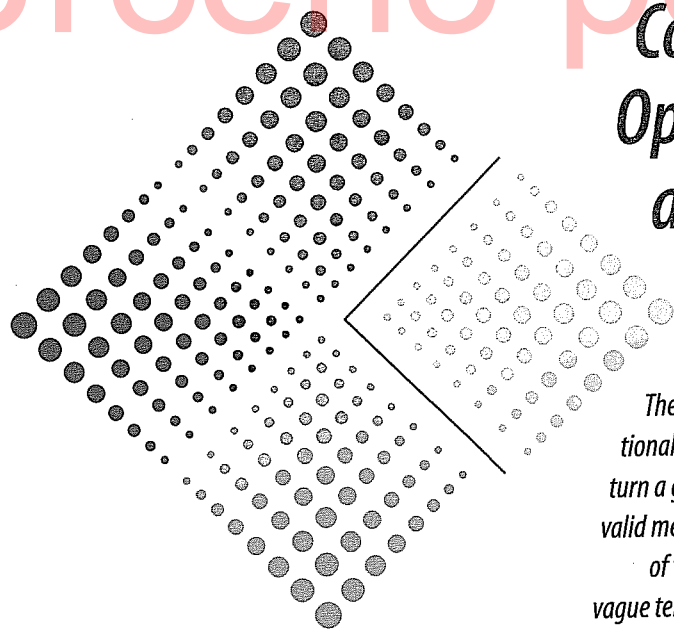


Conceptualization, Operationalization, and Measurement



Holographic Overview

The interrelated steps of conceptualization, operationalization, and measurement allow researchers to turn a general idea for 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

Conceptions, Concepts, and Reality
Conceptions 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 Extremes
A Note on Dimensions
Defining Variables and Attributes

Levels of Measurement

Single or Multiple Indicators
Some Illustrations of Operationalization
Choices
Operationalization Goes On and On

Criteria of Measurement Quality

Precision and Accuracy
Reliability
Validity
Who Decides What's Valid?
Tension between Reliability and Validity

MAIN POINTS

KEY TERMS

REVIEW QUESTIONS AND EXERCISES

ADDITIONAL READINGS

SOCIOLOGY WEB SITE

INFOTRAC COLLEGE EDITION

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.

We begin this chapter by confronting the hidden concern people sometimes have about whether it's truly possible to measure the stuff of life: love, hate, prejudice, religiosity, radicalism, alienation. The answer is yes, but it will take a few pages to see how. Once we establish 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 *measurement* instead, meaning careful, 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 research 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 religiosity as the number of times a person went to church, 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 might consult the list of registered voters to note whether 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 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; other times it simply means whatever religion, if any, you identify yourself with. Perhaps to you it means something else, such as church attendance.

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 have 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, we make statements about them, such as "In Happytown, 55 percent of the adults affiliate with the Republican Party, and 45 percent of them are Episcopalians. Overall, people in Happytown are low in prejudice and high in compassion." 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:

- We personally heard people say nasty things about minority groups.
- We heard people said 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. 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 agreed to go along with the system. That's where "prejudice" came from. We never observed it. We just agreed to use it as a shortcut, a name that represents 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 devel-

ops our own mental image of what the set of real phenomena we've observed represents in general and what these 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 as though file drawers in our minds contained thousands of sheets of paper, with each sheet of paper labeled in the upper right-hand corner. A sheet of paper in each of our minds has the term *prejudice* on it. On your sheet are all the things you've been told about prejudice and everything you've observed that seems to be an example of it. My sheet has what I've been told about it plus all the things I've observed that seem examples of it—and mine isn't the same as yours.

The technical term for those mental images, those sheets of paper in our mental file drawers, is *conception*. 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 will 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*.

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. Now suppose that you have seen Pat help lost children find their parents and put a tiny bird back in its nest. Pat got you to take turkeys to poor families on Thanksgiving and to visit a children's hospital on Christmas. You've seen Pat weep through a movie about a mother overcoming adversities 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 leaf 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 yours. 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 sure were 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 weep 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 just sentimentality."

To further strengthen my case, I tell you that I saw Pat refuse to give money to 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 about comparing the entries we have on our respective sheets labeled "compassionate." We then discover that we have many differing mental images corresponding to that term.

In the big picture, language and communication work only to 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. 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 file systems. If we want to measure "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, whether Pat actually puts the little bird back in its nest, visits the hospital on Christmas, weeps at the movie, or refuses to contribute to saving the whales. All of those behaviors exist, so we can measure them. But is Pat really compassionate? We can't answer that question; we can't measure compassion in any objective sense, because compassion doesn't exist the way 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 2, you'll recognize that some people would object to the degree of "reality" I've allowed in the preceding comments. Did Pat "really" visit the hospital on Christmas? Does the hospital "really" exist? Does Christmas? Though we aren't going to be radically postmodern 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 going 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 made in a questionnaire. The second class, indirect observables, require "relatively more subtle, complex, or indirect observations" (1964:55). We note a person's check mark beside "female" 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 constructs—theoretical creations 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.

Kaplan (1964:49) defines *concept* as a “family of conceptions.” A concept is, as Kaplan notes, a construct, something we create. Concepts like compassion and prejudice are constructs created from your conception of them, my conception of 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 seemingly related observations and experiences. Although the observations and experiences are real, at least subjectively, 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 like *prejudice* is, objectively speaking, only a collection of letters. It has no intrinsic reality beyond that. It has only the meaning we agree to give it.

Usually, 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. Further, the danger is all the greater in the presence of experts who appear to know more than we do about what the terms really mean: It’s easy to yield to authority in such a situation.

