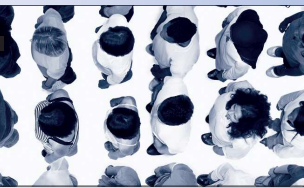


Conceptualization, Operationalization, and Measurement

CHAPTER OVERVIEW

The interested uses of conceptualization, operationalization, and measurement allow researchers to turn a general idea or a research topic into useful and valid measurements in the real world. An essential part of this process involves transforming the relatively vague terms of ordinary language into precise objects of study with well-defined and measurable meanings.



Introduction

Measuring Anything That Exists
Concepts, Indicators, and Reality
Concepts as Constructs
Conceptualization
Indicators and Dimensions
The Interchangeability of Indicators
Real, Nominal, and Operational Definitions
Creating Conceptual Order
An Example of Conceptualization: The Concept of Anomie
Definitions in Descriptive and Explanatory Studies
Operationalization Choices
Range of Variation

Variations between the Extreme
A Note on Dimensions
Defining Variables and Attributes
Levels of Measurement
Single or Multiple Indicators
Some Illustrations of Operationalization Choices
Operationalization Goals
On and On
Criteria of Measurement Quality
Precision and Accuracy
Reliability
Validity
What Decides What's Valid?
Tension between Reliability and Validity
The Ethics of Measurement

EngageNOW for Sociology

Use this online tool to help you make the grade on your next exam. After reading this chapter, go to [EngageNOW](#) to explore the chapter for instructions on how to benefit from EngageNOW.

Keeping Humanity in Focus

In the early 1970s, Elijah Anderson spent three years observing life in a black working-class neighborhood in South Chicago, focusing on how a Caribbean bar and liquor store while some people still believe that impoverished neighborhoods in the inner city are socially chaotic and disorganized, Anderson's study and others like it have demonstrated a different social structure that guides the behavior of its participants. Much of his interest centered on systems of social status and how the 55-year-old regular at Jilly's worked those systems to establish themselves among their peers.

In the second edition of this classic study of urban life, Elijah Anderson returns to Jilly's and the surrounding neighborhood. There he found several changes, largely due to the economic and racial segregation that has brought economic and mental depression to many of the residents. These changes in turn, had also altered the nature of social organization.

For a research method student, this book offers many insights into the process of establishing rapport with people being observed in their natural surroundings. It offers brilliant examples of how concepts are established in qualitative research.



Photo credit: © AP/Wide World

observed that seems to be an example of it. My sheet has what I've been told about it as well as I can mine that's same as yours.

The technical terms for those mental images, those sheets of paper in our mental file drawers, is *concepts*. That is, I have a conception of prejudice, and so do you. We can't communicate these mental images directly, so we use the terms written in the upper right-hand corner of our own mental sheets of paper as a way of communicating about our conceptions and the things we observe that are related to those conceptions. These terms make it possible for us to communicate and eventually agree on what we specifically mean by those terms. In social research, the process of coming to an agreement about what terms mean is *conceptualization*, and the result is called a *concept*. See "Keeping Humanity in Focus" for a glimpse at a project that reveals a lot about conceptualization.

I'm sure you've heard some reference to the many words Eskimos have for snow, as an example of how environment can shape language. Here's an exercise you might enjoy when you're ready to take a break from reading. Search the web for "Eskimo words for snow." You may be surprised by what you find. You're likely to discover wide disagreement on the number of, say, limit, words—ranging from 1 to 400. Several sources, moreover, will suggest that if the Inuit have several words for snow, so does English. Cecil Adams, for example, lists "snow, sleet, sleet, hail, powder, hard pack, Mirazid, firnics, flake, dusting, crink, anahapik."

conceptualization The mental process whereby fuzzy and imprecise notions (concepts) are made more specific and precise. See also *what you study* and *what you mean by "prejudice"*. Are there different kinds of prejudice? What are they?

A Concept in Search of a Label

In the late 1950s, in the heat of the Cold War between the United States and the Communist bloc, American foreign policy came under criticism for being sometimes arrogant and thoughtless regarding the cultures and concerns of other countries—especially the Europeans. U.S. aid to the people of Latin America and the Caribbean, for example, became a hot topic. In 1958, the image of an American in the Caribbean (shown here) became a symbol of the United States' foreign policy. The image was used to illustrate the concept of "American imperialism" in a book by the author, which was published in 1958. The image was used to illustrate the concept of "American imperialism" in a book by the author, which was published in 1958.

This term was taken from a 1958 political novel of the same name, by William L. Anderson and Eugene Bickel. Ironically, however, the meaning of the term was completely reversed in the book. In the novel, Anderson is the hero—a kind of American volunteer to help ally nations in a country (likely similar to Vietnam). Rather than pushing the villages around and suppressing his will, he's a thoughtful and considerate hero. When the local people discuss problems and possible solutions, he looks for ways to help them. The novel is a classic of the genre.

measuring, but this aim brings a pitfall with it: Getting on the "best" way of measuring a variable in a particular study may imply that we discovered the "real" meaning of the concept involved. In fact, concepts have no real, true, or objective meanings—only those we agree are best for that particular purpose.

This discussion implies that concepts, prejudice, and similar constructs can't be measured? Interestingly, a lot is not so. (And a good thing, too, or a bit of an ironic twist.) I've said that we can measure anything that's real. Constructs aren't real in the way that trees are real, but they do have another important virtue: They are useful. That is, they help us organize, communicate about, and understand things that are real. They help us make predictions about real things. Some of those predictions even turn out to be true. Constructs can work, this way because, although not real or observable in themselves, they have a definite relationship to things that are real and observable. The bridge from direct and indirect observables to useful constructs is the process called *conceptualization*.

Conceptualization

As we've seen, day-to-day communication usually operates through a system of vague and general agreements about the use of terms. Although you and I do not agree completely about the use of the term *compassion*, I'm probably safe in assuming that we both mean something by it. A wide range of misunderstandings and conflict—both from the interpersonal to the international—is the price we pay for our imprecision, but somehow we muddle through. Science, however, aims at more than muddling; it cannot operate in a context of such imprecision.

The process through which we specify what we mean when we use particular terms in research is called *conceptualization*. Suppose we want to find out, for example, whether women are more compassionate than men. I suspect many people assume this is the case, but it might be interesting to find out if it's really so. We can't measure "being" the question, for it alone agrees on the answer, without some working agreement about the meaning of *compassion*. They are "working"

Introduction

This chapter and the next deal with how researchers move from a general idea about what they want to study to effective and well-defined measurements in the real world. This chapter discusses the interrelated processes of conceptualization, operationalization, and measurement. Chapter 6 builds on this foundation to discuss types of measurements that are more complex.

Consider a nation such as "satisfaction with college." I'm sure you know some people who are very satisfied, some who are very dissatisfied, and many who are between those extremes. Moreover, you can probably place yourself somewhere along that satisfaction spectrum. While this probably makes sense to you as a general matter, how would you go about measuring how different students were, so you could place them along that spectrum?

There are some comments students make in conversations (such as "This place sucks") that would tip you off as to where they stood. Or, in a more active effort, you can probably think of questions you might ask students to learn about their satisfaction (such as "How satisfied are you with . . . ?"). Perhaps there are certain behaviors (class attendance, use of campus facilities, setting the dean's office on fire) that would suggest different levels of satisfaction. As you think about ways of measuring satisfaction with college, you are engaging in the subject matter of this chapter.

We begin by confronting the hidden concern people sometimes have about whether it's truly possible to measure the stuff of life—love, hate, prejudice, religious radicalism, alienation. The answer is yes, but it will take a few pages to see how. Once we see that researchers can measure anything that exists, we'll turn to the steps involved in doing just that.

Measuring Anything That Exists

Earlier in this book, I said that one of the two pillars of science is observation. Because this word can suggest a casual, passive activity, scientists often use the term *observation* to mean something more deliberate: observations of the real world for the purpose of describing objects and events in terms of the attributes composing a variable. You may have some reservations about the ability of science to measure the really important aspects of human social existence: If you've read reports dealing with something like liberalism or religion or prejudice, you may have been dissatisfied with the way the researchers measured whatever they were studying. You may have felt that they were too superficial, that they missed the aspects that really matter most. Maybe they measured religion as the number of times a person went to religious services, or maybe they measured liberalism by how people voted in a single election. Your dissatisfaction would surely have increased if you had found yourself being misclassified by the measurement system.

Your feeling of dissatisfaction reflects an important fact about social research: Most of the variables we want to study don't actually exist in the way that rocks exist. Indeed, they are made up. Moreover, they seldom have a single, unambiguous meaning.

To see what I mean, suppose we want to study *political party affiliation*. To measure this variable, we first need to decide what we mean by the term. We either the people we were studying were registered as Democrats or Republicans and take that as a measure of their party affiliation. But we could also simply ask someone what party they identify with and take their response as our measure. Notice that these two different measurement possibilities reflect somewhat different definitions

drift, frost, and iceberg" (Straight Dope 2001). This illustrates the ambiguities in the field with regard to the concepts and words we use in everyday communications and that also serve as the grounding for social research.

Let's take another example of a conception. Suppose that I'm going to meet someone named Pat, whom you already know. I ask you what Pat is like. How do you respond? You may tell me that children find their parents and put a tiny bird back in its nest. Pat goes out to take turkeys to poor families on Thanksgiving and to visit a children's hospital on Christmas. You've seen Pat wrap through a movie about a mother overcoming adversity to save and protect her child. As you search through your mental files, you may find all or most of those phenomena recorded on a single sheet labeled "compassionate." You look over the other entries on the page, and you find they seem to provide an accurate description of Pat. So you say, "Pat is compassionate."

Now I head through my own mental file drawer until I find a sheet marked "compassionate." I then look over the things written on my sheet, and I say "Oh, that's nice." I now feel I know what Pat is like, but my expectations reflect the entries on my file sheet, not your. Later, when I meet Pat, I happen to find that my own experiences correspond to the entries I have on my "compassionate" file sheet, and I say that you are very right.

But suppose my observations of Pat contradict the things I have on my file sheet. I tell you that I don't think Pat is very compassionate, and we begin to compare notes. You say, "I once saw Pat wrap through a movie about a mother overcoming adversity to save and protect her child. I look at my 'compassionate' sheet" and can't find anything like that. Looking elsewhere in my file, I locate that sort of phenomenon on a sheet labeled "sentimental." I retort, "That's not compassion. That's sentimentality!" To further strengthen my case, I tell you that I can't relate to the girl in a movie about an organization dedicated to saving whales from extinction. "That represents a lack of compassion," I argue. You search through your files and find saving the whales on two sheets—"environmental activism"

and "cross-species dating"—and you say so. Eventually, we set out comparing the entries we have on our respective sheets to see how they differ. We then discover that many of our mental images corresponding to that term differ.

In the big picture, language and communication work only in the extent that you and I have considerable overlap in the kinds of entries we have on our corresponding mental file sheets. The similarities we have on those sheets represent the agreements existing in our society. As we grow up, we're told approximately the same thing when we're first introduced to a particular term. Through our nationality, gender, race, ethnicity, religion, language, or other cultural factors may shape our understanding of concepts.

Dictionaries formalize the agreements our society has about such terms. Each of us, then, shapes his or her mental images to correspond with such agreements. But because all of us have different experiences and observations, no two people end up with exactly the same set of entries on any sheet in their mental file drawers. But if I say "prejudice" or "compassion," we must first stipulate what, exactly, counts as prejudice or compassion for our purposes.

Returning to the assertion made at the outset of this chapter, we can measure anything that's real. We can measure, for example, that Pat actually lives in the little red brick house on the hilltop on Christmas, wears that movie, or refuses to contribute to saving the whales. All of those behaviors exist, so we can measure them. But if I say "prejudice" or "compassion," we must first agree on what we can measure compassion in any objective sense. To measure this research interest, we need to know that those things I just described exist. Compassion exists only in the form of the agreements we have about how to use the term in communicating about things that are real.

Concepts as Constructs

If you recall the discussions of postmodernism in Chapter 1, you'll recognize that some people would object to the degree of "reality" I've allowed in the preceding comments. Did Pat "really" exist?

much. And yet, the groups' literature often speaks of their compassion for others. You want to explore this seeming paradox.

