WK. I

CHAPTER 3

# THEORIES

*There is nothing so practical as a good theory.*
—Kurt Lewin

*Regardless of whether it is called basic or evaluative, research that neglects theory has little long-run impact.*
—Daniel Glaser

*Nothing can justify a theory except its explaining observed facts.*
—C.S. Peirce

## OBJECTIVES

Without theory, scientists would be no different from newspaper reporters or accountants. Criminal justice would not be a respected social science discipline. In this chapter you will examine the building blocks of theories, and how theories connect to the real world and the world of data. You also will consider a difficult issue: What is the relationship between theory and reality?

## WHO NEEDS THEORY

Theories sound fancy and useless. You may think scientists who make up theory are just big "eggheads" or "nerds," like Fred MacMurray in *The Absent-Minded Professor*, who couldn't even remember his wedding date. You may envision **theorizing**, the process of constructing theories, to be marginally less exciting than watching paint dry. You may have chosen a criminal justice major because you thought you could *do* something with it, perhaps as a police officer, a correctional worker, and administrator, or a reformer in the criminal justice system. You might not view your future career as exciting as Jamie Lee Curtis's in *Blue Steel*, or Danny Glover's in *Lethal Weapon*. Nevertheless, you probably did not think that activities like "theory building" or "theory testing" would play a big part.

Have I correctly anticipated your reactions to the topic of theory? If so, I want you to change your mind or to at least *think* about changing your mind. I will suggest that, without theory, you, and the entire field of criminal justice, are lost. Researchers investigating the criminal justice system and evaluating correctional programs have been led astray when they have failed to use theory to guide their investigations. [1]

## ORGANIZATION OF THE CHAPTER

In the last chapter I continued the scenario in which you were an R&D officer in a police department. You were investigating why some police officers expressed more satisfaction with a community-policing assignment than others. In this chapter I revisit the scenario. Returning to the theory I suggested you might have developed, we inspect its details more closely. You will explore the *components* of theory and the *steps of theory construction*. Finally, you will examine the relationship between scientific theories and the nature of reality.

## THE PROCESS OF THEORIZING AND THE ELEMENTS OF THEORY

### The Theory of Community-Policing Satisfaction

In Chapter 2 the theory I suggested you might have developed, using Holmes's method of scientific inquiry, contained the following initial predictions:

1. More sociable officers found the community police officer role more satisfying.
2. Older officers found the community police officer role less satisfying; younger officers found it more satisfying.
3. Younger officers, believing that crime fighting is always the most important job of a police officer, were less satisfied with the community-policing role than younger officers not holding this view.

The empirical evidence you gathered subsequently supported Predictions 1 and 2, but did not support Prediction 3. Close inspection of the data suggested a fourth prediction:

4. Officers who were more cynical about police work were less satisfied with the community police officer role.

### The Elements of the Theory and Theory Testing

**Level** Your theory focuses on dynamics occurring at a particular *level of analysis*. It concerns *individuals*, not organizations or groups or counties or states or countries. Theories can focus on *different* levels of analysis. (See Box 3.1.)

BOX 3.1

**LEVELS OF EXPLANATION IN CRIMINAL JUSTICE AND CRIMINOLOGICAL THEORY**

Theorists can frame their ideas at a general and abstract level. Such theories represent **macrotheories**, concerned with societal or systemwide dynamics. [2] A criminal justice macrotheory, for example, might explain how the criminal justice system protects the vested interests of the higher classes of society. A criminological macrotheory might explain the different rates of juvenile delinquency across different social classes. Such theories focus on societal or structural concepts. They are sociological in spirit.

By contrast, **microtheories** focus on specific cases, or groups of people, or individuals. [2] The component ideas are more concrete. A criminological theory might explain why certain youths in a neighborhood become delinquent whereas others do not. A criminal justice microtheory might explain why some cases are plea bargained, while other comparable cases are not.

Between these two levels of explanation you will find **bridging theories**. They concentrate *both* on structural factors and individual-level or small-group-level factors. [2] They also can be called *structural–contextual* theories. They consider both social structure, and the specific context in which events are occurring. [1] In criminology such theories consider both the societal and individual factors leading to criminality. In criminal justice such theories consider the social and political milieu of the criminal justice system and its subsystems, as well as subsystem characteristics and connections.

**Where Do Theories Come From?** In this section I portray the development and testing of theory using Einstein's approach to scien-

For a very brief period, medieval scientists were known to have dabbled in the merits of cardboard armor.

An example of a theory that did *not* fit reality.

(*Source*. THE FAR SIDE © 1989 FarWorks, Inc. Distributed by Universal Press Syndicate. Reprinted with permission. All rights reserved.)

tific inquiry. Then I will explain how to follow the process using Holmes's scientific approach.

Theories arise from many sources. We observe, directly and through media and conversation, what happens in the "real world." We absorb ideas and attitudes from the surrounding political, social, and intellectual context. Theories often start with a concern about things that occur in, or could occur in, the real world.

Look at Figure 3.1. It represents a schematic for guiding our discussion on theories and theorizing. Following Einstein's approach, we will move from the top to the bottom of the figure.

Theories have their origin in **features** and *dynamics* in the *real world*. Features, as I use the term here, refer to **attributes** or facets of units or groups of units in the real world. All of the following qualify as features of the real world:

- attitudes toward the death penalty for persons convicted of killing police officers;
- burglary clearance rates in midwestern towns during calendar year 1993; and
- the levels of satisfaction held by the community-policing officers in our running scenario.

**Dynamics** describe connections between features. We often presume that these connections are *causal* in nature. Something happened, and this caused something else to happen. For example, your mother's good high school friend had a police officer for a husband who was killed in the line of duty. Consequently, your mother holds the belief that all convicted killers of police officers deserve the death penalty. In our running scenario, because certain officers had more sociable personalities, they enjoyed their community-policing role more.

The dynamics and features happening and observed in the real world, or that could happen, are *not* theory. They represent one possible origin of a theory.

**In Your Head: Building Blocks and Glue**

*Concepts* Through several different channels, the features and dynamics occurring in the real world, or that could occur in the real world, produce notions in our own heads. We observe, we read, we imagine, we reflect, we discuss, and sometimes we argue with others about a range of topics of interest. During and after these activities we develop **concepts**. These also can be called **constructs**. *These make up the basic building blocks of theory.*

Concepts refer to features of real or possible worlds. But the concepts themselves *do not exist* "out there" in the real world. They are *ideas* about things that exist or could exist.

