analyzed in a uniform manner.

Note that the measure of assertiveness in Table 5.1 (pages 87–88) is an example of a Likert scale. This scale consists of a questionnaire that includes several statements pertinent to assertiveness. Each of these statements uses the same set of response categories (“never or almost never true of me,” “rarely true of me,” “sometimes true of me,” “usually true of me,” and “always or almost always true of me”) that fit the definition of a Likert scale (namely, proceeding in order from one extreme to the opposite extreme).

The general principal is that Likert items should have five to eight values or categories. There are two reasons for this. The first reason is that forced-response questions are criticized because they “force” a person to select a response. With at least five options available, chances are better that respondents can find one with which they can agree. The second reason is that five to eight options provide enough “spread” of answers to be able to detect small changes in attitudes or opinions.

Likert questions with less than five response categories are also permissible, particularly when there is a good rationale. For example, if a question is asked and there are only two, three, or four possible answers, then there should be only that many response categories. An example is “Are you satisfied with your salary?” The responses could be “yes,” “sort of,” “no,” and “not sure.” Or if the question is asking for the correct answer to a topic such as “Is having multiple sex partners an example of unsafe sex?” the only possible responses would be “yes” or “no.” A “maybe” category
would not make sense to include. Perhaps a “don’t know” category could also be included to discourage respondents who don’t know from guessing.

There are other considerations for determining which Likert scale categories to use, based on what you know about a particular sample of participants. You don’t always need to spread out the scale evenly. For example, if you are surveying a specific group of senior citizens and you ask their age, you would not include any age categories below 60 or 65.

**Thurstone and Guttman Scales**

While the Likert scale is the most commonly used scale in questionnaires, other scales have special qualities that can be considered as well. Two scales in particular are the Thurstone and Guttman scales (Babbie, 2001). Because each of these scales adds important properties to the development of the items or statements that are used in a scale, they are briefly mentioned here. A **Thurstone scale** involves consulting a group of outside experts. The statements used in a Thurstone scale are selected by consulting a group of experts on a particular concept or topic and asking them to rate several statements in terms of how strongly each of these statements measures this concept. The statements that are consistently rated most strongly by most of the judges are included in the questionnaire, while the weaker statements are discarded. In this scale, the statements are likely to be more accurate or valid in measuring a concept because of the experts’ input.

The **Guttman scale** has special qualities as well. This scale recognizes that the statements used in a scale can vary in degree of importance. This scale takes into account the fact that some statements of a scale may be more extreme indicators of a concept than others. Measuring political activism offers an example of this.

**Measuring Political Activism**

Political activism is an important stance to take in promoting social and economic justice for our clients. The principles of a Guttman scale can be helpful in constructing a scale to measure the degree to which a social worker is politically active. Some of the **extreme statements** of such a scale might include:

- run for a political office
- become a local precinct captain for a political party
- present a policy position at a legislative hearing or a NASW-sponsored political event
- participate in a protest march or demonstration in Washington, D.C.

Some of the **weaker statements** measuring political activism might include:

- vote in an election
- discuss your political views with another person
- write a letter to a public official
- join PACE (the political action committee for NASW)
As this example suggests, statements like “run for a political office” and “speak at a legislative hearing” are more extreme measures of political activism than “vote in an election” or “discuss your political views with another person” even though all of these statements reflect political activism to some degree. Questionnaires that have such variability in their statements can treat the responses to the more extreme indicators as more important than the responses to the weaker statements. Also, the responses to such a scale might be expected to reflect patterns of interest to a researcher. For example, respondents with favorable responses to the more extreme indicators (e.g., “run for office”) could be predicted to have favorable responses to the weaker indicators (e.g., “vote in an election”). However, the converse pattern would not necessarily be expected: those responding favorably to the weaker statements would not necessarily have favorable responses to the more extreme statements.

Diversity Issues and Standardized Scales

Diversity issues are important to consider when deciding whether or not to use a standardized scale. Witkin (2001) suggests some good diversity questions to think over when considering whether to use a standardized scale.

- To what extent are the cultural experiences of people of color, gay men and lesbians, people with disabilities, and other disadvantaged groups considered by the scale?
- What are the negative consequences of having clients complete a scale, if any? Do they get put into a category that stigmatizes them?
- What can this scale tell me about a client beyond what I already know?

A Goal Attainment Scale for Evaluating Social Work Practice

Scales are useful in social work practice as well as research. A goal attainment scale (GAS) is one such scale that has widespread use in practice and is a useful resource to consider when measuring clients’ progress in reaching their goals. A goal attainment scale is presented in Figure 5.2. The first column of the scale in Figure 5.2 describes five points of measure: “most unfavorable outcome likely,” “less-than-expected success,” “expected level of success,” “more than favorable success,” and “most favorable outcome which is unlikely but plausible.” Note that each of these points of measure are also assigned a quantitative score of 1 through 5. The second and third columns of the scale are left blank, to be filled in by the social worker and client. The second column is to be used for the first goal, and the third column for the second goal. Each of the cells in these two columns are to reflect a descriptor of the level of anticipated success of the respective goal at a follow-up date.
The steps involved in using a GAS are fairly straightforward.