To pursue this research interest, you might arrange to interact with cult members, getting to know them and learning more about their views. You could tell them you were a social researcher interested in learning about their group, or perhaps you would just express an interest in learning more, without saying why.

In the course of your conversations with group members and perhaps attendance of religious services, you would put yourself in situations where you could come to understand what the cult members mean by compassion. You might learn, for example, that members of the group were so deeply concerned about sinners burning in hell that they were willing to be aggressive, even violent, to make people change their evil ways. Within their own paradigm, then, cult members would see bearing up for guys, prostitutes, and abortion doctors as acts of compassion.

Social researchers focus their attention on the meanings that the people under study give to words and actions. Doing so can often clarify the behaviors observed. At least you understand how the cult can see violent acts as compassionate. On the other hand, paying attention to what words and actions mean to the people under study almost always complicates the concepts researchers are interested in. (We'll return to this issue when we discuss the validity of measures, toward the end of this chapter.)

Whenever we take our concepts seriously and set out specifying what we mean by them, we discover disagreements and inconsistencies. Not only do you and I disagree, but so do you as likely to find a common middle ground within our own mental images. If you take a moment to look at what you mean by compassion, you'll probably find that your image contains several kinds of

Indicator An observation that we choose to conduct as a reflection of a variable we wish to study. Thus, for example, attending religious services might be considered an indicator of *religiosity*.

of political party affiliation. They might even produce different results: Someone may have registered as a Democrat years ago but gravitated more and more toward a Republican philosophy over time. Or someone who is registered with neither political party may, when asked, say she is affiliated with the one she feels the most kinship with.

Similar points apply to *religious affiliation*. Sometimes this variable refers to official membership in a particular church, temple, mosque, and so forth; other times it simply means whatever religion, if any, you identify yourself with. Perhaps to you it means something else, such as attendance at religious services.

The truth is that neither *party affiliation* nor *religious affiliation* has any real meaning, if by "real" we mean corresponding to some objective aspect of reality. These variables do not exist in nature. They are merely terms we've made up and assigned specific meanings to for some purpose, such as doing social research.

But you might object, *political affiliation* and *religious affiliation*—and a host of other things social researchers are interested in, such as prejudice or compassion—have some reality. After all, researchers make statements about them, such as "In Happiness, 55 percent of the adults affiliate with the Republican Party, and 45 percent of them are Episcopalians. Overall, people in Happiness are 16 percent in the pro-life and 84 percent in the pro-choice." Even ordinary people, not just social researchers, have been known to make statements like that. If these things don't exist in reality, what is it that we're measuring and talking about?

What indeed? Let's take a closer look by considering a variable of interest to many social researchers (and many other people as well)—*prejudice*.

Conceptions, Concepts, and Reality

As you and I wandered down the road of life, we observed a lot of things and knew they were real through our observations, and we heard reports from other people that seemed real. For example,

Table 5-1

What Social Scientists Measure

	Example
Direct observables	Physical characteristics (sex, height, skin color) of a person Being observed and/or interviewed
Indirect observables	Characteristics of a person as indicated by answers given in a self-administered questionnaire
Constructs	Level of prejudice, as measured by a scale that is created by combining several direct and/or indirect observables

hospital on Christmas? Does the hospital "really" exist? Does Christmas? Though we aren't going to be radically pessimistic in this chapter, I think you'll recognize the importance of an intellectually tough view of what's real and what's not. (When the intellectual game gets tough, the tough become social scientists.)

In this context, Abraham Kaplan (1964) distinguishes three classes of things that scientists measure. The first class is *direct observables*: those things we can observe rather simply and directly, like the color of an apple or the check mark on a questionnaire. The second class, *indirect observables*, require "relatively more subtle, complex, or indirect observation" (1964: 59). We note a person's check mark beside "feminist" in a questionnaire and have indirectly observed that person's gender. History books or minutes of corporate board meetings provide indirect observations of past social actions. Finally, the third class of observables consists of *anomalies*—theoretical constructs that are based on observations but that cannot be observed directly or indirectly. A good example is intelligence quotient, or IQ. It is constructed mathematically from observations of the answers given to a large number of questions on an IQ test. No one can directly or indirectly observe IQ; it is no more a "real" characteristic of people than is compassion or prejudice.

As you see, we use a lot of examples of what social scientists measure. Kaplan (1964: 49) defines *concept* as a "family of constructs." A *concept* is as Kaplan notes, a construct, something we create. Concepts such as compassion and prejudice are constructs that result from your conception of them, my conception of

- I personally heard people say nasty things about minority groups.
- We heard people say that women were inferior to men.
- We read about African Americans being lynched.
- We read that women and minorities earned less for the same work.
- We learned about "ethnic cleansing" and wars in which one ethnic group tried to eradicate another.

With additional experience, we noticed something more: People who participated in lynching were also quite likely to call African Americans ugly names. As a lot of them, moreover, seemed to want women to "stay in their place." Eventually it dawned on us that these several tendencies often appeared together in the same people and also had something in common. At some point, someone had a bright idea: "Let's use the word *prejudiced* as a shorthand notation for people like that. We can use the term even if they don't do all those things—as long as they're pretty much like that."

Being basically agreeable and interested in efficiency, we went along with the idea. That's where "prejudice" came from. We never observed it. We just agreed to use it as a shorthand, a term that represented a collection of apparently related phenomena that we've each observed in the course of life. In short, we made it up.

Here's another clue that prejudice isn't something that exists apart from our rough agreement to use the term in a certain way. Each of us develops our own mental image of what the set of real phenomena we've observed represents in general and what the phenomena have in common. When I say the word *prejudice*, it evokes a mental image in your mind, just as it evokes one in mine. It's so different that the two of us could combine thousands of sheets of paper, with each sheet of paper labeled in the upper right-hand corner. A sheet of paper in which our minds have the same prejudice on it. On your sheet all the things you've been told about prejudice and everything you've

them, and the conceptions of all those who have ever used these terms. They cannot be observed directly or indirectly, because they don't exist. We made them up.

To summarize, *concepts* are constructs derived by mutual agreement from mental images (conceptions). Our conceptions summarize collections of roughly related observations and experiences. Although the observations and experiences are real, at least supposedly, conceptions, and the concepts derived from them, are only mental creations. The terms associated with concepts are merely devices created for the purposes of filing and communication. A term, such as *prejudice*, is objectively speaking, really only a collection of letters. It has no intrinsic reality beyond that. It has only the meaning we've given to it. See "A Concept in Search of a Label" for an example of such an agreement.

Again, however, we fall into the trap of believing that terms for constructs do have intrinsic meaning, that they name real entities in the world. That danger seems to grow stronger when we begin to take terms seriously and attempt to use them precisely. The danger is that we begin to ignore the presence of experts who appear to know more than we do about what the terms really mean. It's easy to do so without in any such a situation.

Once we assume that terms such as *compassion* and *compassion* have real meanings, we begin the tortured task of discovering what those real meanings are. We begin by asking what *compassion* is objectively. We begin by asking what *compassion* is in the world. That danger seems to grow stronger when we begin to take terms seriously and attempt to use them precisely. The danger is that we begin to ignore the presence of experts who appear to know more than we do about what the terms really mean. It's easy to do so without in any such a situation.

Some times conceptualization aimed at identifying different dimensions of a variable leads to a different kind of distinction. We may conclude that the distinction between the same word for meaningfully distinguishable concepts. In the following example, the researchers find (1) that "violence" is a sufficient distinction of (2) that the concept "genocide" itself comprises several distinct phenomena. Let's look at the process they went through to come to this conclusion.

When Daniel Chert and Jennifer Edwards attempted to define the concept of "genocide," they found existing assumptions were not precise enough for their purposes:

The United Nations originally defined it as an attempt to destroy "in whole or in part, a national, ethnic, racial, or religious group." If genocide is distinct from other types of violence, it requires its own unique vocabulary. (2003: 14)

Notice the final comment in this excerpt, as it provides an important insight into why researchers are so careful in specifying the concepts they study. If concepts, such as the Holocaust, are simply another example of violence, like assaults and homicides, then what we know about violence in general might explain genocide. If it differs from other forms of violence, then we may need a different explanation for it. So, the researchers began by suggesting this "genocide" was a concept distinct from "violence" for their purposes.

- The frequency of worry about becoming a victim of three personal crises and two property crises in the immediate neighborhood.
- Estimates of likelihood of falling victims to each crime locally.
- Perceptions of control over the possibility of becoming a victim of each crime locally.
- Perceptions of the seriousness of the consequences of each crime.
- Beliefs about the incidence of each crime locally.
- Perceptions of the extent of social physical violence in the neighborhood.
- Perceptions of community cohesion, including informal social control and trust/social capital.

dimension A specifiable aspect of a concept. "Religiosity" for example, might be specified in terms of belief dimensions, a formal social belief dimension, a knowledge dimension, and so forth.

to Oklahoma in an event known as the “Trail of Tears,” which ultimately killed as many as half of those forced to leave.

2. *Revenge:* When the Chinese of Nanking bravely resisted the Japanese invaders in the early years of World War II, the conquerors felt they had been insulted by those who they regarded as inferior beings. Tens of thousands were slaughtered in the “Rape of Nanking” in 1937–1938.

3. *Fear:* The ethnic cleansing that recently occurred in the former Yugoslavia was at least partly motivated by economic competition and worries that the growing Albanian population of Kosovo was gaining political strength through numbers. Similarly, the Hittite attempt to eradicate the Tutsis of Rwanda grew out of a fear that returning Tutsi refugees would seize control of the country. Often intergroup fear such as these grew out of long histories of animosities, often inflamed in both directions.

4. *Purification:* The Nazi Holocaust, probably the most publicized case of genocide, was intended as a purification of the Aryan race. While Jews were the main target, gypsies, homosexuals, and many other groups were also included. Other examples include the Indonesian witch-hunt against communists in 1965–1966 and the attempt to eradicate all non-Khmer Kambodjians under Pol Pot in the 1970s.

No single theory of genocide could explain these varied forms of mayhem. Indeed, this act of conceptualization suggests four distinct phenomena, each needing a different set of explanations.

Specifying the different dimensions of a concept often paves the way for a more sophisticated understanding of what we’re studying. We might observe, for example, that women are more compassionate in terms of feelings, and men more so in terms of actions—or vice versa. Whichever turned out to be the case, we would not be able to say whether the actions are really more compassionate. Our research would have shown that there is no answer anyway to the question. That is, we cannot advance to our understanding of reality. To get a better feel for concepts, variables, and indicators, go to the General Social Survey

website and explore some of the ways the researchers have measured various concepts (see the link on this book’s website: <http://www.norc.uchicago.edu/habib/>).

The Interchangeability of Indicators

There is another way that the notion of indicators can help us in our attempt to understand reality by means of “unnatural” constructs. Suppose, for the moment, that you and I have compiled a list of 100 indicators of compassion and its various dimensions. Suppose further that we disagree widely on which indicators give the clearest evidence of compassion or its absence. If we pretty much agree on some indicators, we could focus our attention on those, and we would probably agree on the answer they provided. We would then be able to say that some people are more compassionate than others in some dimension. But suppose we don’t really agree on any of the possible indicators. Surprisingly, we can still reach an agreement on whether men or women are the more compassionate. How we do that has to do with the interchangeability of indicators.

The “it works like this. If we disagree totally on the value of the indicators, one solution would be to study all of them. Suppose that women turn out to be more compassionate than men on all 100 indicators—in all the indicators you favor and on all of mine. Then we would be able to agree that women are more compassionate than men, even though we still disagree on exactly what compassion means in general.

The interchangeability of indicators means that if we have different indicators on a nominal and operational definition. A nominal definition is one that is simply assigned to a term without any claim that the definition represents a “real” entity. Nominal definitions are arbitrary—we could define compassion in terms of

we should use the two sets of indicators represent different dimensions of compassion.

You have now seen the fundamental logic of conceptualization and measurement. The discussions that follow are mainly refinements and extensions of what you’ve just read. Before returning to a technical elaboration of measurement, however, we need to fill out the picture of conceptualization by looking at some of the ways social researchers provide standards, consistency, and commonality for the meanings of terms.