Once we assume that terms like *prejudice* and *compassion* have real meanings, we begin the tortured task of discovering what those real meanings are and what constitutes a genuine measurement of them. Regarding constructs as real is called *reification*. The reification of concepts in day-to-day life is quite common. In science, we want to be quite clear about what it is we are actually measuring.

Does this discussion imply that compassion, prejudice, and similar constructs can’t be measured? Interestingly, the answer is no. (And a good thing, too, or a lot of us social researcher types would be out of work.) I’ve said that we can measure any-

thing 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, while 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 occurs 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 *compassionate*, I’m probably safe in assuming that Pat won’t pull the wings off flies. A wide range of misunderstandings and conflict—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 meaningfully study the question, let alone agree on the answer, without some working agreements about the meaning of *compassion*. They are “working” agreements in the sense that they allow us to work on the question. We don’t need to agree or even pretend to agree that a particular specification is ultimately the best one.

Conceptualization, then, produces a specific, agreed-upon meaning for a concept for the purposes of research. This process of specifying exact meaning involves describing the indicators we’ll be

using to measure our concept and the different aspects of the concept, called dimensions.

Indicators and Dimensions

Conceptualization gives definite meaning to a concept by specifying one or more indicators of what we have in mind. An **indicator** is a sign of the presence or absence of the concept we’re studying. Here’s an example.

We might agree that visiting children’s hospitals during Christmas and Hanukkah is an indicator of compassion. Putting little birds back in their nests might be agreed on as another indicator, and so forth. If the unit of analysis for our study is the individual person, we can then observe the presence or absence of each indicator for each person under study. Going beyond that, we can add up the number of indicators of compassion observed for each individual. We might agree on ten specific indicators, for example, and find six present in our study of Pat, three for John, nine for Mary, and so forth.

Returning to our question about whether men or women are more compassionate, we might calculate that the women we studied displayed an average of 6.5 indicators of compassion, the men an average of 3.2. On the basis of our quantitative analysis of group difference, we might therefore conclude that women are, on the whole, more compassionate than men.

Usually, though, it’s not that simple. Imagine you’re interested in understanding a small fundamentalist religious cult, particularly their harsh views on various groups: gays, nonbelievers, feminists, and others. In fact, they suggest that anyone who refuses to join their group and abide by its teachings will “burn in hell.” In the context of your interest in compassion, they don’t seem to have much. And yet, the group’s 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 sinful ways. Within their own paradigm, then, cult members would see beating up gays, prostitutes, and abortion doctors as acts of compassion.

Social researchers focus their attention on the meanings given to words and actions by the people under study. Doing so can often clarify the behaviors observed: At least now 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 complicated 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 about specifying what we mean by them, we discover disagreements and inconsistencies. Not only do you and I disagree, but each of us is likely to find a good deal of muddiness 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 compassion. That is, the entries on your file sheet can be combined into groups and subgroups, say, compassion toward friends, co-religionists, humans, and birds. You may also find several different strategies for making combinations. For example, you might group the entries into feelings and actions.

The technical term for such groupings is **dimension**: a specifiable aspect of a concept. For instance, we might speak of the feeling dimension of compassion and the action dimension of compassion. In a different grouping scheme, we might distinguish compassion for humans from compassion for animals. Or we might see compassion as helping people have what we want for them versus

what they want for themselves. Still differently, we might distinguish compassion as forgiveness from compassion as pity.

Thus, we could subdivide compassion into several clearly defined dimensions. A complete conceptualization involves both specifying dimensions and identifying the various indicators for each.

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 men or women are really more compassionate. Our research would have shown that there is no single answer to the question. That alone represents an advance in our understanding of reality.

The Interchangeability of Indicators

There is another way that the notion of indicators can help us in our attempts to understand reality by means of “unreal” 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 logic 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—on 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 several different indicators all represent, to some degree, the same concept, then all of them will behave the same way that the concept would behave if it were real and could be observed. Thus, given a basic agreement about what “compassion” is, if women are generally more compassionate than men, we should be able to observe that difference by using any reasonable measure of compassion. If, on the other hand, women are more compassionate than men on some indicators but not on others, we should see if 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 turning 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 the meanings of terms with standards, consistency, and commonality.

Real, Nominal, and Operational Definitions

As we have seen, the design and execution of social research requires us to clear away 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 has cautioned,

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.

(1952:6)

In other words, trying to specify the “real” meaning of concepts only leads to a quagmire: It mistakes a construct for a real entity.

The specification of concepts in scientific inquiry depends instead 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—I could define *compassion* as “plucking feathers off helpless birds” if I wanted to—but they can 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 an earlier chapter, specifies precisely how a concept will be measured—that is, the operations we will 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. In 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 Rossman (1995:18) speak of a “conceptual funnel” through which a researcher's interest becomes increasingly focused. Thus, a general interest in social activism could narrow to “individuals who are committed to empowerment and social change” and further focus on discovering “what experiences shaped the development of fully committed social activists.” This focusing process is 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. society. Imagine how the meaning of oppression would shift as you delved more and more deeply into the members' experiences and worldviews. For example, you might start out 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 textual 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.

(Kvale 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 is 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 concepts

under study. Without that commitment, the study could not proceed further.

Even in less-structured research methods, however, it's important to begin with an initial set of anticipated 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:

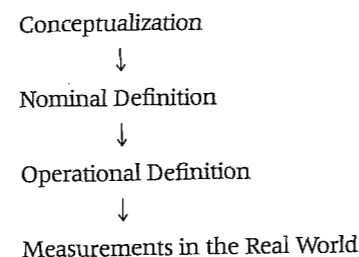
1. What was your total family income during the past 12 months?
2. What is the highest level of school you completed?