Einstein's Mode of Inquiry

| DOMAIN | Features | Dynamic Linking Features |
|---|---|---|
| "REAL" WORLD | Events<br>Charatericstics<br>Attitudes<br>of:<br>People, organizations, other units | Process or dynamics whereby one thing leads to another<br>F1 ⟶ F2 |
| IN YOUR HEAD: THEORY BEFORE OPERATIONALIZATION | Concept of x (Cx)<br>Concept of y (Cy) | Propositions link concepts<br>e.g.; Cx ⟶ Cy<br>Causal logic assumed |
| | Op. G.T. | |
| IN YOUR HEAD: THEORY AFTER OPERATIONALIZATION | Variables reflecting concepts<br>e.g.; $V_{x1}$, $V_{x2}$<br>e.g.; $V_{y1}$, $V_{y2}$<br>("True" scores) | Hypotheses express links between variables<br>e.g.; $V_{x1} \longrightarrow V_{y1}$<br>Causal logic assumed |
| | Meas; E added G.T. | |
| DATA WORLD: WHAT YOU OBSERVE WITH YOUR DATA | Variables as measured<br>e.g.; $V_{x1obs}$, $V_{x2obs}$<br>e.g.; $V_{y1obs}$, $V_{y2obs}$<br>(Observed scores) | Correlations between two or more observed variables<br>e.g.; $V_{x1obs} \longleftrightarrow V_{y1obs}$<br>Associational logic observed |

Holmes's Mode of Inquiry

Note: E = error, G.T. = grounded theorizing, Meas. = measurement process; Op. = operationalization

**FIGURE 3.1**

Although they are based on features of the real world, *they exist only in your head*.

In the case of the running scenario, you developed the concept of satisfaction with the community-policing role based on your conversations with officers, observations of officer-citizen interactions, and perhaps readings about the topic. Your *concept* of officer satisfaction with the community-policing role *relates* to these real world features, but is *separate* and distinct from it.

*Propositions* You connect the different conceptual building blocks—concepts—to each other through *propositions*.[1] **Propositions** are statements about the linkages between concepts. These statements often assume a **causal logic**. Something has caused something

[1] It is possible for a theory to have one building block and no glue. For example, your theory may simply state that over half of all community-policing officers will be more satisfied than dissatisfied with their role as community-policing officers. You can translate this theory into a testable hypothesis, and verify or falsify it. Most theories of interest to criminal justice researchers, however, have two or more concepts.

else. For example, in the running scenario your fourth prediction says that officers who were more cynical about police work would be, as a result, less satisfied with a community-policing role. Graphically, you could express this proposition as follows:

$$C_{\text{Police Cynicism}} \xrightarrow{(-)} C_{\text{Satisfaction with Community Officer Role}}$$

*Propositions Classify Concepts* Once you have connected concepts with propositions, you have divided your concepts into two classes. First, you have **predictor concepts**. Their role in your theory is to cause other concepts. These also can be called *exogenous* concepts, or *independent* concepts. They are exogenous, because their origins are *external* to the theory. They are independent because your theory does not anticipate that they are dependent upon, or caused by, any other concepts.

Second, you have **outcome concepts**. Their role in the theory is to be caused by other concepts. They also can be called *dependent* concepts, because they are influenced, according to your theory, by other concepts. You also can call them *endogenous* if you wish. They are endogenous because their origins are internal to the theory.

Theories often contain more than one proposition. Your theory of community-policing satisfaction is a case in point. Graphically, you could portray the theory, after its first empirical test, as follows:

| *Predictors* | *Outcome* | |
|---|---|---|
| $C_{\text{Sociability}} \xrightarrow{(+)}$ | | (Pred. 1) |
| $C_{\text{Age}} \xrightarrow{(-)}$ | $\rightarrow C_{\text{Satisfaction with CPO Role}}$ | (Pred. 2) |
| $C_{\text{Police Cynicism}} \xrightarrow{(-)}$ | | (Pred. 4) |

Although a proposition necessarily separates two concepts into predictor and outcome concepts, it also can create a third class of concept if the theory included three or more concepts. A proposition can classify a concept as a **mediating concept**. A mediating concept connects the predictor and outcome concepts by "carrying" or "channeling" the impact of the predictor to the outcome.

For example, after collecting your data on CPO satisfaction and examining them closely, you might conclude that the concept of police cynicism *mediates* the connection between age and satisfaction. You may propose that older officers exhibit less satisfaction with the community police officer role *because* their years of service have led them to develop cynical attitudes about police work. Age affects satisfaction because older officers become more cynical. It is this higher level of cynicism, not age per se, that results in less satisfaction with the CPO role. So now you would connect these three concepts so:

$$C_{\text{Age}} \xrightarrow{(+)} C_{\text{Police Cynicism}} \xrightarrow{(-)} C_{\text{Satisfaction with CPO Role}}$$

You can express your entire theory as follows:

$$\left.\begin{array}{l} C_{\text{Sociability}} \xrightarrow{(+)} \\ C_{\text{Age}} \xrightarrow{(+)} C_{\text{Police Cynicism}} \xrightarrow{(-)} \end{array}\right\} \rightarrow C_{\text{Satisfaction with CPO Role}}$$

In sum: concepts are connected by theoretical propositions that are usually causal in nature. The independent concept ($C_x$) or predictor causes the dependent concept ($C_y$) or outcome. Sometimes theories may position mediating constructs ($C_m$) between the independent and dependent constructs.

*Different Causal Logics* Propositions can assume several different causal logics. Here are the possibilities:

1. More of a predictor concept leads to more of an outcome concept. The outcome does not in turn influence the predictor. The

proposition posits a **positive, unidirectional** causal logic:[2]

$$C_x \xrightarrow{(+)} C_y$$

In the officer satisfaction model, more sociability leads to more satisfaction.

2. More of a predictor concept leads to less of an outcome concept. The outcome does not in turn influence the predictor. The proposition posits a **negative, unidirectional** causal logic:

$$C_x \xrightarrow{(-)} C_y$$

In the officer satisfaction model, more cynicism leads to less satisfaction.

3. **More** of a predictor concept leads to more **of an** outcome concept, which in turn **leads** to more of the predictor concept. The proposition posits a **positive feedback, bidirectional** causal logic.[3] This causal logic consists of two distinct propositions.

$$C_x \underset{(+)}{\overset{(+)}{\rightleftarrows}} C_y$$

For example, you might modify the satisfaction theory as follows. It not only predicts that more sociable officers will find the CPO role more satisfying. It also predicts that the more satisfied they are with the role, the more sociable they become. Over time such connections lead to more of each concept.

4. Your proposition can be *bidirectional* but posit **negative feedback between** the two concepts. More of the **predictor** leads to less of the outcome, which in turn leads to more of the predictor. Graphically:

$$C_x \underset{(-)}{\overset{(-)}{\rightleftarrows}} C_y$$

For example, the satisfaction theory might be modified to state the following: Not only do more cynical officers exhibit less satisfaction with the CPO role; in addition, the less satisfied they are with the role, the more cynical they become. Over time such a bidirectional proposition leads to less and less of the outcome.