Real, Nominal, and Operational Definitions

As we have seen, the design and execution of social research requires us to deal with the confusion over concepts and reality. To this end, logicians and scientists have found it useful to distinguish three kinds of definitions: real, nominal, and operational. The first of these reflects the reification of terms. As Carl Hempel cautions:

“A ‘real’ definition, according to traditional logic, is not a stipulation determining the meaning of some expression but a statement of the ‘essential nature’ or the ‘essential attributes’ of some entity. The notion of essential nature, however, is so vague as to render this characterization useless for the purposes of rigorous inquiry.”

In other words, trying to specify the “real” meaning of concepts only leads to a nagging: Is mistake a construct, for a try to specify the

The specification of concepts in scientific inquiry depends on nominal and operational definitions. A nominal definition is one that is simply assigned to a term without any claim that the definition represents a “real” entity. Nominal definitions are arbitrary—we could define compas-

Specification The process through which concepts are made more specific.

as “glancing feathers of helix birds’ life” I wanted—to be that or be more or less useful. For most purposes, especially communication, that last definition of compassion would be pretty useless. Most nominal definitions represent some consensus, or convention, about how a particular term is to be used.

An operational definition, as you may remember from Chapter 4, specifies precisely how a term will be measured—that is, the operations we’ll perform. An operational definition is nominal rather than real, but it has the advantage of achieving maximum clarity about what a concept means in the context of a given study: the midst of disagreement and confusion over what a term “really” means, we can specify a working definition for the purposes of an inquiry. Wishing to examine socioeconomic status (SES) in a study, for example, we may simply specify that we are going to treat SES as a combination of income and educational attainment. In this decision, we rule out other possible aspects of SES: occupational status, money in the bank, property, lineage, lifestyle, and so forth. Our findings will then be interesting to the extent that our definition of SES is useful for our purpose.

Creating Conceptual Order

The clarification of concepts is a continuing process in social research. Catherine Marshall and Gretchen Rossen (1999: 18) speak of a “conceptual funnel” through which concepts are defined, as they become increasingly focused. Thus, a general interest in social activity could narrow to “trends in violence among teenagers,” “social capital,” and “father-foes on discovering” what experiences shaped the development of fully internalized concepts.

Some of these concepts are somewhat different, but they are inescapably linked to the language we use. In some forms of qualitative research, the clarification of concepts is a key element in the collection of data. Suppose you were conducting interviews and observations in a radical political group devoted to combating oppression in U.S. schools. You might have the following question: “What would shift your life, your work, and your identity into the members’ experiences and worldviews.”

TABLE 5.2

Progression of Measurement

Measurement Step	Example: Social Class
Conceptualization	What are the different meanings and dimensions of the concept “social class”?
Nominal definition	For our study, we will define “social class” as representing economic differences: specifically, income.
Operational definition	We will measure economic differences via responses to the survey question “What was your annual income before taxes, last year?”
Measurements in the real world	The interviewer will ask, “What was your annual income before taxes, last year?”

means to specific measurements in a fully structured scientific study.

An Example of Conceptualization: The Concept of Anomie

To bring this discussion of conceptualization in research together, let’s look briefly at the history of a specific social science concept. Researchers studying urban riots are often interested in the part played by feelings of powerlessness. Social scientists sometimes use the word *anomie* in this context. This term was first introduced into social science by Emile Durkheim, the great French sociologist, in his classic 1897 study, *Suicide*. Using only government publications on suicide rates in different regions and countries, Durkheim proved the work of analytic genius. To determine the effects of religion on suicide, he compared the suicide rates of predominantly Protestant countries with those of predominantly Catholic ones. Protestant regions of Catholic countries with Catholic regions of Protestant countries, and so forth. To determine the possible effects of the weather, he divided his body of data into northern and southern countries and regions, and he examined the different suicide rates across the months and seasons of the year. Although he drew conclusions about a surprisingly individualistic and personal act without having any data about the individuals engaging in it, he drew his general level. Durkheim suggested that suicide also reflects the extent to which a so-

ciety’s agreements are clear and stable. Noting that times of social upheaval and change often present individuals with grave uncertainties about what is expected of them, Durkheim suggested that such uncertainties cause anxiety, conflict, and even self-destruction. To describe this social condition of normlessness, Durkheim chose the term *anomie*. Durkheim did not make this word up. Used in both German and French, it literally means “without Durkheim’s use, Robert Merton, in a classic 1936 article entitled “Social Structure and Anomie” (1938), concluded that anomie results from a disparity between the goals and means available to individuals. Merton’s success, for example, is a widely shared goal in our society, yet not all individuals have the resources to attain this goal.

An emphasis on the goal itself, Merton suggested, produces normlessness, because those denied the traditional avenues to wealth stop getting it through legitimate means. Merton’s distinction, then, could be considered a further conceptualization of the concept of anomie.

The “it works like this” definition usually used the concept of anomie as a characteristic of societies, as did Merton after him, or social scientists have used it to describe individuals. In this distinction, some scholars have chosen to use

anomie in reference to its original, social meaning and to use the term *anomia* in reference to the individual characteristic. In a given society, there are some individuals experiencing anomie, and others do not. Elwin Powell, writing 20 years after Merton, provided the following conceptualization of anomie (through using the term *anomia*) as a characteristic of individuals:

When the ends of action become contradictory, inaccessible or insignificant, a condition of anomie arises. Characterized by a general loss of orientation and accompanied by feelings of “emptiness” and apathy anomie can be simply conceived as meaninglessness. (1958: 132)

Powell went on to suggest that there were two distinct kinds of anomie and to examine how the two rose out of different conceptual experiences to result at times in suicide. In his study, however, Powell did not measure anomie per se; he studied the relationship between suicide and occupation, making inferences about the two kinds of anomie. Thus, the study did not provide an operational definition of anomie, only a further conceptualization.

Although many researchers have offered operational definitions of anomie, one name stands out from the two: Powell’s article appeared, Leo Srole (1956) published a set of questionnaire items that he said provided a good measure of anomie as experienced by individuals. I consist of five statements that subjects were asked to agree or disagree with:

1. In spite of what some people say, the lot of the average man is getting worse.
2. It’s hardly fair to bring children into the world with the way things look for the future.
3. Nowadays a person has to live pretty much for today and let tomorrow take care of itself.
4. These days a person doesn’t really know who he can count on.
5. There’s little use writing to public officials because they aren’t really interested in the problems of the average man. (1956: 71)

convention and exclude all people under 14 years of age from the labor force.

This convention alone, however, would not give us a satisfactory definition, because it would count as unemployed such people as high school students, the retired, the disabled, and homemakers. We might follow the census convention further by defining the labor force as “all persons 14 years of age and over who are employed, looking for work, or waiting to be called back to a job from which they have been laid off or furloughed.” If a student, homemaker, or retired person is not looking for work, such a person would not be included in the labor force. Unemployed people, then, would be those members of the labor force, as defined, who are not “looking for work.”

But what does “looking for work” mean? Must a person register with the state employment service or go from door to door asking for employment? Or would it be sufficient to want a job to be open to one’s offer of employment? Conventionally, “looking for work” is defined operationally as saying yes in response to an interviewer’s asking “Have you been looking for a job during the past seven days?” (seven days is the period most often specified, but for some research purposes it might make more sense to shorten or lengthen it.)

As you can see, the construction of a descriptive study about the unemployment rate depends directly on how each issue of definition is resolved. Inconsistency in the period during which people are counted as looking for work would add more unemployed people to the labor force as defined, thereby increasing the reported unemployment rate. If we follow another convention and speak of the civilian labor force and the civilian unemployment rate, we’re excluding military personnel; that is, we’re excluding military personnel who are in the service of the armed forces. If we include such military personnel would be employed—by definition. Thus, the descriptive statement that the unemployment rate in a city is 3 percent, or 9 percent (since it might be, depends directly on the operational definitions used.

This example is relatively clear because there are few competing conventions about the labor force and unemployment. Now, consider how difficult it would be to get agreement about the

definitions you would need in order to say “Forty-five percent of the students at this institution are politically conservative.” Like the unemployment rate, this percentage would depend directly on the definition of what is being measured—in this case, political conservatism. A different definition might result in the conclusion “Five percent of the student body are politically conservative.”

Ironically, definitions are less problematic in the case of explanatory research. Let’s suppose we’re interested in explaining political conservatism. Why are some people conservative and others not? We may specifically like to see how we’re interested in whether conservatism increases with age. What if you and I have 25 different operational definitions of conservatism, and we can’t agree on which definition is best? As we saw in the discussion of indicators, this is not necessarily an insurmountable obstacle to our research. Suppose we found four people to be more conservative than we were. If we had 25 different definitions, clearly, the exact definition would matter much. We would conclude that old people are generally more conservative than young people—even though we couldn’t agree about exactly what *conservative* means.

In practice, explanatory research seldom results in findings quite as straightforward as this example suggests; nonetheless, the general pattern is quite common in actual research. There are consistent ways to measure things, and there are consistent ways to test in consistent research findings. However, such consistency does not appear in an inductive situation. Changing definitions almost inevitably results in different findings. The Importance of Variable Names explores this issue in connection with the variable *alienation participation*.

Operationalization Choices

In discussing conceptualization, I frequently have referred to operationalization, for the two are intimately linked. To receive operationalization is the refinement and specification of abstract concepts, and operationalization is the development of

The Importance of Variable Names

Patricia Fisher

Graduate School of Planning, University of Toronto

Operationalization is one of those things that never ends. It’s a simple task to explain to someone the purpose and importance of operational definitions for variables, and even to describe how operationalization typically takes place. However, until we’ve tried to operationalize a variable ourselves, you may not appreciate some of the subtle difficulties involved. Of considerable importance to the operationalization effort is the particular name that you have chosen for a variable. Let’s consider an example from the field of Urban Planning.

A variable of interest to planners is citizen participation. Planners are concerned that participation in the planning process by citizens is important to the success of any implementation. Citizens participation is a key to planners’ understanding of the real and perceived needs of a community, and such involvement by citizens tends to reinforce their cooperation with and support for planning efforts. Although many different conceptualizations might be offered by different planners, there would be little misunderstanding over what’s meant by citizen participation. The name of the variable seems adequate.

However, if we different planners’s provide very simple operational measures for citizen participation, we are likely to find a variety among the responses that their questionnaires produce. The same might keep a tally of attendance by private citizens at city commission and other local government meetings, or other might maintain a record

of the different types addressed by private citizens at similar meetings, while a third might record the number of local government meetings attended, and so on. Each of these measures would be used by the public officials, and many would be used by special-interest groups, during a particular time period. As skilled researchers, we can carry out a study that compares citizen participation efforts in many different dimensions of citizen participation, and forms of citizen participation. Therefore, the original naming of our variable, citizen participation, which was quite satisfactory from a conceptual point of view, proved inadequate for the purposes of operationalization.

The precise and exact naming of variables is important in research. It’s both essential to a level of good operationalization. Variable names quite often result from an inductive process of forming a conceptual definition, then an operational definition, then measuring that definition in a very specific, but often arbitrary, manner. If a researcher observes that a variable name and its measurement are not a reasonable fit, it is discarded. Sometimes the concept of the variable that you are studying is so different from the original one that you started with, but at the same time, you are measuring what you are measuring. Source: from Patricia Fisher, “The Importance of Variable Names: Copyright © Patricia Fisher, 1999, pp. 10–11.

Let’s suppose you want to measure people’s incomes in a study by collecting the information from either records or interviews. The highest annual incomes people receive turn into the millions of dollars, but not many people get that much. Unless you’re studying the very rich, it probably won’t help much to your study to keep track of extremely high incomes. In such a situation, you might probably want to establish a highest income category with a much lower floor—maybe \$100,000 or more. Although this decision doesn’t affect you or those together whose earnings are in the millions a year with paupers earning a mere \$100,000, it’ll survive it, and that missing poverty won’t be clear about the range of variation that interests them. The question is, to what extent are they willing to combine attributes in fairly gross categories?