To organize our data, we would probably want to specify a system for categorizing the answers people give us. 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: less than high school, high school, college, graduate degree. Finally, we would specify the way a person's responses to these two questions would be combined in creating a measure of SES.

In this way we would create a working and workable definition of SES. Although others might disagree with our conceptualization 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.

Here is a diagram showing the progression of measurement steps from our vague sense of what a term 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 scientific 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 produced a work of analytical 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 compared suicide rates in northern and southern countries and regions, and he examined the different suicide rates across the months and seasons of the year. Thus, he could draw conclusions about a supremely individualistic and personal act without having any data about the individuals engaging in it.

At a more general level, Durkheim suggested that suicide also reflects the extent to which a society'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 confusion, anxiety, and even self-destruction. To describe this societal condition of normlessness, Durkheim chose the term *anomie*. Durkheim did not make this word up. Used in both German and French, it literally meant "without law." The English term *anomy* had been used for at least three centuries before Durkheim to mean disregard for divine law. However, Durkheim created the social scientific concept of *anomie*.

In the years that have followed the publication of *Suicide*, social scientists have found *anomie* a useful concept, and many have expanded on Durkheim's use. Robert Merton, in a classic article entitled "Social Structure and Anomie" (1938), concluded that *anomie* results from a disparity between the goals and means prescribed by a society. Monetary success, for example, is a widely shared goal in our society, yet not all individuals have the resources to achieve it through acceptable means. An emphasis on the goal itself, Merton suggested, produces normlessness, because those denied the traditional avenues to wealth go about getting it through illegitimate means. Merton's discussion, then, could be considered a further conceptualization of the concept of *anomie*.

Although Durkheim originally used the concept of *anomie* as a characteristic of societies, as did Merton after him, other social scientists have used it to describe individuals. To clarify this distinction, some scholars have chosen to use *anomie* in reference to its original, societal meaning and to use the term *anomia* in reference to the individual characteristic. In a given society, then, some individuals experience *anomia*, and others do not. Elwin Powell, writing 20 years after Merton, provided the following conceptualization of *anomia* (though using the term *anomie*) 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.

(1957:132)

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

Although many researchers have offered operational definitions of *anomia*, one name stands out over all. Two years before Powell's article appeared, Leo Srole (1956) published a set of questionnaire items that he said provided a good measure of *anomia* as experienced by individuals. It consists 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.

(1956:713)

In the decades following its publication, the Srole scale has become a research staple for social scientists. You'll likely find this particular operationalization of *anomia* used in many of the research projects reported in academic journals. Srole touches on this in the accompanying box, "The Origins of *Anomia*," which he prepared for this book before his death.

This abbreviated history of *anomie* and *anomia* as social scientific concepts illustrates several points. First, it is a good example of the process through which general concepts become operationalized measurements. This is not to say that the issue of how to operationalize *anomie/anomia* has been resolved once and for all. Scholars will surely continue

The Origins of Anomia

by Leo Srole

Center for Geriatrics and Gerontology,
Columbia University

My career-long fixation on anomie began with reading Durkheim's *Le Suicide* as a Harvard undergraduate. Later, as a graduate student at Chicago, I studied under two Durkheimian anthropologists: William Lloyd Warner and Alfred Radcliffe-Brown. Radcliffe-Brown had carried on a lively correspondence with Durkheim, making me a collateral "descendant" of the great French sociologist.

For me, the early impact of Durkheim's work on suicide was mixed but permanent. On the one hand, I had serious reservations about his strenuous, ingenious, and often awkward efforts to force the crude, bureaucratic records on suicide rates to fit with his unidirectional sociological determinism. On

the other hand, I was moved by Durkheim's unswerving preoccupation with the moral force of the interpersonal ties that bind us to our time, place, and past, and also his insights about the lethal consequences that can follow from shrinkage and decay in those ties.

My interest in anomie received an eyewitness jolt at the finale of World War II, when I served with the United Nations Relief and Rehabilitation Administration, helping to rebuild a war-torn Europe. At the Nazi concentration camp of Dachau, I saw first-hand the depths of dehumanization that macro-social forces, such as those that engaged Durkheim, could produce in individuals like Hitler, Eichmann, and the others serving their dictates at all levels in the Nazi death factories.

Returning from my UNRRA post, I felt most urgently that the time was long overdue to come to an understanding of the dynamics underlying disintegrated social

bonds. We needed to work expeditiously, deemphasizing proliferation of macro-level theory in favor of a direct exploratory encounter with individuals, using newly developed state-of-the-art survey research methodology. Such research, I also felt, should focus on a broader spectrum of behavioral pathologies than suicide.

My initial investigations were a diverse effort. In 1950, for example, I was able to interview a sample of 401 bus riders in Springfield, Mass. Four years later, the Midtown Manhattan Mental Health Study provided a much larger population reach. These and other field projects gave me scope to expand and refine my measurements of that quality in individuals which reflected the macro-social quality Durkheim had called *anomie*.

While I began by using Durkheim's term in my own work, I soon decided that it was necessary to limit the use of that concept to

its macro-social meaning and to sharply segregate it from its individual manifestations. For the latter purpose, the cognate but hitherto obsolete Greek term, *anomie*, readily suggested itself.

I first published the anomie construct in a 1956 article in the *American Sociological Review*,* describing ways of operationalizing it, and presenting the results of its initial field application research. By 1982, the Science Citation Index and Social Science Citation Index had listed some 400 publications in political science, psychology, social work, and sociology journals here and abroad that had cited use of that article's instruments or findings, warranting the American Institute for Scientific Information to designate it a "citation classic."

