4 Planning Your Research Project

Before the construction of a building, an architect develops a detailed set of plans. These plans enable the builder to erect a strong, well-designed structure. Researchers should be no less detailed and precise in the planning of a research design. Plans, specifications, and design: All of these serve well the architect, the builder, and the researcher alike.

Architectural planning and research planning have much in common. Each requires a conceptualization of the overall organization of a project and a detailed specification of the steps to be carried out; only after such meticulous planning can work on the project actually begin. For successful completion, a building requires plans that are clearly conceived and accurately drawn. A research project should be no less completely conceptualized and precisely detailed.

Planning a General Approach

When we talk about a general strategy for solving a research problem, we are talking about a **research design**. The research design provides the overall structure for the procedures the researcher follows, the data the researcher collects, and the data analyses the researcher conducts. Simply put, research design is *planning*.

Nothing helps a research effort be successful so much as carefully planning the overall design. More research effort is wasted by going off half-prepared—with only a vague set of ideas and procedures—than in any other way. You will be much more efficient and effective as a researcher if you identify your resources, your procedures, and the forms that your data will take—always with the central goal of solving your research problem in mind—at the very beginning of your project.

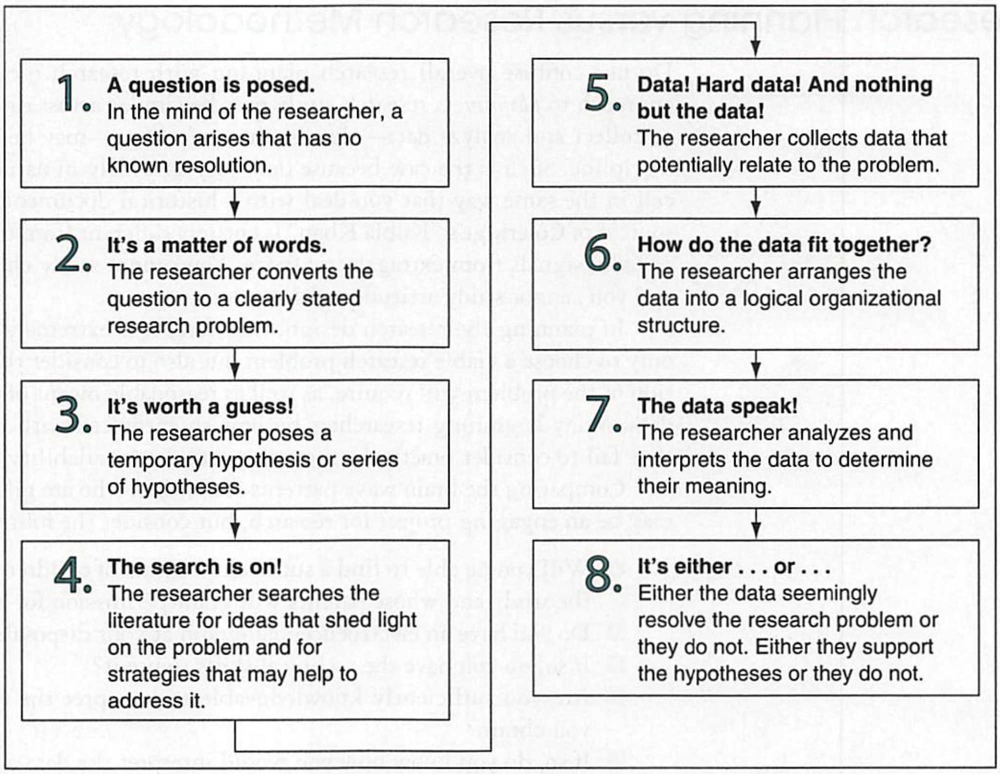
The Basic Format of All Research

The research process follows a basic format. No matter which academic discipline gives rise to the research endeavor, the general research procedure is fundamentally the same. Although you will sometimes hear people refer to "social science research," "nursing research," or "marketing research," in fact the search for information to solve a problem or to answer a question seldom fits squarely within the bounds of a single academic discipline.