5. Finally, your proposition may contain a **bidirectional homeostatic** logic. More of the predictor concept leads to less of the outcome concept, which in turn, leads to less of the predictor concept.

$$C_x \underset{(+)}{\overset{(-)}{\rightleftarrows}} C_y$$

Or, the relationship could work in the opposite way:

$$C_x \underset{(-)}{\overset{(+)}{\rightleftarrows}} C_y$$

Both relationships are *homeostatic* because the predictor concept and the outcome concept tend to balance each other. Thus the two concepts remain at relatively stable levels over time. It is not unlike the connection between the thermostat and the furnace in your house. The temperature drops, leading the thermostat to kick

[2] A fancier term sometimes used for a unidirectional causal logic is *recursive*.

[3] A fancier term sometimes used for a bidirectional relationship is *nonrecursive*.

the heater on. The temperature rises again, shutting off the thermostat and the heater. Over time the temperature stays relatively constant.

*All the Constructs Must Be Linked Together by One or More Propositions* Constructs take their meaning in part from the propositions linking them to other constructs in the theory. [3] These connections represent the **constitutive definition** of the constructs. [4] If a theory has three constructs, and the third construct does not link to the other two constructs by any propositions, that third construct is superfluous. It has no constitutive definition.

**Still in Your Head But in the Realm of Operational Definitions and Hypotheses** The theory you have developed now contains concepts and propositions. Nevertheless, it is still vague. You have not specified how you intend to *operationalize* each concept in the theory. You do not yet know what you would *do* to collect data for each concept.

To guide your data collection efforts you must provide **operational definitions** of the concepts. The operational definition of a concept tells you *how to assess it empirically—how to gather data on the concept.* The process of operationalization—attaching operational definitions to each concept—takes us from the second to the third row of Figure 3.1. Operationalization leads us to a vital intermediate stage in the theorizing process, a stage connecting actual scientific data with theoretical constructs.

Through the process of operationalization we translate concepts into **variables.** Variables, before they are measured, represent the observable and measurable portions of concepts that you have defined operationally. They provide a specific direction for measuring the concepts in your theory.


A developmental theory of delinquency before operationalization. (*Source.* THE FAR SIDE © 1989 FarWorks, Inc. Distributed by Universal Press Syndicate. Reprinted with permission. All rights reserved.)

Concepts —(operationalization)→ Variables

In the running scenario predicting CPO satisfaction, you could operationalize the construct of "satisfaction with the CPO role" in several different ways. You could ask officers directly how satisfied they are with the role. You could ask them about several specific facets of the role, and the levels of satisfaction associated with each. You could ask fellow officers about the target officer's satisfaction with the job. You could look at supervisors' assessments of CPO job commitment.

No one potential variable or set of potential variables is more appropriate than another. To bring the point closer to home: "There is no correct recipe for a white cake. Similarly there is no one correct operational definition for a concept." [5]

A social science theory will have an operational definition for most of its concepts, but it need not necessarily have an operational definition for *all* of its concepts. The theory may have a concept defined solely in terms of other concepts, such as a mediating construct. "It is only necessary that a sufficient number [of constructs] in any system be operationally defined." [4, 6]

In short, you need at least one operational definition for most, and preferably all, of the constructs in your theory. These operational definitions suggest potential variables. You may wish to use more than one operational definition, and thus several variables, for one construct in your theory, or for several constructs.

*Operationalization Translates Propositions Into Hypotheses* The process of attaching operational definitions to most of the constructs in your theory permits you to translate your general propositions into specific hypotheses.

$$\text{Propositions} \xrightarrow{\text{(operationalization)}} \text{Hypotheses}$$

A **hypothesis** states the expected relationship between two or more variables; the statement embodies a causal logic framed so that it can be empirically tested.[4] *Hypotheses state specifically what you expect to observe with your data.* Each prediction (1, 2, and 4) stated in your revised theory of CPO satisfaction can be stated as a specific hypothesis. So the proposition that stated:

> Officers who are more cynical about police work will consequently be less satisfied with their jobs of community-policing officers

can be translated into a hypothesis stating the connection between two variables:

> Those officers who receive higher scores on the Police Cynicism Scale, compared to those officers with lower scores on the form, will provide a lower score on an item asking about satisfaction with the assignment as a community-policing officer.

Your hypotheses must be both *empirically verifiable* and *falsifiable*. A hypothesis is falsifiable if you can state before you examine your data what specific pattern of results would not support your hypothesis.

**Moving From Your Head to the Data World**
You are ready to move from your head to the world of data collection (row 4, Figure 3.1). You have operationalized the constructs in your theory, translating them into specific variables. You have stated the propositions linking your concepts in terms of specific, empirically testable, and falsifiable hypotheses. You are ready to *measure* and collect data. **Measurement** will move you from the conceptual realm to the empirical. The process achieves two results. It will generate observed scores on specific variables for the cases actually studied. It also introduces measurement error.

$$\text{Variables} \xrightarrow{\text{(measurement)}} \text{Observed Scores on Variables}$$

$$\text{Hypotheses} \xrightarrow{\text{(measurement)}} \text{Observed Associations Between Variables}$$

*What Gets Lost in the Translation Process* In the data world you have **observed scores** on specific variables. Officer 1 checked the box "1"-"extremely dissatisfied" when responding to the question on overall satisfaction as a CPO. Officer 2's total score on the cynicism scale you used was 56. Officer 3's score on the sociability scale you used was 38.

[4] Again, it is possible to have a hypothesis about just one variable. Such a hypothesis would specify the expected level of the variable.

As you proceeded from your *idea* of specific variables to the actual *observed scores* on specific variables, measurement error crept in. People may have misread questions. People may have not wanted to report their true feelings. You will be learning more about measurement error in Chapter 6. The only point you need remember here is to make an important distinction between *potential scores on potential variables as conceptualized* and *observed scores on variables actually measured.*

Similarly, the causal logic connecting variables loses something as it too is translated from the conceptual realm to the empirical realm. You framed your hypotheses using a causal logic. All that you can observe empirically, however, are *associations* between measured variables. You might have hypothesized that higher cynicism scores caused lower levels of satisfaction with the CPO role. But the data may show you only that observed scores on the two variables correlate in a negative fashion; they reveal only an *associational* logic. More cynical officers were less satisfied, and less satisfied officers were more cynical. Data do not reveal a causal logic. From the simple association you cannot be sure which variable caused which other variable if both were measured simultaneously.[5] This point underscores an oft-quoted social science maxim: Correlation does not imply causation. For example, a high volume of ice cream sales does not cause high temperatures, although the two measures are correlated.

[5] You can attempt, through complex statistical techniques such as partial correlation, regression analysis, and modeling of simultaneous equations, to make inferences about causality by examining complex patterns of associations between variables. The observed data, nevertheless, do not directly support a causal logic. You are still making causal inferences based on empirical patterns of association.

You also can try to get more directly at causes by using research designs that measure the predictor at an earlier time frame than they measure the outcome. See Chapter 14 on longitudinal research.