*Leo Srole, "Social Integration and Certain Corollaries: An Exploratory Study," *American Sociological Review* 21 (1956): 709–16.

to reconceptualize and reoperationalize these concepts for years to come, continually seeking more useful measures.

The Srole scale illustrates another important point. Letting conceptualization and operationalization be open-ended does not necessarily produce anarchy and chaos, as you might expect. Order often emerges. For one thing, although we could define anomia any way we chose—in terms of, say, shoe size—we're likely to define it in ways not too different from other people's mental images. If you were to use a really offbeat definition, people would probably ignore you.

A second source of order is that, as researchers discover the utility of a particular conceptualization and operationalization of a concept, they're likely to adopt it, which leads to standardized definitions of concepts. Besides the Srole scale,

examples include IQ tests and a host of demographic and economic measures developed by the U.S. Census Bureau. Using such established measures has two advantages: They have been extensively pretested and debugged, and studies using the same scales can be compared. If you and I do separate studies of two different groups and use the Srole scale, we can compare our two groups on the basis of anomia.

Social scientists, then, can measure anything that's real; through conceptualization and operationalization, they can even do a pretty good job of measuring things that aren't. Granting that such concepts as socioeconomic status, prejudice, compassion, and anomia aren't ultimately real, social scientists can create order in handling them. It is an order based on utility, however, not on ultimate truth.

Definitions in Descriptive and Explanatory Studies

As you recall from Chapter 4, 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 a simpler task than is explanation, you may be surprised to learn that definitions are more problematic for descriptive research than for explanatory research. Before we turn to other aspects of measurement, you'll need a basic understanding of why this is so (we'll discuss this point more fully in Part 4).

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 definition of another term: the labor force. If it seems patently absurd to regard a three-year-old child as being unemployed, it is because such a child is not considered a member of the labor force. Thus, we might follow the U.S. Census Bureau's 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

The Importance of Variable Names

by Patricia Fisher

Graduate School of Planning,
University of Tennessee

Operationalization is one of those things that's easier said than done. It is quite simple to explain to someone the purpose and importance of operational definitions for variables, and even to describe how operationalization typically takes place. However, until you've tried to operationalize a rather complex variable, 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 convinced that

participation in the planning process by citizens is important to the success of plan implementation. Citizen participation is an aid to planners' understanding of the real and perceived needs of a community, and such involvement by citizens tends to enhance their cooperation with and support for planning efforts. Although many different conceptual definitions might be offered by different planners, there would be little misunderstanding over what is meant by *citizen participation*. The name of the variable seems adequate.

However, if we asked different planners to provide very simple operational measures for citizen participation, we are likely to find a variety among their responses that does generate confusion. One planner might keep a tally of attendance by private citizens at city commission and other local government meet-

ings; another might maintain a record of the different topics addressed by private citizens at similar meetings; while a third might record the number of local government meeting attendees, letters, and phone calls received by the mayor and other public officials, and meetings held by special interest groups during a particular time period. As skilled researchers, we can readily see that each planner would be measuring (in a very simplistic fashion) a different dimension of citizen participation: extent of citizen participation, issues prompting citizen participation, and form 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 purposes of operationalization.

The precise and exact naming of variables is important in research. It is both essential to

and a result of good operationalization. Variable names quite often evolve from an iterative process of forming a conceptual definition, then an operational definition, then renaming the concept to better match what can or will be measured. This looping process continues (our example above illustrates only one iteration), resulting in a gradual refinement of the variable name and its measurement until a reasonable fit is obtained. Sometimes the concept of the variable that you end up with is a bit different from the original one that you started with, but at least you are measuring what you are talking about, if only because you are talking about what you are measuring!

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 employed.

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 or be open to an 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 conclusion of a descriptive study about the unemployment rate depends directly on how each issue of definition is resolved.

Increasing 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 are excluding military personnel; that, too, increases the reported unemployment rate, because military personnel would be employed—by definition. Thus the descriptive statement that the unemployment rate in a city is 3 percent, or 9 percent, or whatever it might be, depends directly on the operational definitions used.

This example is relatively clear because there are several accepted conventions relating to 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? More specifically, let's suppose we're interested in whether conservatism increases with age. What if you and I have 25 different operational definitions of *conservative*, 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 old people to be more conservative than young people in terms of all 25 definitions. Clearly,

the exact definition would be of small consequence. Suppose we found old people to be more conservative than young people by every reasonable definition of conservatism we could think of. It wouldn't matter what our definition was. 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 unambiguous as this example suggests; nonetheless, the general pattern is quite common in actual research. There are consistent patterns of relationships in human social life that result in consistent research findings. However, such consistency does not appear in a descriptive situation. Changing definitions almost inevitably result in different descriptive conclusions. The box "The Importance of Variable Names"

explores this issue in connection with the variable *citizen participation*.

Operationalization Choices

In discussing conceptualization, I frequently have referred to operationalization, for the two are intimately linked. To recap: Conceptualization is the refinement and specification of abstract concepts, and operationalization is the development of specific research procedures (operations) that will result in empirical observations representing those concepts in the real world.

As with the methods of data collection, social researchers have a variety of choices when operationalizing a concept. Although the several choices are intimately interconnected, I've separated them for the sake of discussion. Realize, though, that operationalization does not proceed through a systematic checklist.

Range of Variation

In operationalizing any concept, researchers must be clear about the range of variation that interests them. The question is, to what extent are we 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 run into the millions of dollars, but not many people get that much. Unless you're studying the very rich, it probably won't add much to your study to keep track of extremely high categories. Depending on whom you study, you'll probably want to establish a highest income category with a much lower floor—maybe \$100,000 or more. Although this decision will lead you to throw together people who earn a trillion dollars a year with paupers earning a mere \$100,000, they'll survive it, and that mixing probably won't hurt your research any, either. The same decision faces you at the other end of the income spectrum. In studies of the general U.S. population, a bottom category of \$5,000 or less usually works fine.