In planning a research design, the researcher in quest of new knowledge cannot be shackled by discipline-specific methodological restraints. The course of a research project will frequently lead the investigator into new and unfamiliar territories that have historically been associated with other content areas. The sociologist trying to resolve a problem in sociology may come face to face with problems that are psychological or economic. The educational researcher exploring the causes of a learning disability may need to consider the domains of neurophysiology, psychopathology, endocrinology, and family counseling. On the way to finding a solution for a problem in criminology, the student in criminal justice may venture into the realms of abnormal psychology and behavioral genetics. Any good researcher must be *eclectic,* willing to draw on whatever sources seem to offer productive methods or evidence for resolving the research problem.

Figure 4.1 presents the basic format for the research process. It is important to note, however, that a research project does not always follow these steps in the exact sequence depicted

FIGURE 4.1
The basic format of the research process



here. Especially in the case of qualitative research—an approach we describe later in the chapter—a researcher may order the steps somewhat differently. For instance, in some forms of qualitative research, the researcher formulates a hypothesis and reads much of the literature only *after* collecting a substantial amount of data (more on this point in Chapter 6).

Instead of limiting their thinking to departmentalized knowledge, researchers might do much better to think of problems as arising out of broad generic areas within whose boundaries all research falls: people, things, records, thoughts and ideas, and dynamics and energy. Let's briefly consider some research problems that may fall within each of these areas.

- *People.* In this category are found research problems relating to children, senior citizens, families, communities, cultural groups, ancestors, employees, management, mental and physiological processes, learning, motivation, social and educational problems, crime, rehabilitation, medical treatments, nutrition, language, religion.
- *Things.* In this category are found research problems relating to animal and vegetable life, viruses and bacteria, inanimate objects (rocks, soil, buildings, machines), matter (molecules, atoms, subatomic matter), stars, galaxies.
- *Records.* In this category are found research problems relating to newspapers, journals, letters, manuscripts, registers, speeches, interviews, minutes, legal documents, financial and corporate statements, census reports, archeological remains, sketches, drawings, paintings, music.
- *Thoughts and ideas.* In this category are found research problems relating to concepts, theories, perceptions, opinions, beliefs, reactions, issues, semantics, poetry, cartoons, caricatures.
- *Dynamics and energy.* In this category are found research problems relating to human interactions, metabolism, chemical reactions, radiation, radio and microwave transmissions, quantum mechanics, thermodynamics, hydrodynamics, hydrologic cycles, atomic and nuclear energy, wave mechanics, atmospheric and oceanic energy systems, solar energy, black holes.

We do not intend the preceding lists to be mutually exclusive or all-inclusive. We merely present them to give you an idea of the many research possibilities that each category suggests.

Research Planning versus Research Methodology

Do not confuse overall research planning with research methodology. Whereas the general approach to *planning* a research study may be similar across disciplines, the techniques one uses to collect and analyze data—that is, the *methodology*—may be specific to a particular academic discipline. Such is the case because data vary so widely in nature. You cannot deal with a blood cell in the same way that you deal with a historical document, and the problem of finding the sources of Coleridge's "Kubla Khan" is entirely different from the problem of finding the sources of radio signals from extragalactic space. You cannot study chromosomes with a questionnaire, and you cannot study attitudes with a microscope.

In planning the research design, therefore, it is extremely important for the researcher not only to choose a viable research problem but also to consider the kinds of data that an investigation of the problem will require, as well as reasonable means of collecting and interpreting those data. Many beginning researchers become so entranced with the glamour of the problem that they fail to consider practical issues related to data availability, collection, and interpretation.

Comparing the brain wave patterns of children who are gifted versus those of average ability may be an engaging project for research, but consider the following issues:

- Will you be able to find a sufficient number of children who are willing to participate in the study and whose parents will grant permission for their children to participate?
- Do you have an electroencephalograph at your disposal?
- If so, do you have the technical skills to use it?
- Are you sufficiently knowledgeable to interpret the electroencephalographic tracings you obtain?
- If so, do you know how you would interpret the data and organize your findings so that you could draw conclusions from them?

Unless the answer to all of these questions is *yes,* it probably is better that you give up this project in favor of one for which you have the knowledge, skills, and resources to carry through to completion. Your research should be *practical* research, built on precise and realistic *planning* and executed within the framework of a clearly conceived and feasible *design.*

General Criteria for a Research Project

A good research project has four important qualities you can use as criteria in evaluating your own research plan. Here, briefly, are those qualities.