Of course, when you are operating in Einstein's mode of inquiry you interpret the associations you observe *in light* of your hypotheses. If the associations revealed by the observed scores on the variables support your hypothesis, you may be justified in inferring that the causal logic embedded within your hypothesis is at work. But then again this justification may not be warranted. (See Box 3.2)

### Einstein, Holmes, and the Theorizing Process

The description I have presented above moves from the real world, to your head, and ends up in the world of data. As such it represents Einstein's logic of scientific inquiry. You move from the general to the particular, and from general expectations, to specific hypotheses, to actual data to test hypotheses. You proceed from the top of Figure 3.1 to the bottom.

It is possible, and perhaps even desirable, however, to approach the process of theorizing in exactly the opposite manner. You can use Holmes's grounded-theory approach and proceed from the bottom up. In this process you move from the world of data, to your head, to the real world.

You would begin at the bottom of Figure 3.1, with data that have been collected. You examine particular measured variables and the data collection process. You scrutinize the observed correlations between variables. Based on the measured variables and observed correlations, you make estimations about the variables before measurement. You may use procedures to try to remove error introduced through the process of measurement. The associations you observe permit you to derive hypotheses about how the variables relate to one another. You have moved from the fourth row in the figure up to the third row, and are ready to generate theory.

CHAPTER 6

# MEASUREMENT AND INQUIRY

*Measurement enables the tool of mathematics to be applied to science.[1]*

## OBJECTIVES

An extremely crucial part of scientific inquiry is *measuring*. This process allows researchers to cross over from the safe territory of speculation to the harsh land of empirical verification. In this chapter you will master some fundamentals of how social scientists measure, what they measure, and the logic behind their actions.

Imagine you wake up one morning and find that none of the clocks in your house tells the usual time. Their faces simply have the message "It is now daytime." You look at the thermometer outside your window to check on the weather. All the numbers are gone. It has a message: "The temperature is not like it was yesterday." Suspecting that something has gone seriously awry somewhere, and regretting the *Twilight Zone* marathon you watched on cable the night before, you do not listen to the morning news or read the morning paper. You dress hurriedly, get in your car, and drive to work. The gas gauge in your car has no needle, no "E" or "F," but it does display the message "Yes, you do have some gas in the tank. Not sure how much." As you drive along to work, your speedometer simply displays the word "Moving" when you are moving, and "Stopped" when you are stopped. There are no numbers or needle. When you get to work and put money in a parking meter, it tells you "Not expired." But it does not tell you how much time you have.

## PURPOSE AND ORGANIZATION OF THE PRESENT CHAPTER

The scenario above depicts what life would be like if measurement did not exist.

> The purpose of the measurement process is to assign observed scores on variables for particular objects, and to understand the mathematical properties of those observed scores. The observed scores assigned describe the relative amounts of a variable or attribute possessed by specific objects. [2]

Without measurement we would be unable to learn how much of something this or that object had; we would not know how much gas there was in the tank, or how much time we had in the parking meter.

In everyday life we take the results of measurement for granted: time, speed, temperature, cost, and so on. If we did not have reports available to us on these matters, our days would be full of much more uncertainty than they already are!

The position of criminal justice researchers, and other social scientists, is similar to that in the above scenario. They are stranded in a land where the time, temperature, and speed of the cars are not known.

They must create, often from scratch, various measurements of the indicators of concepts of interest to them. The process of deciding what and how to measure represents an integral part of the scientific enterprise.

Measurement is *not* simply concerned with the "nuts and bolts" or "engineering" side of criminal justice research. The process of measurement intertwines with the processes of theoretical elaboration, hypothesis testing, and grounded theorizing. [3] Without measurement processes we cannot connect the constructs in a theory to observable data.

In Chapter 3 you learned about the importance of *operationalizing* concepts and choosing *empirical indicators* for constructs. In the last row of Figure 3.1 you obtained measurements, in the form of observed scores, on the variables of interest to you for specific objects. This chapter explains in more detail the process and consequences of moving from the third to the fourth row of Figure 3.1.

In this chapter you begin by reflecting on the transition from variables that exist in our head to variables that exist in the world of observed data. You then consider what we measure when we measure. You may be surprised to discover that criminal justice researchers do not measure what you might have expected. Next you will explore the properties and consequences of different levels of measurement. You will find that, from a mathematical point of view, not all variables are created equal. Then you will explore the properties of scales and indexes, and examine some indexes and scales widely used in criminal justice.

### Return to the Electronic Monitoring Scenario

In many examples here I will refer to the running scenario, introduced in Chapter 1, concerning offenders sentenced to home confinement with electronic monitoring (HCEM). You may recall that, after two embarrassing incidents, your boss ordered you to investigate the cases sentenced to home confinement with electronic monitoring.

You decide that one way to approach the project is to gather, from court records, information about the offenders sentenced to HCEM. You also collect information from the state police. The latter records report if the offender was arrested for a subsequent offense.

## TEST THEORY

Your understanding of the measurement process will benefit from a short introduction to *classical test theory*. It is an area of measurement psychology that explains the different components making up the observed score of one individual on one variable.

Take another look at Figure 3.1 on page 41. Note that, going from the third row of the figure to the fourth row, you pass from variables as operationalized to variables as measured. With measured variables you obtain **observed scores** for particular cases studied. You go from thinking about specific variables to obtaining actual data for them.

Your variables as operationalized are different in important ways from your variables as measured. The process of measuring variables introduces **measurement error.** Measurement error arises from many sources, including the measuring instruments used, the measuring situation, the cases studied, and other factors. It results in nonsystematic discrepancies between how a case should have scored on a variable if the variable had been perfectly measured, and how it actually did score. Measurement error is unavoidable in social scientific research.

Think back to the time you took your SAT or ACT exams for college. Recall the first time you took the test. Think of all the factors leading up to the test that may have made you do less well than you thought you could: not enough sleep the night before, too much talking about it with classmates making you "psyched out" about the whole endeavor, or anxious parents deepening your own insecurities. Think also of the factors during the actual testing situation that may have made you do more poorly than you thought you should have done: someone sitting next to you nervously tapping a pencil, or a confused proctor who failed to give you the correct amount of time on a section of the test. All these factors added up to produce "measurement error"—a score that did not perfectly agree with your actual level of knowledge.

In short, in the third row of Figure 3.1 you are dealing with hypothesized **true scores** on variables. These true scores are hypothetical, unobserved constructs that exist only in your head, or in the head of others envisioning the same operationalized variables. As you proceed from the third row to the fourth row, the process of measuring these variables—obtaining actual scores on actual variables for studied cases using measuring instruments—introduces measurement error and results in observed scores.

This formulation linking true scores and observed scores has been called **classical test theory** [4]. It can be stated in the following equation:

Observed Score = True Score
+ Measurement Error

Turn your attention back to the offenders sentenced to HCEM. Imagine that, in your jurisdiction, all arrestees receive mandatory drug testing through urinalysis. The test used in this scenario detects drug use within the last 48 hours. One observed score produced by the urinalysis tells whether the arrestee tests positive for recent cocaine or crack-cocaine use. A range of factors could produce a discrepancy between the urinalysis results and the arrestee's actual behavior. Samples may have been mislabeled. Instruments at the testing lab may have needed recalibration. Chemicals used in the process may have been past the expiration date. These and other factors may have resulted in disparities between what happened—cocaine or crack was or was not used by the arrestee in the 48 hours before the testing—and the results of the urinalysis test. The discrepancy may occur for one arrestee or for several.