In studies of attitudes and orientations, the question of range of variation has another dimension. Unless you're careful, you may end up measuring only half an attitude without really meaning to. Here's an example of what I mean.

Suppose you're interested in people's attitudes toward expanding the use of nuclear power generators. You'd anticipate that some people consider nuclear power the greatest thing since the wheel, whereas other people have absolutely no interest in it. Given that anticipation, it would seem to make sense to ask people how much they favor expanding the use of nuclear energy and to give them answer categories ranging from "Favor it very much" to "Don't favor it at all."

This operationalization, however, conceals half the attitudinal spectrum regarding nuclear energy. Many people have feelings that go beyond simply not favoring it: They are, with greater or lesser degrees of intensity, actively opposed to it. In this instance, there is considerable variation on the left side of zero. Some oppose it a little, some quite a bit, and others a great deal. To measure the full range of variation, then, you'd want to operationalize attitudes toward nuclear energy with a range from favoring it very much, through no feelings one way or the other, to opposing it very much.

This consideration applies to many of the variables social scientists study. Virtually any public issue involves both support and opposition, each in varying degrees. Political orientations range from very liberal to very conservative, and depending on the people you're studying, you may want to allow for radicals on one or both ends. Similarly, people are not just more or less religious; some are positively antireligious.

The point is not that you must measure the full range of variation in every case. You should, however, consider whether you need to, given your particular research purpose. If the difference between not religious and antireligious isn't relevant to your research, forget it. Someone has defined pragmatism as "any difference that makes no difference is no difference." Be pragmatic.

Finally, decisions on the range of variation should be governed by the expected distribution of attributes among the subjects of the study. In a study of college professors' attitudes toward the

value of higher education, you could probably stop at no value and not worry about those who might consider higher education dangerous to students' health. (If you were studying students, however . . .)

Variations between the Extremes

Degree of precision is a second consideration in operationalizing variables. What it boils down to is how fine you will make distinctions among the various possible attributes composing a given variable. Does it matter for your purposes whether a person is 17 or 18 years old, or could you conduct your inquiry by throwing them together in a group labeled 10 to 19 years old? Don't answer too quickly. If you wanted to study rates of voter registration and participation, you'd definitely want to know whether the people you studied were old enough to vote. In general, if you're going to measure age, you must look at the purpose and procedures of your study and decide whether fine or gross differences in age are important to you. In a survey, you'll need to make these decisions in order to design an appropriate questionnaire. In the case of in-depth interviews, these decisions will condition the extent to which you probe for details.

The same thing applies to other variables. If you measure political affiliation, will it matter to your inquiry whether a person is a conservative Democrat rather than a liberal Democrat, or will it be sufficient to know the party? In measuring religious affiliation, is it enough to know that a person is a Protestant, or do you need to know the denomination? Do you simply need to know whether or not a person is married, or will it make a difference to know if he or she has never married or is separated, widowed, or divorced?

There is, of course, no general answer to such questions. The answers come out of the purpose of a given study, or why we are making a particular measurement. I can give you a useful guideline, though. Whenever you're not sure how much detail to pursue in a measurement, get too much rather than too little. When a subject in an in-depth interview volunteers that she is 37 years old, record "37" in your notes, not "in her thirties." When

you're analyzing the data, you can always combine precise attributes into more general 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 a 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 essential to be clear about which ones are important in our 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 have determined how you're going to collect your data (for example, survey, field research) and 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-logical one. That is, you may need to decide what level of measurement to use. To discuss this point, we 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 "student." Variables, on the other hand, are logical sets of attributes. Thus, *gender* is a variable composed of the attributes female and male.

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, *employment status* is a variable having the attributes employed and unemployed; the list of attributes could also be expanded to include the other possibilities discussed earlier, such as homemaker.

Every variable must have two important qualities. First, the attributes composing it should be exhaustive. For the variable to have any utility in research, we 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 the variable *political party affiliation* in terms of the attributes Republican and Democrat, because some of the people 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 "no 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 is also looking for work. (We might run across a fully employed mud wrestler who is look-

ing 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, birthplace, college major, 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 structures. 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 at least one thing in common and differ from the people in all other groups in that same regard. Where the individual groups form, how close they are to one another, or how the groups are arranged in the room is irrelevant. All that matters is 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, conservatism, alienation, 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 not 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 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 materials it is harder than and which softer than.

Let's pursue the earlier example of grouping the people at a social gathering. 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.

Notice 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 5 feet apart, and the less-than-high-school group 500 feet farther down the line. These actual distances don't have any meaning. The high school group, however, should be between the less-than-high-school group and the college group, or else the rank order will be incorrect.

Interval Measures

For the attributes composing some variables, the actual distance separating those attributes 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 points in the Fahrenheit (and Celsius) scales are arbitrary; zero degrees does not really mean lack of heat. Similarly, minus 30 degrees on either scale doesn't represent 30 degrees less than no heat. (In contrast, the Kelvin scale is based on an absolute zero, which does mean a complete lack of heat.)

About the only interval measures commonly used in social scientific 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. But 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 received a score of 0 on a standard IQ test could not be regarded, strictly speaking, as having no intelligence, although we might feel he or she was unsuited to be a college professor or even a college student. But perhaps a dean . . . ?)

When comparing two people in terms of an interval variable, we can say they are different from one another (nominal), and that one is more than

another (ordinal). In addition, we can say "how much" more.