Universality. The research project should be one that might be carried out by *any* competent person. As a researcher, you are a director and producer—an agent whose function is to collect, organize, and report what the collected data seem to indicate—but another, equally knowledgeable individual might take your place and complete the project with essentially the same results. (We will find an exception to this criterion later in the chapter, when we discuss the nature of data collection in qualitative research.)

Replication. The research should be repeatable. Any other competent researcher should be able to take your problem and, collecting data under the same circumstances and within the same parameters you have used, achieve results comparable to those you have obtained.

Control. The researcher must in some way isolate, or *control,* those factors that are central to the research problem. Control is important for replication: An experiment should be repeated under the identical conditions and in the identical way in which it was first carried out. Control is also important for consistency within the research design. For instance, if you want to compare the effects of two different treatments on some other factor, then you should keep everything else (aside from the specific treatments you are studying) as similar as possible.

Control is more easily achieved in some areas than in others. In the physical sciences, for example, such factors as temperature, pressure, electrical potential, humidity, and the like are highly controllable. Control is more challenging—but certainly possible—in research areas concerned with human behavior.

Control is especially important in experimental research designs. Accordingly, we talk more about this issue in Chapter 9.

Measurement. The data should be able to be measured in some way. This, again, is often easily accomplished in the physical sciences. Measurement is typically less precise and less accurate in the humanities and social sciences. Later in the chapter we discuss strategies for enhancing measurement procedures in these areas.

The Nature and Role of Data in Research

Research is a viable approach to a problem only when data exist to support it. The term *data* is plural (singular is *datum)* and comes from the past participle of the Latin verb *dare,* which means "to give." Data are those pieces of information that any particular situation *gives* to an observer.

What Are Data?

Researchers must always remember that data are not absolute reality—the pure, naked Truth that underlies all the phenomena we observe. Rather, data are *manifestations* of that reality. No one has ever looked upon Truth itself. We are like those individuals who live in a dungeon, across the floor of which a beam of sunlight passes. That light gives us an idea of what the sun must be like, but if we can never behold the sun, we shall never know the difference between it and the shaft of light on the dungeon floor.[1]

The researcher is in a factual dungeon. He or she will never be able to see the original source of the data. For instance, we often see what other people do—the behaviors they exhibit, the things they create, and the effects of their actions on others. But the actual people "inside"—those individuals we will never know!

Research seeks, through data, to discover underlying truths. Yet such is probably an endless pursuit. Experienced researchers are constantly aware that the Truth they most ardently seek is forever just beyond what is represented by the data and, hence, just beyond human grasp. For instance, the scientist probing the nature of subatomic matter may detect tiny entities that make up larger bits of matter. Yet the scientist knows, too, that such entities are probably made up of sub-sub-subentities that beckon yet evade measurement and investigation.

The mind yearns to understand the Truth. To pursue that goal, we have chosen the path of research. But the path always ends at the farthest reaches of the data, which are at the brink of the canyon in whose depths lies the inaccessible, ultimate Truth.

Data Are Transient and Ever Changing

Whenever we look at data analytically, we gain new insights. But at the same time, we also discern new problems that demand further research.

Data are not only elusive but also transient. Data that the researcher is permitted to glimpse may exist for only a split second. Consider, as an example, a sociologist who plans to conduct a survey in order to learn about people's attitudes and opinions in a certain city. The sociologist's research assistants begin by administering the survey in a particular city block. By the time they move on to the next block, the data they have collected are already out of date. Some people in the previous block who indicated that they held a particular opinion have now changed their minds and have a somewhat different opinion. They may have seen a television program or heard a discussion that has changed it. Some people have moved away, and others have moved in; some

[1]For readers interested in philosophy, our dungeon analogy is based loosely on Plato's Analogy of the Cave, which he used in Book VII of *The Republic*.

have died, and others have been born. Tomorrow, next week, next year—what we thought we had "discovered" may have changed completely.

Thus is the transient nature of data. We catch merely a fleeting glance of what seems to be true at one point in time but is not necessarily true the next. Even the most carefully collected data may have an elusive quality about them; at a later point in time they may have no counterpart in reality whatsoever. Data are volatile: They evaporate quickly.

Primary Data versus Secondary Data

The researcher's only perceptions of Truth are various layers of truth-revealing facts. In the layer closest to the Truth are **primary data**; these are often the most valid, the most illuminating, the most truth-manifesting. Farther away is a layer consisting of **secondary data**, which are derived not from the Truth itself, but from the primary data.