Let’s suppose you want to measure people’s incomes in a study by collecting the information from either records or interviews. The highest annual incomes people receive turn into the millions of dollars, but not many people get that much. Unless you’re studying the very rich, it probably won’t help much to your study to keep track of extremely high incomes. In such a situation, you might probably want to establish a highest income category with a much lower floor—maybe \$100,000 or more. Although this decision doesn’t affect you or those together whose earnings are in the millions a year with paupers earning a mere \$100,000, it’ll survive it, and that missing poverty won’t be clear about the range of variation that interests them. The question is, to what extent are they willing to combine attributes in fairly gross categories?

Let’s suppose you want to measure people’s incomes in a study by collecting the information from either records or interviews. The highest annual incomes people receive turn into the millions of dollars, but not many people get that much. Unless you’re studying the very rich, it probably won’t help much to your study to keep track of extremely high incomes. In such a situation, you might probably want to establish a highest income category with a much lower floor—maybe \$100,000 or more. Although this decision doesn’t affect you or those together whose earnings are in the millions a year with paupers earning a mere \$100,000, it’ll survive it, and that missing poverty won’t be clear about the range of variation that interests them. The question is, to what extent are they willing to combine attributes in fairly gross categories?

Let’s suppose you want to measure people’s incomes in a study by collecting the information from either records or interviews. The highest annual incomes people receive turn into the millions of dollars, but not many people get that much. Unless you’re studying the very rich, it probably won’t help much to your study to keep track of extremely high incomes. In such a situation, you might probably want to establish a highest income category with a much lower floor—maybe \$100,000 or more. Although this decision doesn’t affect you or those together whose earnings are in the millions a year with paupers earning a mere \$100,000, it’ll survive it, and that missing poverty won’t be clear about the range of variation that interests them. The question is, to what extent are they willing to combine attributes in fairly gross categories?

Let’s suppose you want to measure people’s incomes in a study by collecting the information from either records or interviews. The highest annual incomes people receive turn into the millions of dollars, but not many people get that much. Unless you’re studying the very rich, it probably won’t help much to your study to keep track of extremely high incomes. In such a situation, you might probably want to establish a highest income category with a much lower floor—maybe \$100,000 or more. Although this decision doesn’t affect you or those together whose earnings are in the millions a year with paupers earning a mere \$100,000, it’ll survive it, and that missing poverty won’t be clear about the range of variation that interests them. The question is, to what extent are they willing to combine attributes in fairly gross categories?

For example, you might start thinking of oppression in physical and perhaps economic terms. The more you learned about the group, however, the more you might appreciate the possibility of psychological oppression.

The same point applies even to contexts where meanings might seem more fixed. In the analysis of social materials, for example, social researchers sometimes speak of the “hermeneutic circle,” a cyclical process of ever-deeper understanding.

The understanding of a text takes place through a process in which the meaning of the separate parts is determined by the global meaning of the text as it is anticipated. The closer determination of the meaning of the separate parts may eventually change the originally anticipated meaning of the totality, which again influences the meaning of the separate parts, and so on. (Erick 1996: 47)

Consider the concept “prejudice.” Suppose you needed to write a definition of the term. You might start out thinking about racial/ethnic prejudice. At some point you would realize you should probably allow for gender prejudice, religious prejudice, anti-gay prejudice, and the like in your definition. Examining each of these specific types of prejudice would affect your overall understanding of the general concept. As your general understanding changed, however, you would likely see each of the individual forms somewhat differently.

The continual refinement of concepts occurs in all social research methods. Often you will find yourself refining the meaning of important concepts even as you write up your final report. Although conceptualization is a continuing process, it’s vital to address it specifically at the beginning of any study design, especially rigorously structured research designs such as surveys and experiments. In a survey, for example, operationalization results in a commitment to a specific set of questionnaire items that will represent the concept you study. Without that commitment, the study could not proceed.

Even in less-structured research methods, however, it’s important to begin with an initial set of articulated meanings that can be refined during data collection and interpretation. No one seriously believes we can observe life with no preconceptions; for this reason, scientific observers must be conscious of and explicit about these conceptual starting points.

Let’s explore initial conceptualization the way it applies to structured inquiries such as surveys and experiments. Though specifying nominal definitions focuses our observational strategy, it does not allow us to observe. As a next step we must specify exactly what we are going to observe, how we will do it, and what interpretations we are going to place on various possible observations. All these further specifications make up the operational definition of the concept.

In the example of socioeconomic status, we might decide to ask survey respondents two questions, corresponding to the decision to measure SES in terms of income and educational attainment:

1. What was your total family income during the past 12 months?
2. What is the highest level of school you completed?

To organize our data, we’d probably want to specify a system for categorizing the answers people give. For income, we might use categories such as “under \$5,000,” “\$5,000 to \$10,000,” and so on. Educational attainment might be similarly grouped in categories like high school, high school, college, graduate degree. Finally, we would specify the way a person’s responses to these two questions would be combined to create a measure of SES.

In this way we would create a working and workable definition of SES. Although others might disagree with our operationalization and operationalization, the definition would have one essential scientific virtue: It would be absolutely specific and unambiguous. Even if someone disagreed with our definition, that person would have a good idea how to interpret our research results, because what we meant by SES—reflected in our analyses and conclusions—would be precise and clear.

Table 5.2 shows the progression of measurement steps from our vague sense of what a term

means to specific measurements in a fully structured scientific study.

Definitions in Descriptive and Explanatory Studies

As you’ll recall from Chapter 2, two general purposes of research are description and explanation. The distinction between them has important implications for definition and measurement. If it seems that description is simpler than explanation, you may be surprised to learn that definitions are more

problematic for descriptive research than for explanatory research. Before we turn to that set of examples of measurement, you’ll need a clear understanding of why this is so (we’ll discuss this point more fully in Part 6).

It’s easy to see the importance of clear and precise definitions for descriptive research. If we want to describe and report the unemployment rate in a city, our definition of being unemployed is obviously critical. That definition will depend on our conceptualization and operationalization. The distinction between them has important implications for definition and measurement. If it seems that description is simpler than explanation, you may be surprised to learn that definitions are more

problematic for descriptive research than for explanatory research. Before we turn to that set of examples of measurement, you’ll need a clear understanding of why this is so (we’ll discuss this point more fully in Part 6).

It’s easy to see the importance of clear and precise definitions for descriptive research. If we want to describe and report the unemployment rate in a city, our definition of being unemployed is obviously critical. That definition will depend on our conceptualization and operationalization. The distinction between them has important implications for definition and measurement. If it seems that description is simpler than explanation, you may be surprised to learn that definitions are more

problematic for descriptive research than for explanatory research. Before we turn to that set of examples of measurement, you’ll need a clear understanding of why this is so (we’ll discuss this point more fully in Part 6).

It’s easy to see the importance of clear and precise definitions for descriptive research. If we want to describe and report the unemployment rate in a city, our definition of being unemployed is obviously critical. That definition will depend on our conceptualization and operationalization. The distinction between them has important implications for definition and measurement. If it seems that description is simpler than explanation, you may be surprised to learn that definitions are more

problematic for descriptive research than for explanatory research. Before we turn to that set of examples of measurement, you’ll need a clear understanding of why this is so (we’ll discuss this point more fully in Part 6).

It’s easy to see the importance of clear and precise definitions for descriptive research. If we want to describe and report the unemployment rate in a city, our definition of being unemployed is obviously critical. That definition will depend on our conceptualization and operationalization. The distinction between them has important implications for definition and measurement. If it seems that description is simpler than explanation, you may be surprised to learn that definitions are more

problematic for descriptive research than for explanatory research. Before we turn to that set of examples of measurement, you’ll need a clear understanding of why this is so (we’ll discuss this point more fully in Part 6).

It’s easy to see the importance of clear and precise definitions for descriptive research. If we want to describe and report the unemployment rate in a city, our definition of being unemployed is obviously critical. That definition will depend on our conceptualization and operationalization. The distinction between them has important implications for definition and measurement. If it seems that description is simpler than explanation, you may be surprised to learn that definitions are more

problematic for descriptive research than for explanatory research. Before we turn to that set of examples of measurement, you’ll need a clear understanding of why this is so (we’ll discuss this point more fully in Part 6).

It’s easy to see the importance of clear and precise definitions for descriptive research. If we want to describe and report the unemployment rate in a city, our definition of being unemployed is obviously critical. That definition will depend on our conceptualization and operationalization. The distinction between them has important implications for definition and measurement. If it seems that description is simpler than explanation, you may be surprised to learn that definitions are more

problematic for descriptive research than for explanatory research. Before we turn to that set of examples of measurement, you’ll need a clear understanding of why this is so (we’ll discuss this point more fully in Part 6).

It’s easy to see the importance of clear and precise definitions for descriptive research. If we want to describe and report the unemployment rate in a city, our definition of being unemployed is obviously critical. That definition will depend on our conceptualization and operationalization. The distinction between them has important implications for definition and measurement. If it seems that description is simpler than explanation, you may be surprised to learn that definitions are more

problematic for descriptive research than for explanatory research. Before we turn to that set of examples of measurement, you’ll need a clear understanding of why this is so (we’ll discuss this point more fully in Part 6).

It’s easy to see the importance of clear and precise definitions for descriptive research. If we want to describe and report the unemployment rate in a city, our definition of being unemployed is obviously critical. That definition will depend on our conceptualization and operationalization. The distinction between them has important implications for definition and measurement. If it seems that description is simpler than explanation, you may be surprised to learn that definitions are more

problematic for descriptive research than for explanatory research. Before we turn to that set of examples of measurement, you’ll need a clear understanding of why this is so (we’ll discuss this point more fully in Part 6).

It’s easy to see the importance of clear and precise definitions for descriptive research. If we want to describe and report the unemployment rate in a city, our definition of being unemployed is obviously critical. That definition will depend on our conceptualization and operationalization. The distinction between them has important implications for definition and measurement. If it seems that description is simpler than explanation, you may be surprised to learn that definitions are more

problematic for descriptive research than for explanatory research. Before we turn to that set of examples of measurement, you’ll need a clear understanding of why this is so (we’ll discuss this point more fully in Part 6).

It’s easy to see the importance of clear and precise definitions for descriptive research. If we want to describe and report the unemployment rate in a city, our definition of being unemployed is obviously critical. That definition will depend on our conceptualization and operationalization. The distinction between them has important implications for definition and measurement. If it seems that description is simpler than explanation, you may be surprised to learn that definitions are more

problematic for descriptive research than for explanatory research. Before we turn to that set of examples of measurement, you’ll need a clear understanding of why this is so (we’ll discuss this point more fully in Part 6).

It’s easy to see the importance of clear and precise definitions for descriptive research. If we want to describe and report the unemployment rate in a city, our definition of being unemployed is obviously critical. That definition will depend on our conceptualization and operationalization. The distinction between them has important implications for definition and measurement. If it seems that description is simpler than explanation, you may be surprised to learn that definitions are more

problematic for descriptive research than for explanatory research. Before we turn to that set of examples of measurement, you’ll need a clear understanding of why this is so (we’ll discuss this point more fully in Part 6).

It’s easy to see the importance of clear and precise definitions for descriptive research. If we want to describe and report the unemployment rate in a city, our definition of being unemployed is obviously critical. That definition will depend on our conceptualization and operationalization. The distinction between them has important implications for definition and measurement. If it seems that description is simpler than explanation, you may be surprised to learn that definitions are more

problematic for descriptive research than for explanatory research. Before we turn to that set of examples of measurement, you’ll need a clear understanding of why this is so (we’ll discuss this point more fully in Part 6).

It’s easy to see the importance of clear and precise definitions for descriptive research. If we want to describe and report the unemployment rate in a city, our definition of being unemployed is obviously critical. That definition will depend on our conceptualization and operationalization. The distinction between them has important implications for definition and measurement. If it seems that description is simpler than explanation, you may be surprised to learn that definitions are more

categories, but you can never separate any variations you lumped together during observation and measurement.

A Note on Dimensions

We've already discussed dimensions as characteristic of concepts. When researchers get down to the business of creating operational measures of variables, they often discover—or worse, never notice—that they're not exactly clear about which dimensions of a variable they're really interested in. Here's an example.