1. Consider whether or not the GAS is the most appropriate design to use with a particular client. Usually, this design is appropriate if more than one client contact is expected and the client’s goals are measurable.

2. Make sure that all of the pertinent issues of informed consent are covered and understood by the client (i.e., discuss with the client the purpose of the selected tool; how it works; how the client, worker, and agency can benefit; the expectations
you have for the client in participating in this evaluation; ways of ensuring confidentiality; options for the client to withdraw, etc.).

3. Help the client identify the problem.

4. Select a mutually agreed-on goal to overcome this problem.

5. Set a realistic date when the goal is to be achieved. Note that the form in Figure 5.2 provides space at the bottom for both the initial date when the GAS is constructed and the follow-up date when the goal is expected to be achieved.

6. Work with the client to construct specific predictions for a series of five levels of reaching the goal (most unfavorable, less than expected, expected level of success, more than expected, and most favorable outcome) by filling in the blank cells. Usually, the easiest way to do this is to first identify the “expected level of success,” followed by the outer limits of the “most unfavorable” and “most favorable outcome.” Finally the “less than expected” and “more than expected” levels can be identified. Now the GAS is ready to be used. As mentioned earlier, the form in Figure 5.2 can be used and provides space for two goals.

7. Measure the client’s actual level of attainment in reaching the goal during the beginning session as a baseline measure.

8. During the final session after the service has been provided, measure the client’s actual level of attainment in reaching the goal. Discuss the progress that has been made and how the GAS was helpful in measuring this progress.

Examples of how a few students and their clients have constructed goal attainment scales offer illustrations.

### Specific Goal Attainment Scales for Three Different Client Goals

**Goal 1: Stop Smoking**

1 = no intention of quitting smoking and smokes more  
2 = recognizes need to stop smoking but does not 
3 = recognizes need to stop and switches to a lighter brand  
4 = recognizes need to stop and cuts back 50 percent  
5 = fully recognizes need to stop and does

**Goal 2: Get Employed**

1 = remains unemployed  
2 = employed part-time  
3 = gainfully employed full-time and making $20,000/year  
4 = gainfully employed full-time and making $30,000/year  
5 = gainfully employed full-time and making $50,000/year

**Goal 3: Improve Parenting Skills**

1 = ignores parenting skills altogether  
2 = reads a book on parenting skills  
3 = learns one new parenting skill  
4 = applies one new parenting skill  
5 = applies two new parenting skills
The goal attainment scale has many special features that make it a useful evaluation tool, such as:

1. While this is scale is standardized in terms of the five levels of success, it can also be described in an individualized way related to the needs of each client’s personal circumstances in the blank cells.
2. The GAS can be used in a participatory way with a client, involving the client in identifying and thinking about personal goals, defining measures of success in meeting the goals, and measuring progress. When the tool is used in this way, it can be a great motivator in getting clients to work on their goals. Also, when discrepancies become evident between the worker and client about the scale or measures of it, they can be brought up for useful discussion.
3. It can be used as both a qualitative and quantitative measure.
4. It measures degrees of progress (five levels) rather than simply two levels (yes or no).
5. Each point on the scale can be multidimensional or include more than one variable (e.g., note how the goal to stop smoking in the above example can include recognizing the need to stop, switching brands, cutting back, and stopping smoking at each of the five levels).
6. It is a good tool for visualizing progress.
7. It can serve as a formal or informal contract with the client.
8. The qualitative aspects of the scale can be standardized and used with a group of clients with similar circumstances.
9. It can be used as an assessment tool as it is being constructed, which can facilitate greater understanding about the goal.
10. It can be used in almost any form that works for the client. For example, the tool could be simplified by reducing it to a three-point scale.

The GAS has limitations to consider and external barriers that sometimes need to be overcome. These limitations and barriers include:

1. It does not work well when only one session of service is provided.
2. It is difficult to use when there is a language barrier.
3. The agency may not encourage goal setting.
4. It is easier to use with some problems than others. For example, some counseling goals may be more complicated and difficult to operationalize into a five-point scale.
5. It does not work as well in a context other than a face-to-face interview.
6. The client may have a cognitive barrier.
7. It may not have meaning to the client or could be an unnecessary distraction.
8. Some clients may be suspicious of how it will be used beyond the worker-client relationship. This issue may need extensive discussion and assurance of confidentiality.
Values and Ethics, Diversity, At-Risk Groups, and Social Justice

Values and ethics are important to consider when it comes to measurement issues. The chapter has emphasized the ethical importance of developing measures that fit the people being studied. Validity, reliability, and triangulation offer important standards to use in doing this. Actually ethics may come into play when a decision is to be made about having either numerical or word forms of measurement. While some concepts can be easily measured using numerical scores, other relevant concepts cannot. Qualitative measures often need to be considered when addressing some of the more complicated and controversial concepts relevant to social work. An important question to ask in a critical consumer role is “Is important meaning manifested in the data lost by forcing qualitative findings into numerical or statistical formats?”