Earlier we used the analogy of the researcher as someone who sits in a dungeon and tries to understand the sun only by looking at a shaft of sunlight that falls on the floor. This direct beam of sunlight represents the *primary data.* Although the shaft is not the sun itself, it has come directly from the sun. But now imagine that the imprisoned researcher sees the sunlight, not as a direct beam, but as a pattern of shimmering light on the floor. The sunlight (primary data) has fallen onto a shiny surface and then been reflected—distorted by the imperfections of the shiny surface—to an image that is in some ways similar but in other ways dissimilar to the original shaft of light. This reflection of light is secondary data.

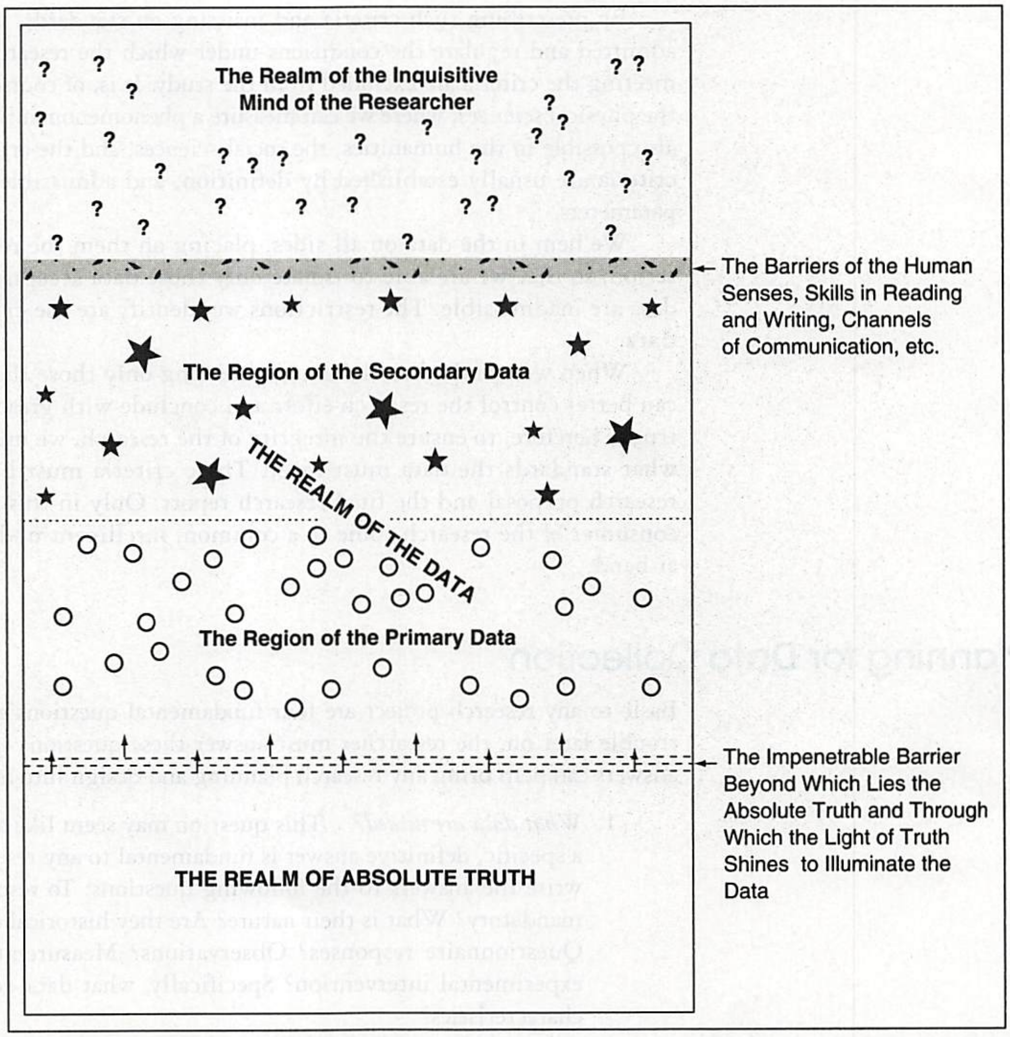
As another example, consider the following incident: You see a car veer off the highway and into a ditch. You have witnessed the entire event. Afterward, the driver says he had no realization that an accident might occur until the car went out of control. Neither you nor the driver will ever be able to determine the absolute truth underlying the accident. Did the driver have a momentary seizure of which he was unaware? Did the car have an imperfection that the damage from the accident obscured? Were other factors involved that neither of you noticed? The answers lie beyond an impenetrable barrier. The true cause of the accident may never be known, but the things you witnessed, incomplete as they may be, are primary data that emanated directly from the accident itself.