The amount of measurement error may vary across cases studied. One worker in a lab may be much more careful with testing procedures for Cases 1–10 than is a second worker, who is testing samples for Cases 11–20. In addition, the amount of measurement error also can vary across different observed variables.

Imagine that all arrestees in your jurisdiction not only undergo urinalysis, but also are interviewed about their drug usage habits. You might imagine that many arrestees who used cocaine or crack-cocaine in the 48 hours before arrest would not report it. The discrepancy between observed scores and true scores would be greater for these observed scores based on self-report than it would for the observed scores based on urinalysis. [5] In sum, measurement error varies across cases studied, variables examined, and time.

## WHAT DO WE MEASURE WHEN WE MEASURE?

### An Argument

Several years ago, some work colleagues and I were dining at a restaurant. The conversation became merry as the beer and wine bottles were emptied. At one end of the table the conversation took up the following topic. One colleague maintained that "Anything worthwhile can be measured scientifically, i.e., quantified." Another colleague countered that this was not so, and proceeded to propose items defying quantification. Concepts such as "love" and "hate" and "justice" were bandied about. Lively discussion followed about whether indicators of such concepts could be assigned numbers, and the meaning of numbers that would be assigned. Others around the table took sides with one colleague or the other. Take a minute or so and consider the matter. Which side do you think was correct?

## Defining Terms

To be as clear as possible about what our empirical indicators do and do not measure, we need to define our terms. [3]

1. **Objects or cases** refer to things that carry or possess properties. These include events, individuals, and items in the world about us: a crime, an arrest, a police officer, a judge, several judges, trials, bail hearings, a parole board, or a prison facility are all examples of objects. In our running scenario, the offenders sentenced to HCEM are objects. The dataset, introduced in Chapter 5, contains 100 officers trained in community policing; we call these objects.
2. **Attributes or variables** refer to potentially measurable properties of objects or cases. You can grade amounts of these properties; varying amounts can be arranged relative to one another. These are *features (Figure 3.1) that can be measured.*

   The offenders sentenced to HCEM in our running scenario have many attributes that might be of interest to you as you investigate why some offenders succeed on HCEM while others recidivate. The particular variables or indicators you choose will derive from or refer to concepts embodied in the theory you are testing or developing. Possible variables include: number of prior arrests, number of prior convictions, age, drug problems, age at first arrest, years of education completed, extent of prior criminal career, rehabilitation potential, and so on.
3. **Magnitudes** refer to *possible scores* reflecting "particular amounts of an attribute."[1] Three previous arrests, or an age of 24 years, or 8.5 years of school completed would all be examples of specific magnitudes. These are specific points or locations on different attributes. Magnitudes represent the *range of possible scores* on a variable, as well as *specific* scores within that range. These scores, however, have not yet been assigned to particular objects. In the police officer dataset used in Chapter 5, magnitudes for the variable "Satisfaction with the CPO Role" range from 1 to 10.
4. When a possible score on a variable (a magnitude) does apply to a specific object, we have a **quantity** of an attribute. It is a *particular* observed score on a *specific* variable for a *particular* case. In the dataset for the 100 officers trained in community policing, Case 1 reports a satisfaction score of "1"—this is a quantity. Quantities are the outputs produced by the empirical indicators used by the criminal justice researcher. These quantities allow the researcher to sort the objects investigated, whether that be people or institutions, on the underlying dimension. The dataset for 100 officers is sorted based on their satisfaction scores, with the lowest-scoring officers appearing first.

Observed scores on particular variables, such as satisfaction with the CPO role, for specific cases, are the end point of the measurement process and provide the basic "inputs" for many types of research analyses.

---

[1] Do not confuse the term *magnitude*, defined here, with a procedure called *magnitude scaling*, defined later in the chapter.

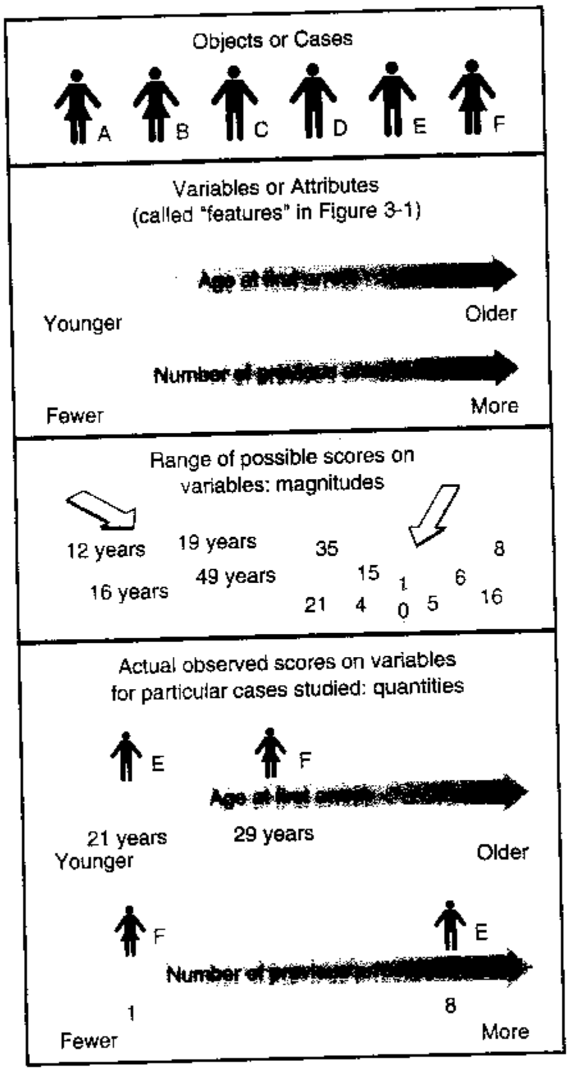


**FIGURE 6.1** Terms used in measurement process.

Figure 6.1 provides a graphical explanation for the definitions of these different components of the measurement process.

## Thinking of the Measurement Process in These Terms

Thus, the measurement tasks confronting the researcher involve:

- Selecting cases (objects) to be studied.
- Identifying measurable features or variables (attributes) of those objects. Each variable represents an underlying continuum or dimension along which the objects can be sorted or scored.
- Developing measuring instruments to assess possible scores (magnitudes of attributes) on these variables.
- Collecting information on actual observed scores of actual objects on the variables selected. These are the actual quantities of the attributes assigned to particular objects.
- Understanding the mathematical properties of these observed scores.

Let me try to sum up an answer to the question "What do we measure when we measure?" *The process of measuring involves generating numbers (quantities) that describe the relative amounts of an observed attribute (an observed variable) possessed by specific objects.* [2]

Let's go back to our restaurant party question. Can we scientifically measure any thing?