Ratio Measures

Most of the social scientific variables meeting the minimum requirements for interval measures also meet the requirements for ratio measures. In **ratio measures**, the attributes composing a variable, besides having all the structural characteristics mentioned previously, are based on a true zero point. The Kelvin temperature scale is one such measure. Examples from social scientific research include age, length of residence in a given place, number of organizations belonged to, number of times attending church during a particular period of time, number of times married, and number of Arab friends.

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 or lie) 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 and that each different group has 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 will be able to 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 hapless party goers 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.

Comparing two people in terms of a ratio variable, then, allows us to conclude (1) they are different (or the same), (2) one is more than the other, (3) how much they differ, and (4) the ratio of one

to another. Figure 5-1 summarizes this discussion by presenting a graphic illustration of the four levels of measurement.

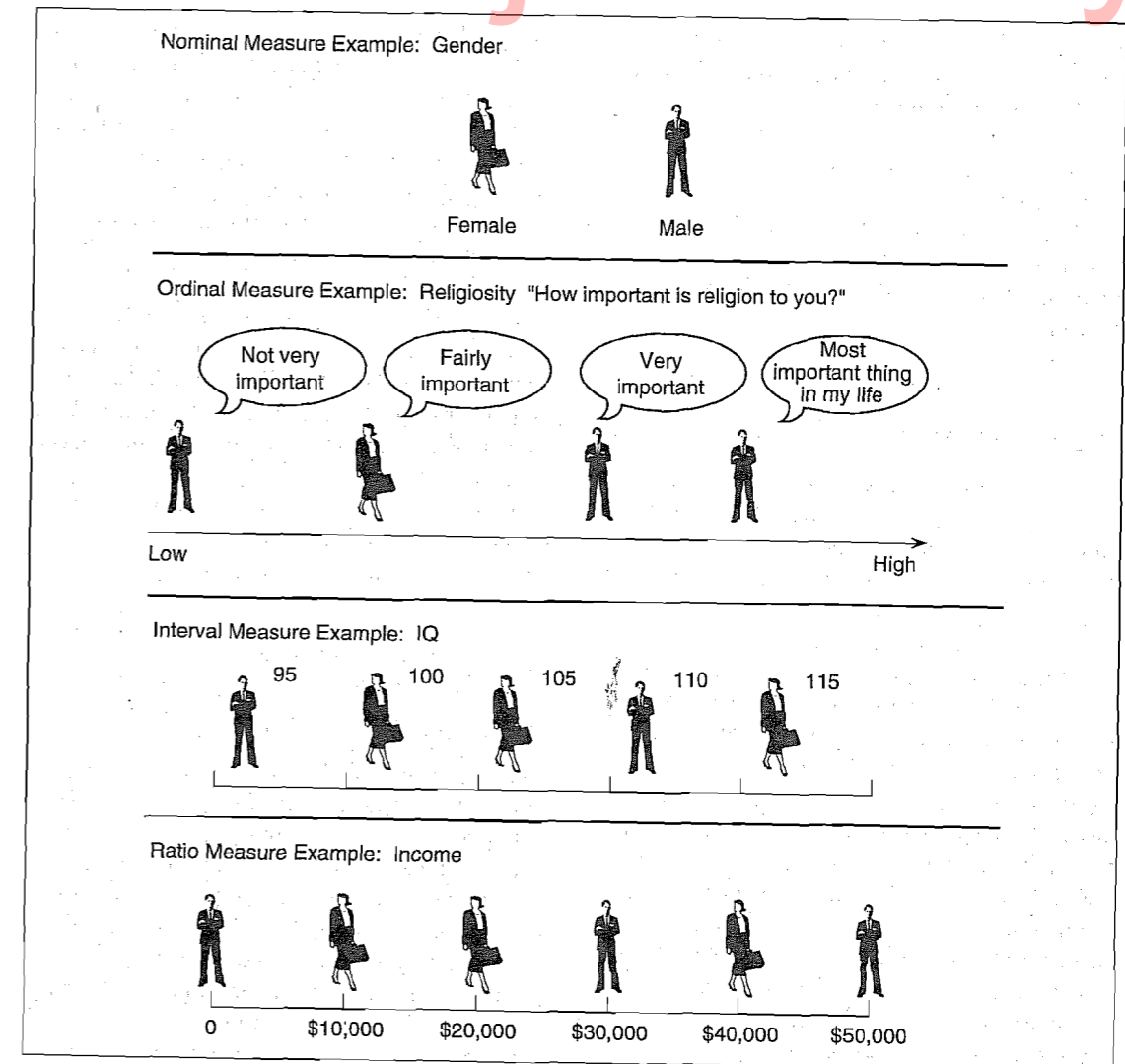
Implications of Levels of Measurement

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), I should draw your attention to some of the practical implications of the differences that have been distinguished. These implications appear primarily in the analysis of data (discussed in Part 4), but you need to anticipate such implications when you're structuring any research project.

Certain quantitative analysis techniques require variables that meet certain minimum levels of measurement. To the extent that the variables to be examined in a research project are limited to a particular level of measurement—say, ordinal—you should plan your analytical techniques accordingly. More precisely, you should anticipate drawing research conclusions appropriate to the levels of measurement used in your variables. For example, you might reasonably plan to determine and report the mean age of a population under study (add up all the individual ages and divide by the number of people), but you should not plan to report the mean religious affiliation, because that is a nominal variable, and the mean requires ratio-level data. (You could report the modal—the most common—religious affiliation.)