Now along comes a newspaper reporter who interviews both you and the driver. The reporter writes an account of the accident for the local paper. When your sister reads the account in the paper the next morning, she gets, as it were, the sunlight-reflection-on-the-floor version of the event. The newspaper article provides secondary data. The data are, of necessity, distorted—perhaps only a little, perhaps quite a bit—by the channels of communication through which they must pass to her. The reporter's writing skills, your sister's reading skills, and the inability of language to reproduce every nuance of detail that a firsthand observation can provide —all of these factors distort what you actually observed.

What we have been saying about data and their relation to Truth is represented by Figure 4.2. Lying farthest away from the researcher—and, hence, least accessible—is The Realm of Absolute Truth. It can be approached by the researcher only by passing through two intermediate areas that we have labeled The Realm of the Data. Notice that a barrier exists between The Realm of Absolute Truth and The Region of the Primary Data. Small bits of information leak through the barrier and manifest themselves as data. Notice, too, the foggy barrier between The Realm of the Data and The Realm of the Inquisitive Mind of the Researcher. This barrier is comprised of many things: the distortions and insensitivities of the human senses, the weaknesses of instrumentation, the inability of language to communicate thoughts exactly, the inability of two human beings to witness the same event and report it in precisely the same way, and so on.

Researchers must never forget the overall idea underlying Figure 4.2. Keeping it in mind can prevent them from making exaggerated claims or drawing unwarranted conclusions. No researcher can ever glimpse Absolute Truth, nor can anyone even perceive the data that reflect that Truth except through imperfect senses and imprecise channels of communication. Such a humiliating awareness helps researchers be cautious in the interpretation and reporting of research findings—for instance, by using such words and phrases as *perhaps, it seems, one might conclude, it would appear to be the case,* and *the data are consistent with the hypothesis that.*

FIGURE 4.2
The relation between data and Truth



Criteria for the Admissibility of Data

Not all data that come to the researcher's attention are acceptable for use in a research project. Data can be defective. If they are, they may affect the validity of the researcher's conclusions. The imperfections in the data stem from the imperfections and irregularities of nature. If researchers include in the mass of data those that are imperfect or irregular—those that are distorted, inaccurate, hopelessly entangled with irrelevant variables, and so on—they corrupt the entire body of data.

As previously mentioned, any research effort should be *replicable;* that is, it should be able to be repeated by any other researcher at any other time under precisely the same conditions. To ensure such precision, certain criteria must be adopted, certain limits established, and certain standards set up that all data must meet in order to be admitted for study.

For example, imagine that an agronomist wants to determine the effect of ultraviolet light on growing plants. *Ultraviolet* is a vague term: It encompasses a range of light waves that vary considerably in nanometers. The agronomist must narrow the parameters of the data so that they will fall within certain specified limits. What precisely does she or he mean by *ultraviolet light?* Within what nanometer range will ultraviolet emission be acceptable? At what intensity? For what length of time? At what distance from the growing plants? What precisely does the researcher mean by the phrase "effect of ultraviolet light on growing plants"? All plants? A specific genus? A particular species? The agronomist must be quite specific about all of these things so that another researcher can achieve very similar conditions in replicating the study.

By prescribing such criteria and insisting on standards, we can control the types of data admitted and regulate the conditions under which the research effort proceeds. Any data not meeting the criteria are excluded from the study. It is, of course, much easier to control data in the physical sciences, where we can measure a phenomenon fairly objectively, but such control is also possible in the humanities, the social sciences, and the arts. In these latter disciplines, the criteria are usually established by definition, and admissible data must meet the definitive parameters.