Let's suppose you're studying people's attitudes toward government, and you want to include an examination of how people feel about corruption. Here are just a few of the dimensions you might examine:

- Do people think there is corruption in government?
- How much corruption do they think there is?
- How certain are they in their judgment of how much corruption there is?
- How do they feel about corruption in government as a problem in society?
- What do they think causes it?
- Do they think it's inevitable?
- What do they feel should be done about it?
- What are they willing to do personally to eliminate corruption in government?
- How certain are they that they would be willing to do what they say they would do?

The list could go on and on—how people feel about corruption in government has many dimensions. It's important to be clear about which ones are important in your inquiry; otherwise, you may measure how people feel about corruption when you really wanted to know how much they think there is or vice versa.

Once you've determined how you're going to collect your data (for example, surveying field agents), you have decided on the relevant range of variation, the degree of precision needed between the extremes of variation, and the specific

dimensions of the variables that interest you. You may have another choice: a mathematical logic of measurement. To do this, you need to take another look at attributes and their relationship to variables.

Defining Variables and Attributes

An attribute, you'll recall, is a characteristic or quality of something. *Female* is an example. So is *old* or *adult*. Variables, on the other hand, are logical sets of attributes. Thus, *gender* is a variable composed of the attributes *female* and *male*. What could be simpler?

Actually, some would insist that as it is the proper name of the variable composed of the physical attributes *female* and *male*, while *gender* is a social identity and behavioral variable composed of the attributes *female* and *male*. In most social science research, biological differences are less important than how people treat those differences in terms of their own behavior as well as their expectations and treatment of others. Despite this distinction, the two terms are commonly used interchangeably, both in everyday language and by social scientists. As long as the terms are defined for the purposes of research, there should be little confusion.

In any case, the conceptualization and operationalization processes can be seen as the specification of variables and the attributes composing them. Thus, in the context of a study of unemployment, *employed* and *unemployed* are the attributes. *Employed* and *unemployed* are the variables. This could also be expanded to include the other possible uses of the terms.

Every variable must have two important qualities. First, the attributes composing it should be exhaustive. For the variable to have any utility in research, you must be able to classify every observation in terms of one of the attributes composing the variable. We'll run into trouble if we conceptualize *age* as *young* and *old* because we've omitted *middle-aged*. *Republican* and *Democrat*, being two of the parties we set out to study will identify with

the Green Party, the Reform Party or some other organization, and some (often a large percentage) will tell us they have no party affiliation. We could make the list of attributes exhaustive by adding *other* and *affiliation*. Whatever we do, we must be able to classify every observation.

At the same time, attributes composing a variable must be mutually exclusive. Every observation must be able to be classified in terms of one and only one attribute. For example, we need to define *employed* and *unemployed* in such a way that nobody can be both at the same time. That means being able to classify the person who is working at a job but also looking for work. (We might run across a fully employed mail worker who is looking for the glamour and excitement of being a social researcher.) In this case, we might define the attributes so that *employed* takes precedence over *unemployed*, and anyone working at a job is employed regardless of whether he or she is looking for something better.

Levels of Measurement

Attributes operationalized as mutually exclusive and exhaustive may be related in other ways as well. For example, the attributes composing variables may represent different levels of measurement. In this section, we'll examine four levels of measurement: nominal, ordinal, interval, and ratio.

Nominal Measures

Variables whose attributes have only the characteristics of exhaustiveness and mutual exclusiveness are *nominal measures*. Examples include *gender*, *religious affiliation*, *political party affiliation*, *hair color*, *eye color*, *age*, and *hair color*. Although the attributes composing each of these variables—as *male* and *female* compose the variable *gender*—are distinct from one another (and exhaust the possibilities of gender among people), they have no additional structure. *Nominal measures* merely offer names or labels for characteristics.

Imagine a group of people characterized in terms of one such variable and physically grouped by the applicable attributes. For example, say

we've asked a large gathering of people to stand together in groups according to the states in which they were born: all those born in Vermont in one group, those born in California in another, and so forth. The variable is *place of birth*; the attributes are *born in California*, *born in Vermont*, and so on. All the people standing in a given group have that same group form, how close they are to one another, or how the groups are arranged in the room is irrelevant. All that matters that all the members of a given group share the same state of birth and that each group has a different shared state of birth. All we can say about two people in terms of a nominal variable is that they are either the same or different.

Ordinal Measures

Variables with attributes we can logically rank-order are *ordinal measures*. The different attributes of ordinal variables represent relatively more or less of the variable. Variables of this type are *social class*, *education*, *attainment*, *prejudice*, *intellectual sophistication*, and the like. In addition to saying whether two people are the same or different in terms of an ordinal variable, you can also say one is "more" than the other—that is, more conservative, more religious, older, and so forth.

In the physical sciences, *hardness* is the most frequently cited example of an ordinal measure. We may say that one material (for example, diamond) is harder than another (say, glass) if the former can scratch the latter and vice versa. By attempting to scratch various materials with other materials, we might eventually be able to arrange several materials in a row, ranging from

nominal measure A variable whose attributes have only the characteristics of exhaustiveness and mutual exclusiveness. In other words, we have a level of measurement describing a variable that has no attributes that are merely different, as distinguished from ordinal, interval, or ratio measures. *Gender* is an example of a nominal measure.

the softest to the hardest. We could never say how hard a given material was in absolute terms; we could only say how hard in relative terms—which means it's harder than and which softer than. Let's pursue the earlier example of grouping the people at a social gathering. In this time imagine that we ask all the people who have graduated from college to stand in one group, all those with only a high school diploma to stand in another group, and all those who have not graduated from high school to stand in a third group. This manner of grouping people satisfies the requirements for exhaustiveness and mutual exclusiveness discussed earlier. In addition, however, we might logically arrange the three groups in terms of the relative amount of formal education (the shared attribute) each had. We might arrange the three groups in a row, ranging from most to least formal education. This arrangement would provide a physical representation of an ordinal measure. If we knew which groups two individuals were in, we could determine that one had more, less, or the same formal education as the other.

Note in this example that it is irrelevant how close or far apart the educational groups are from one another. The college and high school groups might be five feet apart, and the less-than-high-school group 500 feet farther down the line. These actual distances don't have any meaning. The less-than-high-school group and the college group, or the high school and the less-than-high-school group, are the only comparisons that matter. It would be incorrect to infer that someone with an IQ of 150 is 50 percent more intelligent than someone with an IQ of 100. (A person who has a score of 100 on a standard IQ test could just be regarded, strictly speaking, as having no intelligence, although we might feel he or she was unimpaired to be a college professor or even a college student. But perhaps a dean . . .)

Interval Measures

For the attributes composing some variables, the actual distance separating those attributes

ordinal measure A level of measurement describing a variable with attributes that can be rank-ordered along some dimension. An example is *academic achievement*. The attributes *high*, *medium*, and *low* represent a level of measurement describing a variable that has equal distances between adjacent attributes. The Fahrenheit temperature scale is an example of this, because the distance between 17 and 18 is the same as that between 89 and 90.

does have meaning. Such variables are *interval measures*. For these, the logical distance between attributes can be expressed in meaningful standard intervals.

For example, in the Fahrenheit temperature scale, the difference, or distance, between 80 degrees and 90 degrees is the same as that between 40 degrees and 50 degrees. However, 80 degrees Fahrenheit is not twice as hot as 40 degrees, because the zero point in the Fahrenheit scale is arbitrary; zero degrees does not really mean lack of heat. Similarly, minus 30 degrees on this scale doesn't represent 30 degrees less than no heat. (This is true for the Celsius scale as well. In contrast, the Kelvin scale is based on an absolute zero, which does mean a complete lack of heat.)

Although the interval measures commonly used in social science research are constructed measures such as standardized intelligence tests that have been more or less accepted. The interval separating IQ scores of 100 and 110 may be regarded as the same as the interval separating scores of 110 and 120 by virtue of the distribution of observed scores obtained by many thousands of people who have taken the tests over the years. It would be incorrect to infer that someone with an IQ of 150 is 50 percent more intelligent than someone with an IQ of 100. (A person who has a score of 100 on a standard IQ test could just be regarded, strictly speaking, as having no intelligence, although we might feel he or she was unimpaired to be a college professor or even a college student. But perhaps a dean . . .)

When comparing two people in terms of an interval measure, we can say they are more than each other (nominal), and that one is more than the other (ordinal). In addition, we can say "how much" more.

Ratio Measures

Most of the social science variables meeting the minimum requirements for interval measures also meet the requirements for ratio measures. In ratio measures, the attributes composing the variable have all the structural characteristics mentioned previously, are based on a true zero point.

The Kelvin temperature scale is one such measure. Examples from social science research include *age*, *length of residence in a given place*, *number of paintings owned*, *number of times attending religious services during a particular period of time*, *number of times married*, and *number of kids found*.

Returning to the illustration of methodological party games, we might ask a gathering of people to group themselves by age. All the one-year-olds would stand (or sit) in the together, the two-year-olds together, the three-year-olds, and so forth. The fact that members of a single group share the same age is not sufficient to make that group a different shared age satisfies the minimum requirements for a nominal measure. Arranging the several groups in a line from youngest to oldest meets the additional requirements of an ordinal measure and lets us determine if one person is older than, younger than, or the same age as another. If we space the groups equally far apart, we satisfy the additional requirements of an interval measure and can say how much older one person is than another. Finally, because one of the attributes included in age represents a true zero (babies carried by women about to give birth), the phalanx of babies party games also meets the requirements of a ratio measure, permitting us to say that one person is twice as old as another. (Remember this in case you're asked about it in a workbook assignment.)

Another example of a ratio measure is *income*, which extends from an absolute zero to approximately infinity. If you happen to be the founder of Microsoft, you'll have a very high level of income.

Comparing two people in terms of a ratio variable, then, allows us to conclude (1) whether they are different (in the same way that we can say one is more than the other), (2) how much they differ, and (3) what the ratio of one to another is. Figure 5-1 summarizes this discussion by presenting a brief illustration of the four levels of measurement.

Implications of Levels of Measurement

Because it's unlikely that you'll undertake the physical grouping of people just described try it, once, and you won't be invited to many parties.

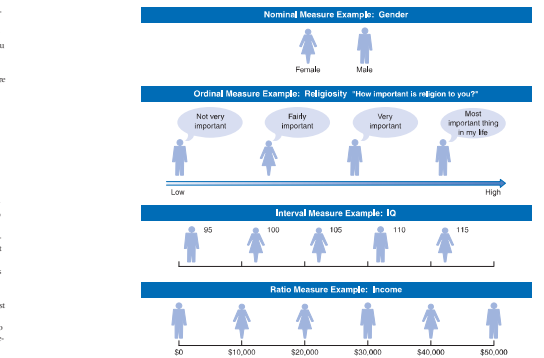


FIGURE 5-1 Levels of Measurement. Often you can choose among different levels of measurement—nominal, ordinal, interval, or ratio—depending on the amount of information.

age might be used as a nominal-level variable for certain research purposes. People might be grouped as being born during the Depression or not. Another nominal measurement, based on birth date rather than just age, would be the ordinal or grouping of people by astrological signs.

The level of measurement you'll seek, then, is determined by the analysis uses you've planned for a given variable, keeping in mind that some variables are inherently limited to a certain level. If a variable is to be used in a variety of ways, requiring

a study of the wealth of corporations, rather than seek more precise information, you may use Dun & Bradstreet ratings to rank corporations. Whenever your research purposes are not altogether clear, however, seek the highest level of measurement possible. As we've discussed, although ratio measures can later be reduced to ordinal ones, you cannot convert an ordinal measure to a ratio one. Most generally, you cannot convert a lower-level measure to a higher-level one. That is a one-way street worth remembering.