At the same time, you can treat some variables as representing different levels of measurement. Ratio measures are the highest level, descending through interval and ordinal to nominal, the lowest level of measurement. A variable representing a higher level of measurement—say, ratio—may also be treated as representing a lower level of measurement—say, ordinal. Recall, for example, that age is a ratio measure. If you wished to examine only the relationship between age and some ordinal-level variable—say, self-perceived religiosity: high, medium, and low—you might choose to treat age as an ordinal-level variable as well. You might characterize the subjects of your study as being young, middle-aged, and old, specifying what age range composed each of these groupings. Finally, age

FIGURE 5-1
Levels of Measurement



might be used as a nominal-level variable for certain research purposes. People might be grouped as being born during the depression of the 1930s or not. Another nominal measurement, based on birth date rather than just age, would be the grouping of people by astrological signs.

The level of measurement you'll seek, then, is determined by the analytical 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 different levels of measurement, the study should be designed to achieve the highest level required. For example, if the subjects in a study are asked their exact ages, they can later be organized into ordinal or nominal groupings.

You need not necessarily measure variables at their highest level of measurement, however. If you're sure to have no need for ages of people at higher than the ordinal level of measurement, you

have been collected in a structured mass survey. As we've seen, in less-structured methods such as field research, the identification and specification of relevant concepts is inseparable from the ongoing process of observation.

As a researcher, always be open to reexamining your concepts and definitions. The ultimate purpose of social research is to clarify the nature of social life. The validity and utility of what you learn in this regard doesn't depend on when you first figured out how to look at things any more than it matters whether you got the idea from a learned textbook, a dream, or your brother-in-law.

Criteria of Measurement Quality

This chapter has come some distance. It began with the bald assertion that social scientists can measure anything that exists. Then we discovered that most of the things we might want to measure and study don't really exist. Next we learned that it's possible to measure them anyway. Now we conclude the chapter with a discussion of some of the yardsticks against which we judge our relative success or failure in measuring things—even things that don't exist.

Precision and Accuracy

To begin, measurements can be made with varying degrees of precision. As we saw in the discussion of operationalization, precision concerns the fineness of distinctions made between the attributes that compose a variable. The description of a woman as "43 years old" is more precise than "in her forties." Saying a street-corner gang 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 would dictate. 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 any additional effort invested in learning her precise age is wasted. The operationalization of con-

cepts, then, must be guided partly 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 Stowe, Vermont"—but suppose 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 estimates 300, we have to conclude 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 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 does asking people to estimate it.

Reliability, however, does not ensure accuracy any more than precision does. Suppose I've set my bathroom scale to shave 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, just be warned 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 small teams of workers complete the whole process.

How should we measure morale? Following one strategy, we could observe the workers in each factory, noticing such things as whether they joke with one another, whether they smile and laugh a lot, and so forth. We could ask them how they like their work and even ask them whether they think they would prefer their current arrangement or the other one being studied. By comparing what we observed in the different factories, we might reach a conclusion about which assembly process produces the higher morale.

Now let's look at some reliability problems inherent in this method. First, how you and I are feeling when we do the observing will likely color what we see. We may misinterpret what we see. We may see workers kidding each other but think 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. If several observers evaluated the same behavior, on the other hand, they too might arrive at different conclusions about the workers' morale.

Here's another strategy for assessing morale. 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 for 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 different interviewers, because of their own attitudes and demeanors, get different answers from respondents. Or, if we were 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 hundreds of editorials and classifying them in terms of their position on the issue. Unfortunately, different coders will code the same editorial differently. 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.

Each of these examples illustrates problems of reliability. Similar problems arise whenever we ask people to give us information about themselves. Sometimes we ask questions that people don't know the answers to: How many times have you been to church? 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

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 less likely than dissatisfied ones to cheat 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 as likely to cheat on their spouses as are the dissatisfied ones, 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 doesn't tap the quality you want it to measure, 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 a particular comparison variable (driving record, marital fidelity) represents in a given situation. It is less important to distinguish the two types of validity tests than to understand the logic of validation that they have in common: If we have been successful in measuring some variable, then our measures should relate in some logical way to other measures.

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 alone but also needs to cover subtraction, multiplication, division, and so forth. Or, if we are measuring 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 hit, 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 What's Valid?

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.

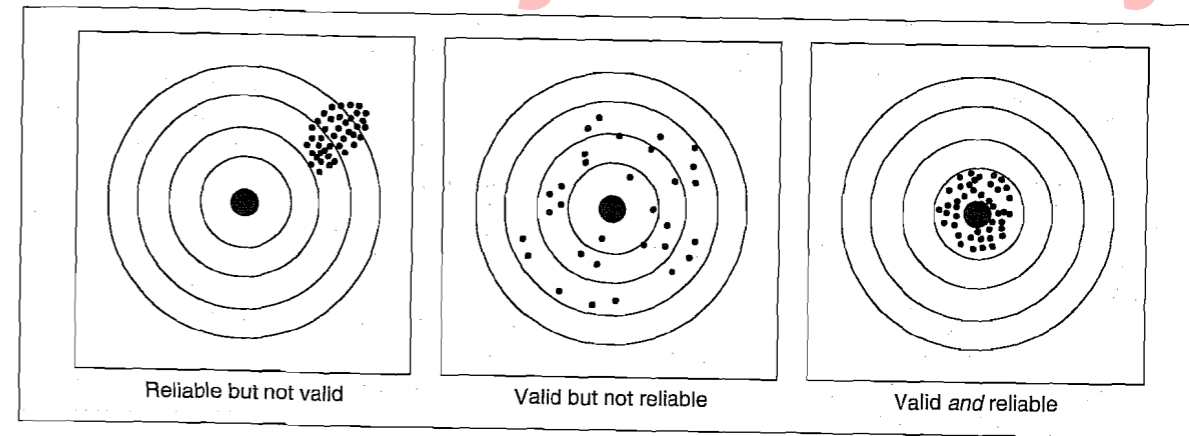
Social researchers sometimes criticize themselves and one another for implicitly assuming they are somewhat superior to those they study. For example, researchers often seek to uncover motivations that the social actors themselves are unaware of. You think you bought that new Burpo-Blasto because of its high performance and good looks, but *we* know you're really trying to achieve a higher social status.