In simple terms: no. We can never measure criminals or crimes or institutions or courts or parole boards or parole officers or judges; or love or justice or equity or due process. Measurement refers to measuring *attributes* or features of objects. Measurement results in assigning observed scores on particular variables to particular cases. The cases *themselves* are not measured. We focus on specific features of these objects, and operationalize indicators of concepts as variables. The operational variables we define will pay attention to some features of these objects, and ignore others.

Selection of Variables to Indicate Concepts + Particular Cases Studied + Measuring Instruments → Observed Scores on Selected Variables for Selected Cases

It might help you to think of a person connected to a special device that only allows for listening to certain features of a piece of music. If listening to a classical symphony, for instance, the device allows the listener to attend to the tempo of the tympani, or the pitch of the notes played by the first violins, or the rhythm of the notes played by the flutes. But when using this device, the listener cannot hear the whole symphony at once. Only certain features can be heard at one time.

The measurement process proceeds like this device. It transmits to the researcher just observed scores on certain variables for the cases being studied. As a result, the researcher does not attend to all possible variables of the cases he is examining, and is not examining the cases themselves.

## CLASSIFICATION AND LEVELS OF MEASUREMENT

### Levels of Measurement

Returning to our scenario, you now have observed scores on a range of variables for these offenders sentenced to HCEM under your jurisdiction's current program. These different scores not only represent different indicators of concepts. They also possess different mathematical properties. The different mathematical properties limit what you can and cannot do with observed scores. The *difference* between the four levels of measurement you will examine is as follows: *they vary in the amount of information they provide about the quantities of the attribute they reflect.* [6] These distinctions may arise from one of two sources: the nature of the attribute itself, or the nature of the measuring instrument.

We can never know empirically the "true" quantity of an attribute possessed by an object, or the "true" quantities of an attribute possessed by a series of objects. We can theorize about these "true scores." As shown in Figure 3.1, these "true scores" exist in your head after you have decided how you will operationalize an indicator of a concept. When you actually *measure*, you cross over to the data world and you obtain *observed scores* on variables for those particular cases.

Consider, for example, that you are interested in how *serious* or *severe* the crimes are for which your offenders sentenced to HCEM were convicted. You hypothesize that each crime possesses a certain "true" quantity of the attribute of *seriousness* or *severity*. You could obtain *measured* quantities of seriousness in several different ways. You could ask a friend to order the crimes, ranging from most serious to least serious. You could ask 10 police officers to do the same and take the middlemost ranking for each crime as the measured quantity for each crime. Or you could have several thousand people across the nation rate them on a severity scale, as researchers did in a project to be described below. These are all different ways of coming up with *measured* quantities that do, you hope, correspond with the *true* quantities of seriousness. All measures, whatever their "level," are trying to achieve a correspondence between observed scores and "true" scores. What varies across the levels of measurement is the degree and type of correspondence.

Let's discuss those four different levels of measurement.[2]

**Nominal Level of Measurement** Nominal measures are really *classifications* of objects into one category or another.[3] The other

[2] There are more than the levels of measurement discussed here. For example, there are also partial orderings, and orderings with natural origins. I do not discuss those and other levels of measurement here, simply because they are not widely noted.

[3] There are strong differences of opinion among measurement experts on whether nominal measures are really measurements at all. S. S. Stevens says they are. Warren Torgerson says they are not. Stevens says they are because you can assign scores to variables measured at the nominal level. Torgerson says they are not because the scores you assign are arbitrary. Personally, I agree with Torgerson on this matter. In deference to the traditional pedagogical approach in this matter, however, I include nominal scales as one of the four levels of measurement.

three levels of measurement, which you will encounter shortly, include variables signifying *how much* of an attribute the object possesses. Nominal measures simply report to which category an object belongs.

In categorizing or classifying objects, nominal measures are both exhaustive and mutually exclusive. They are exhaustive because all objects are sorted into one category or another. They are exclusive because an object cannot simultaneously have two categories of the attribute.

The information you have collected on offenders sentenced to HCEM contains several variables measured at the nominal level. For example, you might have information from a presentence investigation (PSI) reporting if the offender has a drug problem. From this information you can construct a variable "Has a drug problem." If the PSI does not mention a drug problem, the case scores 0 on this variable. If the PSI does mention a drug problem, the case scores 1 on the variable. The cases scoring 1 on this variable are not *higher* on the attribute than the cases scoring 0. They are simply in a different category; they have the attribute, and those scoring 0 do not.

Your attribute "Has a drug problem" has been coded as a **dummy variable**. Nominal measures coded 0 or 1 are dummy variables. Criminal justice and sociological researchers use them often. In the example here you measure the attribute in question with one dummy variable. Sometimes, however, you need more than one dummy variable to measure an attribute. Ethnicity is a case in point.

The files of offenders on HCEM may include members of three racial groups: Hispanics, African Americans, and Whites. If you are particularly interested in Hispanics as compared to others, and African Americans as compared to others, you can construct two dummy variables for each case. Hispanics are coded 1 on the "Hispanic" dummy variable and 0 on the "African American" dummy variable. African Americans are coded 0 on the "Hispanic" dummy variable and 1 on the "African American" dummy variable. Whites are coded 0 on both variables. See Table 6.1.

Nominal measures and classifications, despite being simpler than the other levels of measurement, can play important roles in theory construction. Classifying cases into different "types"—groups that are qualitatively different from each other—and describing the factors associated with each type is a legitimate and important theoretical enterprise. For example, in the scenario here you may attempt to describe three "types" of offenders on HCEM: those who fail by committing a new crime and are rearrested, those who fail only because they violate the technical conditions of the sentence and consequently get sentenced to incarceration, and those who succeed. You might argue that there are fundamental *qualitative* differences between the three groups.

**Ordinal Level of Measurement** *An Ordinal Scaling Process Ranks or Orders Cases (a Set of Objects) on a Variable (an Attribute)* The magnitudes or observed scores assigned are: the object ranked 1 has more of variable A than the object ranked 2, the object 2 has more of variable A than the object ranked 3, and so on. The rankings show which object, relative only to the other objects in the series, has more of variable A.

For example, with your offenders sentenced to HCEM you might be interested in characterizing the *level of prior criminal activity*. You want to know how court personnel who prepare PSIs *perceive* this attribute. You could select 20 offenders sentenced to HCEM

**TABLE 6.1**
**Scores on Dummy Variables Capturing Ethnicity for Different Offenders**

| Case Number | Ethnicity | Dummy Variables and Observed Scores | |
|---|---|---|---|
| | | **Hispanic** | **African American** |
| 27 | African American | 0 | 1 |
| 47 | Hispanic | 1 | 0 |
| 63 | White | 0 | 0 |

and ask several court personnel to review the files and *rank order* the cases on the variable: "Level of prior criminal activity." The observed scores on the variable generated by this process will be **ranks**. Each court worker, acting as a rater, will assign a ranking of 1 (highest level of prior criminal activity) to 20 (lowest level of prior criminal activity) to the 20 cases.