Typically a research project will tap variables at different levels of measurement. For example, William Billings and Denise Billig (1999) set out to examine the world of film and television, using a nonmetric, longitudinal approach to take a moment to remind yourself what that means. In what they referred to as the "culture industry" the authors found that reputation (an ordinal variable) is the best predictor of screenwriters' future productivity. More interestingly, they found that screenwriters who were represented by "core" (or elite) agencies were not only far more likely to find jobs (a nominal variable), but also jobs that paid more (a ratio variable). In other words, the researchers found that agencies' reputations (ordinal) was a key independent variable for predicting a screenwriter's career success. The researchers also found that being older (ratio), female (nominal), an ethnic minority (nominal), and having more years of experience were disadvantageous for a writer's career. On the other hand, higher earnings from previous years (measured in several categories) led to more success in the future. In Billings and Billig's terms, "success breeds success" (1999: 8).

Single or Multiple Indicators

With so many alternatives for operationalizing social science variables, you may find yourself worrying about making the right decision. To counter this feeling, let us add a momentary dash of certainty and stability.

Many social science variables have fairly obvious, straightforward measures. No matter how you

cut it, gender usually turns out to be a matter of male or female: a nominal-level variable that can be measured by a single observation—either by looking (well, not always) or by asking a question (usually). In a study involving the size of the embryo, it's wise to think about whether you have foster children, as well as blended families, but it's usually pretty easy to find out how many children a family has. For most research purposes, the resident population of a country is the resident population of that country—you can look it up in an almanac and know the answer. A great many variables, however, have observational indicators. In these cases, you will want to make several observations for a given variable. You can then combine the several pieces of information you've collected, creating a composite measurement of the variable in question. Chapter 6 is devoted to ways of doing that, so here let's just discuss one simple illustration.

Consider the concept "college performance." All of us have noticed that some students perform well in college courses and others don't. In studying these differences, we might ask what characteristics and experiences are related to high levels of performance. In many research situations (see just ahead), how should we measure overall performance? Each grade in any single course is a potential indicator of college performance; it also may not apply to the student's general performance. The solution to this problem is to formally establish that it is, of course, obvious: the grade point average. We assign numerical scores to each letter grade, total the points earned by a given student, and divide by the number of courses that compose a variable. The description of a woman as "34 years old" is more precise than "in her forties." Saying a street-corner game was "formed" in the summer of 1996 is more precise than saying "during the 1990s."

As a general rule, precise measurements are superior to imprecise ones, as common sense dictates. There are no conditions under which imprecise measurements are intrinsically superior to precise ones. Even so, exact precision is not always necessary or desirable. If knowing that a woman is in her forties satisfies your research requirements, then an additional effort invested in learning her precise age is wasted. If you're interested in whether you need to be guided party by an understanding of the degree of precision required. If your needs are not clear, be more precise rather than less.

Don't confuse precision with accuracy. However, describing someone as "born in New England" is less precise than "born in Boston or Vermont"—but those the person in question was actually born in Boston. The less-precise description, in this instance, is more accurate, a better reflection of the real world.

Precision and accuracy are obviously important qualities in research measurement, and they probably need no further explanation. When social scientists construct and evaluate measurements, however, they pay special attention to two technical considerations: reliability and validity.

Reliability

In the abstract, **reliability** is a matter of whether a particular technique, applied repeatedly to the same object, yields the same result each time. Let's say you want to know how much I weigh. (No, I don't know why.) As one technique, say you ask two different people to estimate my weight. If the first person estimates 150 pounds and the other

cut it, gender usually turns out to be a matter of male or female: a nominal-level variable that can be measured by a single observation—either by looking (well, not always) or by asking a question (usually). In a study involving the size of the embryo, it's wise to think about whether you have foster children, as well as blended families, but it's usually pretty easy to find out how many children a family has. For most research purposes, the resident population of a country is the resident population of that country—you can look it up in an almanac and know the answer. A great many variables, however, have observational indicators. In these cases, you will want to make several observations for a given variable. You can then combine the several pieces of information you've collected, creating a composite measurement of the variable in question. Chapter 6 is devoted to ways of doing that, so here let's just discuss one simple illustration.

Consider the concept "college performance." All of us have noticed that some students perform well in college courses and others don't. In studying these differences, we might ask what characteristics and experiences are related to high levels of performance. In many research situations (see just ahead), how should we measure overall performance? Each grade in any single course is a potential indicator of college performance; it also may not apply to the student's general performance. The solution to this problem is to formally establish that it is, of course, obvious: the grade point average. We assign numerical scores to each letter grade, total the points earned by a given student, and divide by the number of courses that compose a variable. The description of a woman as "34 years old" is more precise than "in her forties." Saying a street-corner game was "formed" in the summer of 1996 is more precise than saying "during the 1990s."

As a general rule, precise measurements are superior to imprecise ones, as common sense dictates. There are no conditions under which imprecise measurements are intrinsically superior to precise ones. Even so, exact precision is not always necessary or desirable. If knowing that a woman is in her forties satisfies your research requirements, then an additional effort invested in learning her precise age is wasted. If you're interested in whether you need to be guided party by an understanding of the degree of precision required. If your needs are not clear, be more precise rather than less.

Don't confuse precision with accuracy. However, describing someone as "born in New England" is less precise than "born in Boston or Vermont"—but those the person in question was actually born in Boston. The less-precise description, in this instance, is more accurate, a better reflection of the real world.

Precision and accuracy are obviously important qualities in research measurement, and they probably need no further explanation. When social scientists construct and evaluate measurements, however, they pay special attention to two technical considerations: reliability and validity.

Reliability

In the abstract, **reliability** is a matter of whether a particular technique, applied repeatedly to the same object, yields the same result each time. Let's say you want to know how much I weigh. (No, I don't know why.) As one technique, say you ask two different people to estimate my weight. If the first person estimates 150 pounds and the other

Some Illustrations of Operationalization Choices

To bring together all the operationalization choices available to the social researcher and to show the potential in those possibilities, let's look at some of the distinct ways you might address various research problems. The alternative ways of operationalizing variables in each case should demonstrate the opportunities that social research can present to our ingenuity and imaginations. To simplify matters, I have not attempted to describe all the research conditions that would make one alternative superior to the others, though in a given situation they would not all be equally appropriate.

Here are specific research questions, then, and some of the ways you could address them. We'll begin with an example discussed earlier in the chapter. It has the added advantage that one of the variables is straightforward to operationalize.

- Are women more compassionate than men?
 - Select a group of subjects for study, with equal numbers of men and women. Present them with hypothetical situations that involve someone's being in trouble. Ask them what they would do if they were confronted with that situation. What would they do, for example, if they came across a small child who was lost and crying for his or her parents? Consider any answer that involves helping or comforting the child as an act of compassion. See whether the men or women are more likely to indicate they would be compassionate.
 - Set up an experiment in which you pay a small child to pretend that he or she is lost. Put the child to work on a busy sidewalk and observe whether men or women are more likely to offer assistance. Also be sure to count the total number of men and women who walk by, because there may be more of one than the other. In such a case, simply calculate the percentage of men and the percentage of women who help.
 - Select a sample of people and do a survey in which you ask them what organizations

they belong to. Calculate whether women or men are more likely to belong to those that seem to reflect compassionate feelings. To account for the size in which the group belongs to more organizations than the other does, do this for each person you study, calculate the percentage of his or her organizational memberships that reflect compassion. See if men or women have a better informed average world affairs?

- Prepare a short quiz on world affairs and arrange to administer it to the students in a sociology class and in an accounting class at a comparable level. If you want to compare sociology and accounting majors, you may wish to ask students what they are majoring in.
- Get the instructor of a course in world affairs to give you the average grades of sociology and accounting students in the course.
- Take a petition to sociology and accounting classes that urges that "the United Nations headquarters be moved to New York City" and see how many signatures you can get. Sign the petition and how many in form you that the UN headquarters is already located in New York City.

- Do people consider New York or California the better place to live?
 - Consulting the *Statistical Abstract of the United States* or a similar public source, see whether migration rates into and out of each state. See if you can find the numbers moving directly into New York or California and vice versa.
 - The national polling companies—Gallup, Harris, Roper, and so forth—often ask people if they would like to live in New York or California. Look up their recent results in the library or through your local newspaper.
 - Compare suicide rates in the two states.

- Who are the most popular instructors on your campus, those in the social sciences, the natural sciences, or the humanities?
 - Design a questionnaire that asks you to

Measuring College Satisfaction

Query in this chapter, we considered college satisfaction as an example of a concept people often talk about casually. (Usually such a concept, however, we need to engage in the process of operationalization and operationalization.) I think that the process itself, though, you might try your hand at depending on my comments.

What are some of the dimensions of college satisfaction? Here are a few you could, but be sure to add your own:

- Academic quality: faculty, courses, majors
- Physical facilities: classrooms, dorms, cafeteria, grounds
- Activities and extracurricular activities
- Cost and availability of financial aid
- Sociality of students, faculty, staff
- Security, crime on campus

- If your school has a provision for student evaluation of instructors, review some recent results and compute the average rating of each of the three groups.
- Begin visiting the introductory courses given in each group of disciplines and measure the attendance rate of each class.
- In December, select a group of faculty in each of the three divisions and ask them to keep a record of the numbers of holiday greeting cards and presents they receive from admiring students. See who wins.

The point of these examples is not necessarily to suggest respectable research projects but to illustrate the many ways variables can be operationalized.

"Measuring College Satisfaction" briefly overviews the preceding steps in terms of a concept mentioned at the outset of this chapter.

Operationalization Goes On and On

Although I've discussed conceptualization and operationalization as activities that precede data collection and analysis—for example, you must design questionnaire items before you send out

Measuring College Satisfaction

Query in this chapter, we considered college satisfaction as an example of a concept people often talk about casually. (Usually such a concept, however, we need to engage in the process of operationalization and operationalization.) I think that the process itself, though, you might try your hand at depending on my comments.

What are some of the dimensions of college satisfaction? Here are a few you could, but be sure to add your own:

- Academic quality: faculty, courses, majors
- Physical facilities: classrooms, dorms, cafeteria, grounds
- Activities and extracurricular activities
- Cost and availability of financial aid
- Sociality of students, faculty, staff
- Security, crime on campus

- If your school has a provision for student evaluation of instructors, review some recent results and compute the average rating of each of the three groups.
- Begin visiting the introductory courses given in each group of disciplines and measure the attendance rate of each class.
- In December, select a group of faculty in each of the three divisions and ask them to keep a record of the numbers of holiday greeting cards and presents they receive from admiring students. See who wins.

The point of these examples is not necessarily to suggest respectable research projects but to illustrate the many ways variables can be operationalized.

"Measuring College Satisfaction" briefly overviews the preceding steps in terms of a concept mentioned at the outset of this chapter.

Operationalization Goes On and On

Although I've discussed conceptualization and operationalization as activities that precede data collection and analysis—for example, you must design questionnaire items before you send out

Measuring College Satisfaction

Query in this chapter, we considered college satisfaction as an example of a concept people often talk about casually. (Usually such a concept, however, we need to engage in the process of operationalization and operationalization.) I think that the process itself, though, you might try your hand at depending on my comments.

What are some of the dimensions of college satisfaction? Here are a few you could, but be sure to add your own:

- Academic quality: faculty, courses, majors
- Physical facilities: classrooms, dorms, cafeteria, grounds
- Activities and extracurricular activities
- Cost and availability of financial aid
- Sociality of students, faculty, staff
- Security, crime on campus

- If your school has a provision for student evaluation of instructors, review some recent results and compute the average rating of each of the three groups.
- Begin visiting the introductory courses given in each group of disciplines and measure the attendance rate of each class.
- In December, select a group of faculty in each of the three divisions and ask them to keep a record of the numbers of holiday greeting cards and presents they receive from admiring students. See who wins.

The point of these examples is not necessarily to suggest respectable research projects but to illustrate the many ways variables can be operationalized.

"Measuring College Satisfaction" briefly overviews the preceding steps in terms of a concept mentioned at the outset of this chapter.

Operationalization Goes On and On

Although I've discussed conceptualization and operationalization as activities that precede data collection and analysis—for example, you must design questionnaire items before you send out

Measuring College Satisfaction

Query in this chapter, we considered college satisfaction as an example of a concept people often talk about casually. (Usually such a concept, however, we need to engage in the process of operationalization and operationalization.) I think that the process itself, though, you might try your hand at depending on my comments.