This implicit sense of superiority would fit comfortably with a totally positivistic approach (the biologist feels superior to the frog on the lab table), but it clashes with the more humanistic and typically qualitative approach taken by many social scientists. We'll explore this issue more deeply in Chapter 10.

In seeking to understand the way ordinary people make sense of their worlds, ethnomethodologists have urged all social scientists to pay more respect to these natural social processes of conceptualization and shared meaning. At the very least, behavior that may seem irrational from the scientist's paradigm may make logical sense when viewed through the actor's paradigm.

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.

FIGURE 5-2
An Analogy to Validity and Reliability



Tension between Reliability and Validity

Clearly, we want our measures to be both reliable and valid. However, there is often a tension 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 would counting grievances. 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 strain in research measurement. Most of the really interesting concepts we want to study have many subtle nuances, and it's hard to specify precisely what we mean by them. 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 anomie/anomia, 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 filed with the union; anomie is much 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 two quite different approaches to social research: quantitative, nomothetic, structured techniques such as surveys and experiments on the one hand, and qualitative, idiographic 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.

MAIN POINTS

- Conceptions are mental images we use as summary devices for bringing together observations and experiences that seem to have something

in common. We use terms or labels to reference these conceptions.

- Concepts are constructs; they represent the agreed-upon meanings we assign to terms. Our concepts don't exist in the real world, so they can't be measured directly, but it's possible to measure the things that our concepts summarize.
- Conceptualization is the process of specifying observations and measurements that give concepts 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.
- 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 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 inter-related choices: specifying the range of variation that is appropriate for the purposes of a 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 measures that 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 to that variable so they can capture the greatest amount of information.
- Operationalization begins in the design phase of a study and continues through all phases of

the research project, including the analysis of data.

- 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 or 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 yardsticks 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 meaning our general concepts have. This problem is inevitable. The best solution is to use several different measures, tapping the different aspects of a concept.

KEY TERMS

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

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 in and exclude from your conceptualization.
2. Locate a research report in a book or journal article. Identify the key variable studied by the researcher(s) and describe how the variable was operationalized for measurement.

3. 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 (strongly approve, approve, disapprove, strongly disapprove)
 - f. Region of birth (Northeast, Midwest, and so on)
 - g. Political orientation (very liberal, somewhat liberal, somewhat conservative, very conservative)
4. In a newspaper or magazine, find an instance of invalid and/or unreliable measurement. Justify your choice.
5. Go to Holocaust Studies: Prejudice (<http://www.socialstudies.com/c/ZcCwFuEspbb41/Pages/holo.html>) and browse through the materials described there. Make a list of the various dimensions of prejudice that you find there.

ADDITIONAL READINGS

- Bohrnstedt, George W. 1983. "Measurement." Pp. 70–121 in *Handbook of Survey Research*, edited by Peter H. Rossi, James D. Wright, and Andy B. Anderson. New York: Academic Press. This essay offers the logical and statistical grounding of reliability and validity in measurement.
- Grimes, Michael D. 1991. *Class in Twentieth-Century American Sociology: An Analysis of Theories and Measurement Strategies*. New York: Praeger. This book provides an excellent, long-term view of conceptualization as the author examines a variety of theoretical views of social class and the measurement techniques appropriate to those theories.
- Lazarsfeld, Paul F., and Morris Rosenberg, eds. 1955. *The Language of Social Research*, Section I. New York: Free Press of Glencoe. An excellent and diverse classic collection of descriptions of specific


measurements in past social research. These 14 articles present useful and readable accounts of actual measurement operations performed by social researchers, as well as more conceptual discussions of measurement in general.

Miller, Delbert. 1991. *Handbook of Research Design and Social Measurement*. Newbury Park, CA: Sage. A powerful reference work. This book, especially Part 6, cites and describes a wide variety of operational measures used in earlier social research. In several cases, the questionnaire formats used are presented. Though the quality of these illustrations is uneven, they provide excellent examples of possible variations.

Silverman, David. 1993. *Interpreting Qualitative Data: Methods for Analyzing Talk, Text, and Interaction*, Chapter 7. Newbury Park, CA: Sage. This chapter deals with the issues of validity and reliability specifically in regard to qualitative research.

U.S. Department of Health and Human Services. 1992. *Survey Measurement of Drug Use*. Washington, DC: Government Printing Office. An extensive review of techniques devised and used for measuring various kinds of drug use.


SOCIOLOGY WEB SITE

 See the Wadsworth Sociology Resource Center, Virtual Society, for additional links, Internet exercises by chapter, quizzes by chapter, and Microcase-related materials:

<http://www.sociology.wadsworth.com>

INFOTRAC COLLEGE EDITION

SEARCH WORD SUMMARY

 Go to the Wadsworth Sociology Resource Center, Virtual Society, to find a list of search words for each chapter. Using the search words, go to InfoTrac College Edition, an online library of over 900 journals where you can do online research and find readings related to your studies. To aid in your search and to gain useful tips, see the Student Guide to InfoTrac College Edition on the Virtual Society Web site:

<http://www.sociology.wadsworth.com>