Diversity issues are also important in selecting concepts and measures of them. Several concepts were mentioned in the chapter that reflect some of the diversity evident in our clients’ social circumstances. Poverty, substance abuse, and crime were three such concepts discussed. Seccombe (1999) reminds us that we also need to listen to the voices of the research participants as we decide on concepts to study. The concepts that Seccombe investigated were evident in these voices—concepts like the stigma associated with welfare and the difficulties of managing a welfare budget.

Diversity comes into play when developing measures as well. For example, the wording used in measures should be in a form that is readily understood by the participants. Also, when response categories are identified, we should ask if we have considered the characteristics of all of the respondents in a study. For example, to what extent are the cultural experiences of people of color, gay men and lesbians, people with disabilities, and other disadvantaged groups considered when selecting a standardized scale or developing a new one?

The illustration about political activism in this chapter may help you to think more about this concept and its implications for social justice issues. As the illustration suggests, this concept is a challenging one to measure. The political activism of social workers is also a pertinent topic to explore, since participation in the political process is one of the important ways in which social and economic justice can be promoted. Hopefully, some will feel that this is a topic that could be more fully researched in social work circles.

Discussion Questions

1. A researcher decides to measure the variable “marital status” by asking the participants “What is your marital status?” This forced-response question is followed by the following three response categories: married, unmarried, and single. Identify two things that are problematic about this set of response categories. Base your response on whether or not they fulfill the three properties of nominal-level variables.
2. What is the level of measurement of each of the following variables?
- race
- social class
- sexual orientation
- age
- number of children in the household
- distance in miles from home to work
- a Likert scale (strongly agree, somewhat agree, somewhat disagree, or strongly disagree)
- numerical grade-point average
- letter grade for a course

3. How would you define and measure the complex multidimensional concepts listed below? Identify three or four interrelated questions (or statements) that collectively could measure each of these concepts. Note that an answer is already provided for the first concept on stereotypes about welfare recipients.
- stereotypes about welfare recipients held by people generally—Answer: Measure this variable by asking if the participants strongly agree, somewhat agree, somewhat disagree, or strongly disagree with each of the following statements: (1) Most welfare recipients misuse their money, (2) Most welfare recipients choose not to work, (3) Most welfare recipients neglect their children. (All of these statements are stereotypes.)
- knowledge about safe sex held by teenagers
- nutritional meals provided to children by their parents
- a variety of play opportunities available for children under 12 years of age in a neighborhood
- children provided adequate protection from physical harm in their homes
- adequate child care available for preschool children so a parent can go to work
- adequate transportation available so a parent can go to work

4. Assume that you want to determine if the staff members of your agency are satisfied with their jobs. You have defined job satisfaction to include the following dimensions of satisfaction: work assignments, salary, medical insurance, relationships with coworkers, and quality of supervision. Develop one Likert scale question for each of these dimensions of job satisfaction.

5. Determine the reliability of the assertiveness scale (Table 5.1) using the test-retest method. Do this by taking the test. Then two hours later conceal your first set of answers and take the test again. Afterward, compare your first and second answers for each question to determine if they are identical or no more than one score apart. Select one question in which your scores were quite different. Check to see if the question has ambiguous wording. Rewrite the question so that it is no longer ambiguous to you.

6. Develop a scale of political activism for social workers. Do this by crafting twelve political behaviors that describe various levels of political activism. Craft four behaviors that reflect an extreme level of political activism, four that reflect a moderate level, and four that reflect a weak level. Feel free to use the information in the example (page 107), which identifies some of the behaviors that could be used in a scale that measures political activism. Also develop a set of response categories that can be used (e.g., a Likert scale, a choice of yes-no categories, etc.).
7. Find a measure of anxiety with at least moderate reliability in the *Mental Measurements Yearbook*. What types of evidence are used to evaluate the validity of this measure of anxiety?

8. Weight is a ratio level of measurement. What values of weight could you use if you wanted to measure it as an ordinal variable? What values could you use if you wanted to measure it as a nominal variable?

9. Help a client to identify a goal. (If you are not currently seeing clients, identify a goal for a family member or friend and complete the rest of the exercise.) Next, construct a goal attainment scale with this client that delineates five levels of progress toward achieving a goal. If possible, return to the goal attainment scale during a later follow-up interview to measure and discuss the client’s progress.

10. Create or make up a worker-client scenario that could be relevant for using a goal attainment scale. Have two students role-play this scenario in class, both during an initial interview when the GAS is being constructed and during a follow-up interview when progress toward the goal is being measured. In the initial interview, have the social worker concentrate on explaining the goal attainment scale and involving the client in constructing levels of achievement for the follow-up interview. At the follow-up interview, have the social worker concentrate on discussing how both the worker and client rate the client’s level of achievement and discuss any differences in opinion about the level that the client has achieved.

References