What are some of the dimensions of college satisfaction? Here are a few you could, but be sure to add your own:

- Academic quality: faculty, courses, majors
- Physical facilities: classrooms, dorms, cafeteria, grounds
- Activities and extracurricular activities
- Cost and availability of financial aid
- Sociality of students, faculty, staff
- Security, crime on campus

- If your school has a provision for student evaluation of instructors, review some recent results and compute the average rating of each of the three groups.
- Begin visiting the introductory courses given in each group of disciplines and measure the attendance rate of each class.
- In December, select a group of faculty in each of the three divisions and ask them to keep a record of the numbers of holiday greeting cards and presents they receive from admiring students. See who wins.

The point of these examples is not necessarily to suggest respectable research projects but to illustrate the many ways variables can be operationalized.

"Measuring College Satisfaction" briefly overviews the preceding steps in terms of a concept mentioned at the outset of this chapter.

Operationalization Goes On and On

Although I've discussed conceptualization and operationalization as activities that precede data collection and analysis—for example, you must design questionnaire items before you send out

estimates 300, we have to conclude that the technique of having people estimate my weight isn't very reliable.

Suppose, as an alternative, that you use a bathroom scale as your measurement technique. I step on the scale twice, and you note the same result each time. The scale has presumably reported the same weight for me both times, indicating that the scale provides a more reliable technique for measuring a person's weight than asking people to estimate it does.

Reliability, however, does not ensure accuracy. More than precision does. Suppose I've set my bathroom scale to show five pounds off my weight just to make me feel better. Although you would (reliably) report the same weight for me each time, you would always be wrong. This new element, called bias, is discussed in Chapter 8. For now, we just want to warn that reliability does not ensure accuracy.

Let's suppose we're interested in studying morale among factory workers in two different kinds of factories. In one set of factories, workers have specialized jobs, reflecting an extreme division of labor. Each worker contributes a tiny part to the overall process performed on a long assembly line. In the other set of factories, each worker performs many tasks, and all teams of workers complete the whole process.

How should we measure morale? Following one strategy, we could observe the workers in each factory, noting such things as how they joke with one another, whether they smile and laugh a lot, and so forth. We could ask them whether they like their work and even ask them whether they think they would rather work in the other factory than the other one being studied. By comparing what we observed in the different factories, we might conclude that morale is higher in the factory with more specialized jobs. Notice that I've just described a qualitative measurement procedure.

Now let's look at some reliability problems inherent in this method. First, how many times do we feel when we do the observing will likely occur, and how many times will we miss it? We may see workers kidding each other but they're having an argument. We may catch them on an off day. If we were to observe the same group

of workers several days in a row, we might arrive at different evaluations on each day. Further, even if several observers evaluated the same behavior, they might arrive at different conclusions about the workers' morale.

Here's another strategy for assessing morale, a quantitative approach. Suppose we check the company records to see how many grievances have been filed with the union during some fixed period. Presumably this would be an indicator of morale, the more grievances, the lower the morale. This measurement strategy would appear to be more reliable: Counting up the grievances over and over, we should keep arriving at the same number. If you find yourself thinking that the number of grievances doesn't necessarily measure morale, you're worrying about validity, not reliability. We'll discuss validity in a moment; the point now is that the last method is more like my bathroom scale—it gives consistent results.

In social research, reliability problems crop up in many forms. Reliability is a concern every time a single observer is the source of data, because we have no certain guard against the impact of that observer's subjectivity. We can't tell for sure how much of what's reported originated in the situation observed and how much in the observer.

Subjectivity is not only a problem with single observers, however. Survey researchers have known for a long time that in different interviews, because of their own attitudes and demons, get different answers from respondents. One way to conduct a study of newspapers' editorial positions on some public issue, we might create a team of coders to take on the job of reading and coding of editorials and classifying them in terms of their position on the issue. Unfortunately, different coders will make different editorial decisions, or we might want to classify a few hundred specific occupations in terms of some standard coding scheme, say a set of categories created by the Department of Labor or by the Census Bureau. You and I would not place all those occupations in the same categories.

To test the reliability of their information, Sacks and his colleagues had all 207 subjects complete a baseline questionnaire that asked about their

Sometimes we ask questions that people don't know the answers to: How many times have you been to religious services? Sometimes we ask people about things they consider totally irrelevant: Are you satisfied with China's current relationship with Albania? In such cases, people will answer differently at different times because they're making up answers as they go. Sometimes we explore issues so complicated that a person who had a clear opinion in the matter might arrive at a different interpretation of the question when asked a second time.

So how do you create reliable measures? If your research design calls for asking people for information, you can be careful to ask only about things the respondents are likely to know the answer to. Ask about things relevant to them, and be clear in what you're asking. Of course, these techniques don't solve every possible reliability problem. Fortunately, social researchers have developed several techniques for cross-checking the reliability of the measures they devise.

Test-Retest Method

Sometimes it's appropriate to make the same measurement more than once, a technique called the test-retest method. If you don't expect the sought-after information to change, then you should expect the same response both times. If answers vary, the measurement method may be so erratic that the variation is unacceptable. Here's an Illustration.

In their research on Health Hazard Appraisal (HHA), a part of preventive medicine, Jeffrey Sacks, Mark Krueber, and Jeffrey Newman (1980) wanted to determine the risks associated with various background and lifestyle factors, making it possible for physicians to counsel their patients appropriately. By knowing patients' life situations, physicians could advise them on their potential for survival and on how to improve it. This purpose, of course, depended heavily on the accuracy of the information gathered about each individual in the study.

To test the reliability of their information, Sacks and his colleagues had all 207 subjects complete a baseline questionnaire that asked about their

characteristics and behavior. Three months later, a follow-up questionnaire asked the same subjects for the same information, and the results of the two surveys were compared. Overall, only 15 percent of the subjects reported the same information in both studies.

Sacks and his colleagues report the following: Almost 10 percent of subjects reported a different height at follow-up examination. Paternal age was changed by over one in three subjects. One parent reportedly aged 20 chronological years in three months. One in five ex-smokers and ex-drinkers have apparently difficulty in reliably recalling their previous consumption pattern.

(1980: 730)

Some heart enamel, diabetes, epilepsy, arteriosclerosis, and thoughts of suicide. One subject, deceased in the first questionnaire, was apparently alive and well in time for the second. One subject had one ovary missing in the first study but present in the second. In another case, an ovary present in the first study was missing in the second study—and had been for ten years! One subject was reportedly 95 years old in the first study and 50 years old three months later. (You have to wonder whether the physician-sonneteer could ever have nearly in his parents that their parents' memories did.) This, test-retest revealed that this data-collection method was not especially reliable.

Split-Half Method

As a general rule, it's always good to make more than one measurement of any subject or complex social concept, such as prejudice, alienation, or social class. This procedure lays the groundwork for another check on reliability. Let's say you've created a questionnaire that contains items you believe measure prejudice against women. Using the split-half technique, you would randomly assign those items to two sets of five. Each set should provide a good measure of prejudice against women, and the two sets should classify respondents the

same way. If the two sets of items classify people differently, you may suspect a problem of reliability in your measure of the variable.

Using Established Measures

Another way to help ensure reliability in getting information from people is to use measures that have proved their reliability in previous research. If you want to measure anomia, for example, you might want to follow Srole's lead.

The heavy use of measures, though, does not guarantee their reliability. For example, the Scholastic Assessment Tests (SATs) and the Minnesota Multiphasic Personality Inventory (MMPI) have been accepted as established standards in their respective domains for decades. In recent years, though, they've needed fundamental overhauls to reflect changes in society, eliminating outdated topics and gender bias in wording.

Reliability of Research Workers

As we've seen, it's also possible for measurement unreliability to be generated by research workers: interviewers and coders, for example. There are several ways to check on reliability in such cases. To guard against interviewer unreliability in surveys, for example, a supervisor will call a subsample of the respondents on the telephone and verify selected pieces of information.

Replication works in other situations, too. If you're worried that newspaper editors or reporters may not be classified reliably, you could have each independently coded by several coders. Those who are classified consistently can then be evaluated more carefully and resolved.

Finally, clarity, specificity, training and practice can prevent a great deal of unreliability and grief. If you and I spend some time reaching a clear agreement on how to evaluate editorial positions on an issue—discussing various positions and reading through several together—we could probably do a good job of classifying them in the same way independently.

The reliability of measurements is a fundamental issue in social research, and we'll return to it more than once in the chapters ahead. For

now, however, it's recall that even total reliability doesn't ensure that our measures actually measure what we think they measure. Now let's plunge into the question of validity.

Validity

In conventional usage, **validity** refers to the extent to which an empirical measure adequately reflects the real meaning of the concept under consideration. A measure of social class should measure social class, not political orientations. A measure of political orientations should measure political orientations, not sexual permissiveness. Validity means that we are actually measuring what we say we are measuring.

Whom? I've already committed us to the view that concepts don't have real meanings. How can we ever say whether a particular measure adequately reflects the concept's meaning, then?

Ultimately, of course, we can't. At the same time, as we've already seen, all of social life, including social research, operates on agreements about the terms we use and the concepts they represent. There are several criteria of success in making measurements that are appropriate to these agreed-on meanings of concepts.

First, there's something called **face validity**. Particular empirical measures may or may not jibe with our common agreements and our individual

validity A term describing a measure that accurately reflects the concept it is intended to measure. For example, your IQ would seem a more valid measure of your intelligence than the number of face validity, criterion-related validity, construct validity, internal validity, and external validity (see Chapter 6). This must not be confused with reliability.

face validity That quality of an indicator that makes it seem a reasonable measure of what it measures. That the frequency of attendance at religious services is a more valid indicator of religiosity seems to make sense without a lot of explanation. It's face validity.

mental images concerning a particular concept. For example, you and I might quarrel about whether counting the number of grievances filed with the union will adequately measure morale. Still, we'd surely agree that the number of grievances has something to do with morale. That is, the measure is valid "on its face," whether or not it's adequate. If I were to suggest that we measure morale by finding out how many books the workers took out of the library during their off-duty hours, you'd undoubtedly raise a more serious objection: That measure wouldn't have much face validity.

Second, I've also pointed to many of the more formally established agreements that define some concepts. The Census Bureau, for example, has created operational definitions of such concepts as family, household, and employment status that seem to have a workable validity in most studies using these concepts.

Three additional types of validity also specify particular ways of testing the validity of measures. The first, **criterion-related validity**, sometimes called **concurrent validity**, is based on some external criterion. For example, the validity of College Board exams is shown in their ability to predict success in college. The validity of a written driver's test is determined, in this sense, by the relationship between the scores people get on the test and their subsequent driving records. In these examples, criterion-related and driving ability are the criteria.

To test your understanding of criterion-related validity, think about how you can think of behavior that might be used to validate each of the following attitudes:

Is very religious
Supports equality of men and women

criterion-related validity The degree to which a measure relates to some external criterion. For example, the validity of College Board tests is shown in their ability to predict the college success of students. Also called *predictive validity*.

construct validity The degree to which a measure measures other variables as expected within a system of theoretical relationships.

Supports fair militia rights groups

Is concerned about the environment
Some possible validations would be, respectively, attends religious services, votes for women candidates, belongs to the NRA, and belongs to the Sierra Club.

Sometimes it's difficult to find behavioral criteria that can be taken to validate measures as directly as in such examples. In those instances, however, we can often approximate such criteria by applying a different test. We can consider how the variable in question relates to other variables. **Construct validity** is based on the logical relationships among variables.

Suppose, for example, that you want to study the sources and consequences of marital satisfaction. As part of your research, you develop a measure of marital satisfaction, and you want to assess its validity.

In addition to developing your measure, you'll have developed certain theoretical expectations about the way the variable marital satisfaction relates to other variables. For example, you might reasonably conclude that satisfied husbands and wives will be more likely than dissatisfied ones to cheer on their spouses. If your measure relates to marital fidelity in the expected fashion, that constitutes evidence of your measure's construct validity. If satisfied marriage partners are more likely to cheer on their spouses as the dissatisfied ones are, however, that would challenge the validity of your measure.