Variables measured at the ordinal level have the following measurement properties:

- Only the *relative* quantities of the variable (attribute) possessed by the objects are indicated (**ordering property**).
- Since you only know about relative observed scores or quantities, you have no information about the *specific absolute level* of an observed score.
- Thus you don't know if the difference in level of prior criminal activity between the offender ranked sixth and the offender ranked seventh is miniscule or gigantic compared to the difference between the offender ranked eighth and the offender ranked ninth. Therefore, you cannot compare differences in ranks.
- Since ordinal scales are crude, there may be *true* differences in the quantities of an attribute possessed by two objects, but the ordinal scale may not reflect that. For example, Rater R-1 may rank the level of prior criminal activity of offenders A and G as equally high. He may say they are tied for first place, and assign both the rank of 1. But this tied observed score may not reflect actual differences on the dimension of interest.

*Summary* With ordinal measurement we can order or rank objects on an underlying dimension or attribute. The observed scores on the variable merely sort the objects on the variable. You cannot compare differences in ranks.

**Interval Level of Measurement** If quantities of an attribute are measured on an interval scale, then we know, besides the ordering of the objects on the attribute, two more points.

First, observed scores measured at the interval level of measurement have a linear relationship with the "true scores" on the variable. For each unit increase of "true scores," observed scores increase a constant amount. Because of this relationship we also


"Now! *That* should clear up a few things around here!"

**PICTURE 6.1 What level of measurement is represented here?** (*Source.* THE FAR SIDE © 1989 FarWorks, Inc. Distributed by Universal Press Syndicate. Reprinted with permission. All rights reserved.)

know something else about interval-level measures.

You can interpret the *differences* between the observed scores of different objects. The difference between any pair of observed quantities, when measured on an interval scale, can be calculated, and compared to the difference observed between any other pair of observed quantities. The difference between every pair of scores will be equal to, less than, or greater than the difference between every other pair of observed scores. This measurement property is called the **distance property**.

Consequently, mathematical operations can be meaningfully applied to differences between observed scores on a variable measured at the interval level. The differences between pairs of observed scores can be discussed as statements about differences in the underlying "true scores."

For example, in the jursidiction in which you work, information from all arrestees is used to complete an inventory. Its purpose is to predict the chances that an arrestee will commit a serious crime if released on bail. The inventory has been extensively tested on many arrestees. Scores on the inventory range from 0 to 100 and have been standardized, so that scores of 50 have the same meaning in different groups of arrestees, scores of 75 have the same meaning in different groups of arrestees, and so on. It is a standardized, widely used inventory.

Since observed scores on this variable represent interval level measurement, you can meaningfully discuss differences between pairs of scores. In your sample of offenders sentenced to HCEM you may find that Case A scores 10 on the inventory, Case B scores 30, Case C scores 30 and Case D scores 50. The difference between the scores of A and B is equal to the difference between the scores of C and D.

**Ratio Level of Measurement** Variables measured at the ratio level of measurement possess all the features of ordinal measures (ordering) and interval measures (ordering and distance), plus one additional characteristic. *The variable has a unique zero point or natural origin.* Consequently there is an even closer relationship between measured quantities and true quantities of an attribute. Not only do observed scores increase linearly as true scores increase. You also know that, when the observed score = 0, barring problems of measurement error, the "true score" = 0.

Thus, we can now apply mathematical operations to the *measured quantities directly.* For example, if Case A has six prior felony convictions and Case B has two prior felony convictions, you can meaningfully say that A has three times more prior felony convictions than B.

Criminal justice abounds with a range of ratio scaled quantities. To cite just a few: number of victimization incidents in the last 12 months, number of previous arrests, length of incarceration in months, number of technical parole violations, length of probation in months, number of cases handled by a district attorney's office in a month, and so on. As a very rough rule of thumb, if a measured quantity refers to a *count* of criminal justice events, or the *length of time* of a criminal justice event, it is probably measured at the ratio level of measurement.

**Summary** Figure 6.2 summarizes what we know about the types of possible relationships between "true scores" on a variable and observed scores on a variable, at the three highest levels of measurement. Each panel displays two possible relationships between true scores and observed scores at that level of measurement. Each graph assumes that observed scores include only a modest amount of measurement error. Figure 6.3 summarizes the properties of observed scores for variables at all four different levels of measurement. You need to be able to identify the level at which a quantity is measured, so that you can know whether the operations the researcher is applying to the numbers in hand are correct or not.

You are undoubtedly wondering if a researcher can *change* the level at which an attribute is measured, after collecting the data. For example, if you have observed scores on a variable measured at the ordinal level, can you change observed scores on the variable into an interval level of measurement? Or, can observed scores measured at the interval level be transformed into observed scores measured at the ordinal level?

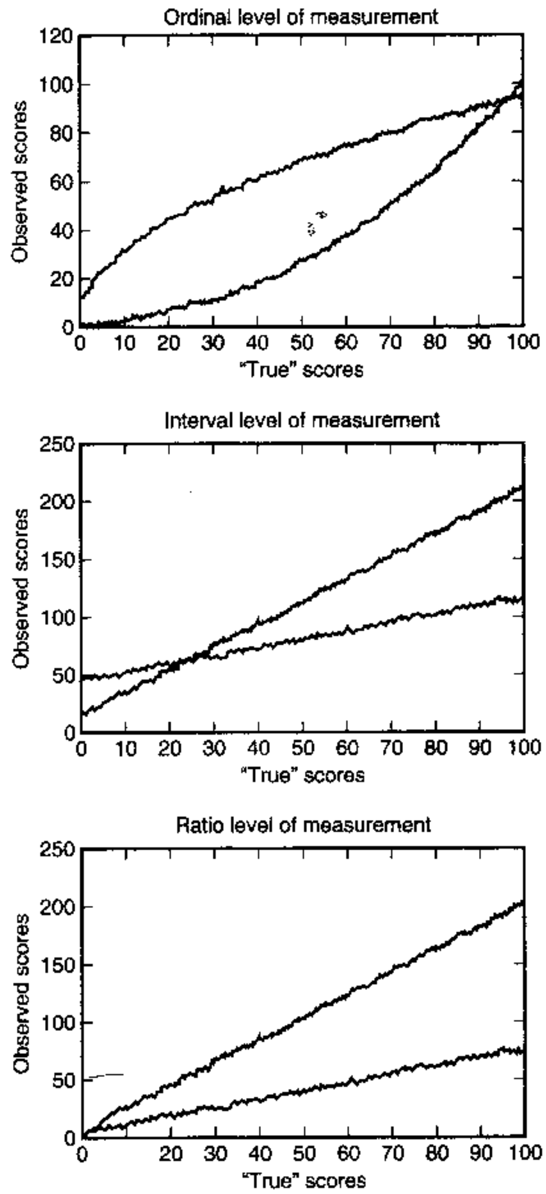


**FIGURE 6.2**
Possible relationships between observed scores on a variable, and "true scores" on a variable, under three levels of measurement: ordinal (top), interval (middle), and ratio (bottom). Each panel graphs two possible different relationships at that level of measurement. Measurement error, for each set of observed scores, is assumed to be minimal.