Tests of construct validity, then, can offer a weight of evidence that your measure either does or does not measure the construct you're interested in without providing definitive proof. Although I have suggested that tests of construct validity are less compelling than those of criterion validity, there is room for disagreement about which kind of test is a particular comparison variable. *Driving record, marital fidelity*) represents in a given situation. It's less important to distinguish the variables and reliability tests than to understand the logic of validation that they have in common. If we're successful in measuring some variable as expected within a system of theoretical relationships, we've made a valid measure.

tests to varied meanings. Suppose, for example, that you are interested in sampling public opinion on the abortion issue in the United States. Notice the difference it would make if you conceptualized one side of the debate as "pro-choice" or as "pro-abortion." If your personal bias made you want to minimize support for having an abortion, you might be tempted to frame the concept and the measurements based on it in terms of people being "pro-abortion," thereby eliminating all those who were not especially kind of abortion per se but felt a woman should have the right to make that choice for herself. To pursue this strategy, however, would violate accepted research ethics.

Consider the choices available to you in conceptualizing attitudes toward the U.S. invasion of Iraq in 2003. Imagine the different levels of support you would "discover" if you framed the position as an unprovoked invasion of a sovereign nation, as a retaliation for the September 11, 2001, attack on the World Trade Towers (many Americans will believe Saddam Hussein masterminded that attack), as a defensive act against a perceived threat, as part of a global war on terrorism, or in any of the other ways this issue has been portrayed. There is no one, correct way to operationalize this issue, but it would be unethical to seek to slant the results through a biased definition of the issue.

MAIN POINTS

Introduction

- The interrelated processes of conceptualization, operationalization, and measurement allow researchers to move from a general idea about what they want to study to effective and well-defined measurements in the real world.

Measuring Anything That Exists

- Concepts are mental images we use as summary devices for bringing together observations and experiences that we want to something in common. We use terms or labels to represent these concepts.
- Concepts are constructs that reference the agreed-on meanings we assign to terms. Our

concepts don't exist in the real world, so they can't be measured directly, but we can measure the things that our concepts summarize.

Conceptualization

- Conceptualization is the process of specifying what you mean by a concept and giving it a definite meaning for the purposes of a research study.

- Conceptualization includes specifying the indicators of a concept and describing its dimensions. Operational definitions specify how variables relevant to a concept will be measured.

Definitions in Descriptive and Explanatory Studies

- Precise definitions are even more important in descriptive than in explanatory studies. The degree of precision needed varies with the type and purpose of a study.

Operationalization Choices

- Operationalization is an extension of conceptualization that specifies the exact procedures that will be used to measure the attributes of variables.
- Operationalization involves a series of interrelated choices: specifying the range of variation that is appropriate for the purposes of the study, determining how precisely to measure variables, accounting for relevant dimensions of variables, clearly defining the attributes of variables and their relationships, and deciding on an appropriate level of measurement.
- Researchers must choose from four levels of measurement, which capture increasing amounts of information: nominal, ordinal, interval, and ratio. The most appropriate level depends on the purpose of the measurement.
- A given variable can sometimes be measured at different levels. When, in doubt, researchers should use the highest level of measurement appropriate that a variable so they can capture the greatest amount of information.
- Operationalization begins in the design phase of a study and continues throughout the phases of data collection and analysis.

Criteria of Measurement Quality

- Criteria of the quality of measures include precision, accuracy, reliability, and validity.

- Whereas reliability means getting consistent results from the same measure, validity refers to getting results that accurately reflect the concept being measured.
- Researchers can test to improve the reliability of measures through the test-retest method, the split-half method, the use of established measures, and the examination of work performed by research workers.
- The preconditions for assessing a measure's validity include face validity, criterion-related validity, construct validity, and content validity.
- Creating specific, reliable measures often seems to diminish the richness of measuring our general concepts have. This problem is inevitable. The best solution is to use several different measures, tapping the different aspects of a concept.

The Ethics of Measurement

- Conceptualization and measurement must never be guided by bias or preferences for particular research outcomes.

KEY TERMS

The following terms are defined in context in the chapter and at the bottom of the page where the term is introduced, as well as in the comprehensive glossary at the back of the book.

conceptualization
construct validity
content validity
criterion-related validity
dimension
face validity
indicator
interval measure
nominal measure
ordinal measure
ratio measure
reliability
specification
validity

PROPOSING SOCIAL RESEARCH: MEASUREMENT

This chapter has taken us deeper into the matter of measurement. In previous exercises, you've identified the concepts and variables you want to address in your research. Now you'll need to get more specific in terms of conceptualization and operationalization. You should conclude this portion of the proposal with a description of how, precisely, you will make distinctions regarding your variables. If you want to compare liberal and conservative, for example, how exactly will you identify subjects' political orientations?

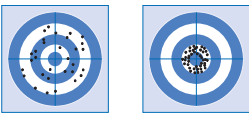


FIGURE 5-2 An analogy to validity and reliability. A good measurement technique should be both valid (measuring what it is intended to measure) and reliable (yielding a consistent measurement dependably).

Reliable but not valid

It is analog to validity and reliability. A good measurement technique should be both valid (measuring what it is intended to measure) and reliable (yielding a consistent measurement dependably).

Valid but not reliable

It is analog to validity and reliability. A good measurement technique should be both valid (measuring what it is intended to measure) and reliable (yielding a consistent measurement dependably).

Valid and reliable

It is analog to validity and reliability. A good measurement technique should be both valid (measuring what it is intended to measure) and reliable (yielding a consistent measurement dependably).

Finally, **content validity** refers to how much a measure covers the range of meanings included within a concept. For example, a test of mathematical ability cannot be limited to addition but also needs to cover subtraction, multiplication, division, and so forth. Or, if we're testing for prejudice, do our measurements reflect all types of prejudice, including prejudice against racial and ethnic groups, religious minorities, women, the elderly, and so on?

Figure 5-2 presents a graphic portrayal of the difference between validity and reliability. If you think of measurement as analogous to repeatedly shooting at the bull's-eye on a target, you'll see that reliability looks like a "tight pattern," regardless of where the shots land, because reliability is a function of consistency. Validity, on the other hand, is a function of shots being arranged around the bull's-eye. The failure of reliability in the figure is randomly distributed around the target; the failure of validity is systematically off the mark. Notice that neither an unreliable nor an invalid measure is likely to be very useful.

Who Decides Validity?

Our discussion of validity began with a reminder that we depend on agreements to determine what's real, and we've just seen some of the ways social scientists can agree among themselves that they have made valid measurements. There is yet another way of looking at validity.

content validity The degree to which a measure covers the range of meanings included within a concept.

The ease or difficulty of this exercise may vary with the type of data collection you're planning. It will probably be easier to accomplish in the case of quantitative studies, such as surveys, where you can report the questionnaire items you use for measurements. In qualitative research, however, you'll have more opportunities to modify the way variables are measured as the study unfolds, taking advantage of insights gained "in the trenches." Even so, you'll still need to begin with some clear ideas about how you'll begin your measurements.

Criteria such as precision, accuracy, validity, and reliability matter greatly in all kinds of social research projects.

REVIEW QUESTIONS AND EXERCISES

1. Pick a social science concept such as liberalism or alienation, then specify that concept so that it could be studied in a research project. Be sure to specify the indicators you'll use as well as the dimensions you wish to include and to exclude from your conceptualization.

2. What level of measurement—nominal, ordinal, interval, or ratio—describes each of the following variables?

a. Race (white, African American, Asian, and so on)

b. Order of finish in a race (first, second, third, and so on)

c. Number of children in families

d. Populations of nations

e. Attitudes toward nuclear energy (totally approve, approve, disagree, strongly disagree)

f. Region of birth (Northeast, Midwest, and so on)

g. Political orientation (very liberal, somewhat liberal, somewhat conservative, very conservative)

3. To operationalize the variable *pragmatic*, use your favorite web browser to search for this term. After reviewing several of the websites resulting from your search, make a list of some different forms of pragmatism that might be studied in an omnibus project dealing with that topic.

4. In a good dictionary, look up *truth* and *false*, then copy out the definitions. Note the key terms used in those definitions (such as *realism*) and copy out the definitions of those terms, and copy out these

of a culture for those who live outside that culture. He recognizes that the livable observer will never grasp those meanings fully, however, and warns, "Cultural analysis is intrinsically incomplete." He then elaborates:

There are a number of ways to escape this—turning culture into folklore and collecting it, turning it into traits and cataloging it, turning it into institutions and classifying it, turning it into structures and trying with it. But they are escapes. The fact is that to commit oneself to a semantic concept of culture and an interpretive approach to the study of it is to commit oneself to a view of ethnographic assertion as, to borrow W. B. Gallie's very own famous phrase, "essentially contestable." Anthropology, or at least interpretive anthropology, is a science whose progress is marked less by a perfection of consensus than by a refinement of debate. What goes better is the precision with which we use each other.

(1972: 29)

Ultimately, social researchers should look both to their colleagues and to their subjects as sources of agreement on the most useful meanings and measurements of the concepts they study. Sometimes one source will be more useful, sometimes the other. But neither one should be dismissed.

Tension between Reliability and Validity

Clearly, we want our measures to be both reliable and valid. However, a tension often arises between the criteria of reliability and validity, forcing a trade-off between the two.

Recall the example of measuring morale in different factories. The strategy of immersing yourself in the day-to-day routine of the assembly line, observing what goes on, and talking to the workers would seem to provide a more valid measure of morale than counting grievances would. It just seems obvious that we'd get a clearer sense of whether the morale was high or low using this first method.

As I pointed out earlier, however, the counting strategy would be more reliable. This situation reflects a more general tension in research measurement. Most of the really interesting concepts we want to study have many subtle nuances, so specifying precisely what we mean by them is hard. Researchers sometimes speak of such concepts as having a "richness of meaning." Although scores of books and articles have been written on the topic of anomia/anomie, for example, they still haven't exhausted its meaning.

Very often, then, specifying reliable operational definitions and measurements seems to rob concepts of their richness of meaning. Positive morale is much more than a lack of grievances and a sense of well-being. Anomie is more than what is measured by the five items created by Leo Srole. Yet, the more variation and richness we allow for a concept, the more opportunity there is for disagreement on how it applies to a particular situation, thus reducing reliability.

To some extent, this dilemma explains the persistence of more qualitative different approaches to research: quantitative, nonnumeric, structured techniques such as surveys and experiments on the one hand, and qualitative, ethnographic methods such as field research and historical studies on the other. In the simplest generalization, the former methods tend to be more reliable, the latter more valid.

By being forewarned, you'll be effectively forearmed against this persistent and inevitable dilemma. If there is no clear agreement on how to measure a concept, measure it several different ways. If the concept has several dimensions, measure them all. Above all, know that the concept does not have any meaning other than what you and I give it. The only justification for giving any concept a particular meaning is utility. Measure concepts in ways that help us understand the world around us.

The Ethics of Measurement

Measurement decisions can sometimes be judged by ethical standards. We have seen that most of the concepts of interest to social researchers are

find information on this online tool, as well as instructions on how to access it of its great resources, in the front of the book.

2. As you review, take advantage of the EngageNOW personalized study plan, based on your quiz results. Use this study plan with interactive exercises and other resources to master the material.

3. When you've finished with your review, take the posttest to confirm that you're truly up to move on to the next chapter.

SPSS EXERCISES

See the booklet that accompanies your text for exercises using SPSS (Statistical Package for the Social Sciences). There are exercises offered for each chapter, and you'll also find a detailed primer on using SPSS.

Online Study Resources

If your book came with an access code card, visit www.cengage.com/login/register. To purchase access, please visit www.cengage.com.

Go to your book's website at www.cengage.com/soc101 to find additional resources, including the EngageNOW personalized study plan, based on your quiz results. Use this study plan with interactive exercises and other resources to master the material.

3. When you've finished with your review, take the posttest to confirm that you're truly up to move on to the next chapter.

WEBSITE FOR THE PRACTICE OF SOCIAL RESEARCH 12TH EDITION

Go to your book's website at www.cengage.com/soc101 to find additional resources, including the EngageNOW personalized study plan, based on your quiz results. Use this study plan with interactive exercises and other resources to master the material.

3. When you've finished with your review, take the posttest to confirm that you're truly up to move on to the next chapter.