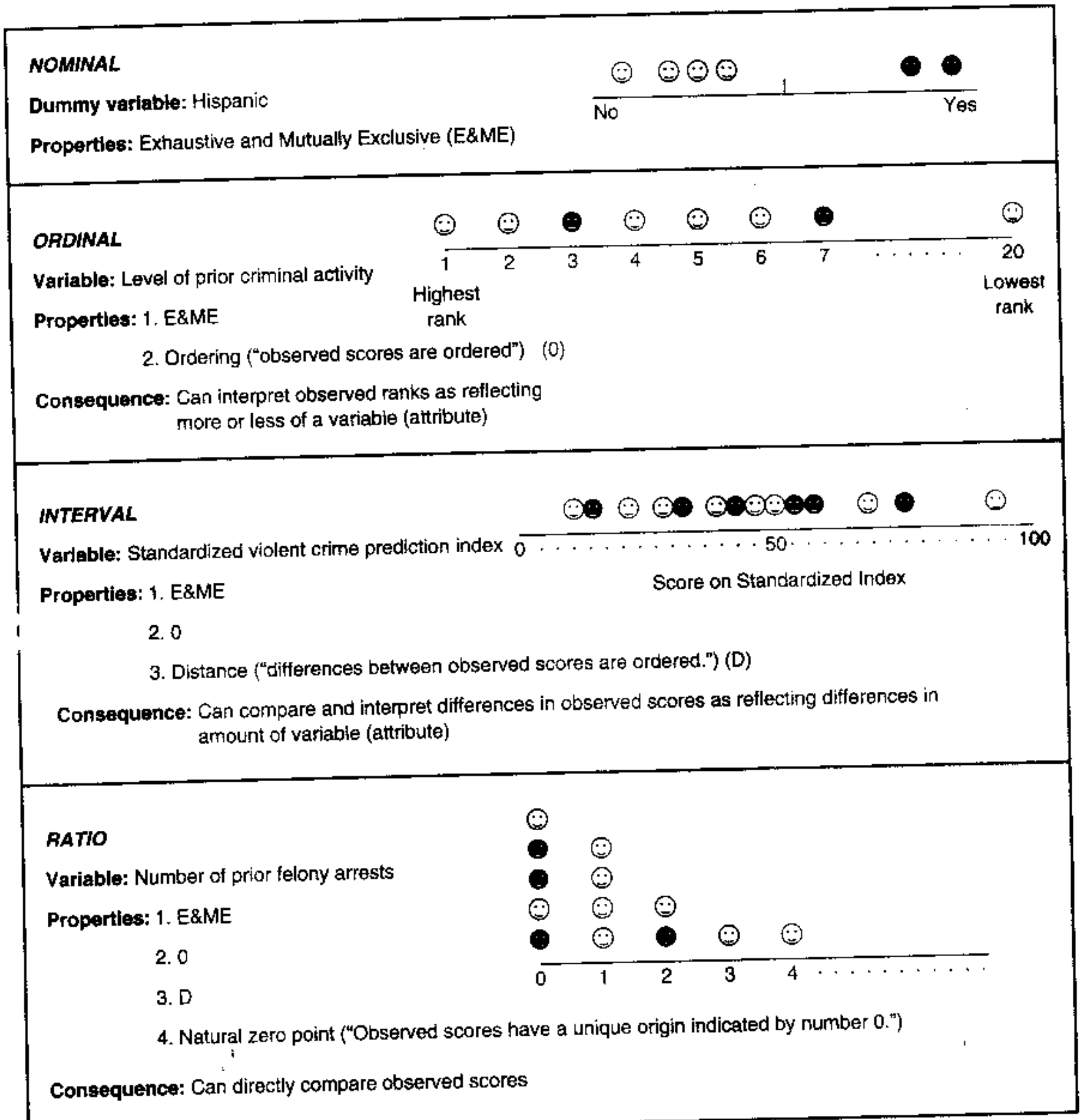


**FIGURE 6.3**
Properties of observed scores of variables at different levels of measurement. Quotes from [1].

Yes, such transformations are permissible, but only in one direction. You can convert observed scores on a variable from higher levels of measurement into lower levels (e.g., ratio to ordinal). You cannot, however, do the reverse. The information required for the latter treatment is simply not in the data.

## INDEXES AND SCALING

In the remainder of the chapter you will examine two specific *types* of operational indicators—indexes and scales. **Indexes** are summary indicators, comprised of more than one variable. As a researcher you do not make distinctions between the different variables that make up the index. Most often, each variable you use in the index contributes about equally to the index summary score. Scores on an index represent a theoretical construct more general than scores on a single observed variable. After you construct an index you can array cases along the general variable or attribute tapped by the index.

The goal of **scale construction** is also to sort out the people being studied on the variable or attribute of interest. But the process involves two steps. First, the researcher learns the relationship of different items to one another. Then, based on their responses to particular items, people or items or both can be sorted on the variable. In short, with scales the researcher has to first *order* or *organize* the items being used before beginning to collect people's responses to the items. You will examine this process more closely later in the chapter.

### Test Theory and Multiple Variables

Your insight into the conceptual foundations of scale and index construction will benefit if we spend more time with test theory. The equation on p. 102 showed that an observed score on one item reflected both a true score and measurement error. You can expand this equation to a situation where you have more than one variable.

As noted above, in the HCEM scenario, arrestees in your jurisdiction receive mandatory drug testing. Tests examine for four different substances: cocaine or crack-cocaine, heroin, amphetamines ("speed"), and marijuana. Each arrestee's urine sample is tested for each of these drugs, resulting in four different test results. Arrestees could test positive for any one of the four drugs.

You want to examine the drug-testing scores of convicted offenders assigned to HCEM. You want to develop one overall variable for the offenders showing the extent of their drug problem. You conceive of a construct: general involvement in drug abuse. You expect that you could operationalize this indicator of a concept by adding up the results of the different drug-testing results to produce one general drug test score.

Operationally, this is easy. You have four different drug test results for each offender. Each observed score is a dummy variable at the nominal level of measurement, showing drug positive or not (1 or 0), for each drug type, for each case. You add the observed scores on each case to obtain scores on an index "General Drug Involvement."

When you construct this index you are making some assumptions about each variable contributing to the index. By postulating the concept "General Drug Involvement" and choosing each of these four variables to contribute to an indicator of it, you assume that the true score of each variable has two portions. One portion is specific to the drug tested. The other emerges from a tendency toward or against general drug involvement; the first represents the *specificity* of the item, the second reflects what is *shared* between the specific concept (e.g., cocaine use) and the more general concept (general drug involvement). Therefore, if you are thinking of spe-