We hem in the data on all sides, placing on them the restrictions of criterion after criterion, so that we are able to isolate only those data acceptable for our use. The rest of the data are inadmissible. The restrictions we identify are the **criteria for the admissibility of data.**

When we standardize the data, admitting only those that comply with our criteria, we can better control the research effort and conclude with greater certainty what appears to be true. Therefore, to ensure the integrity of the research, we must set forth *beforehand* precisely what standards the data must meet. These criteria must be set forth clearly in both the research proposal and the final research report. Only in this way can the researcher and the consumer of the research come to a common, intelligent understanding of the investigation at hand.

Planning for Data Collection

Basic to any research project are four fundamental questions about the data. To avoid serious trouble later on, the researcher must answer these questions specifically and concretely. Clear answers can help bring any research planning and design into focus.

1. *What data are needed?* This question may seem like an overly simple one, but in fact a specific, definitive answer is fundamental to any research effort. On a sheet of paper, write the answers to the following questions: To resolve the problem, what data are mandatory? What is their nature? Are they historical documents? Interview excerpts? Questionnaire responses? Observations? Measurements made before and after an experimental intervention? Specifically, what data do you need, and what are their characteristics?
2. *Where are the data located?* Those of us who have taught courses in research methodology are constantly awed by the fascinating problems that students identify for research projects. But then we ask a basic question: "Where will you get the data to resolve the problem?" Some students either look bewildered and remain speechless or else mutter something such as, "Well, they must be available *somewhere.*" Not *somewhere,* but *precisely where?* If you are doing a study of documents, where are the documents you need? At exactly which library and in what collection will you find them? What society or what organization has the files you must examine? Where are these organizations located? Specify geographically—by town, street address, and postal code! Suppose a nurse or a nutritionist is doing a research study about Walter Olin Atwater, whose work has been instrumental in establishing the science of human nutrition in the United States. Where are the data on Atwater located? The researcher can go no further until that basic question is answered.
3. *How will the data be obtained?* To know where the data are located is not enough; you need to know how you might acquire them. With privacy laws, confidentiality agreements, and so on, obtaining the information you need might not be as easy as you think. You may indeed know what data you need and where you can find them, but an equally important question is, How will you get them? Careful attention to this question marks the difference between a viable research project and a pipe dream.
4. *How will the data be interpreted?* This is perhaps the most important question of all. The three former hurdles have been overcome. You have the data in hand. But you must also spell out precisely what you intend to do with the data to solve the research problem or one of its subproblems.

Now go back and look carefully at how you have worded your research problem. Will you be able to get data that might adequately provide a solution to the problem? And if so, might they reasonably lend themselves to interpretations that shed light on the problem? If the answer to either of these questions is *no,* you must obviously rethink the nature of your problem. If, instead, both answers are *yes,* a next important step is to consider how you might best *measure* what you observe.

Identifying Good Measurement Instruments

As you zero in on the nature of the data necessary for solving your research problem, you will probably discover that you must pin down your observations by measuring them in some way. In some cases you will be able to use one or more existing instruments—perhaps an oscilloscope to measure patterns of sound, a published personality test to measure a person's tendency to be either shy or outgoing, or a rating scale that a previous researcher has developed to assess parents' childrearing practices. In other situations you may have to develop your *own* measurement instruments—perhaps a survey to assess people's opinions about welfare reform, a paper-and-pencil test to measure what students have learned from a particular instructional unit, or a checklist to evaluate the quality of a new product.

Measurement instruments provide a basis on which the entire research effort rests. Just as a building with a questionable foundation is unlikely to be safe for habitation, so, too, will a research effort employing faulty measurement tools provide little of value in solving the problem under investigation.

Defining Measurement

What exactly *is* measurement? Typically we think of measurement in terms of such objects as rulers, scales, gauges, and thermometers. In research, **measurement** takes on a somewhat different meaning:

> Measurement is limiting the data of any phenomenon—substantial or insubstantial—so that those data may be interpreted and, ultimately, compared to a particular qualitative or quantitative standard.

Let's look more closely at this definition. The first five words are *measurement is limiting the data.* When we measure something, we set a limit that constrains the data in some way. We erect a barrier beyond which those data cannot go. What is a foot, a mile, a pound? Each is a unit of measure governed by a numerical constraint: 12 inches constrain a foot; 5,280 feet, a mile; and 16 ounces, a pound.

Now let's look at the next six words: *of any phenomenon—substantial or insubstantial.* This phrase is all-inclusive. Nothing exists that the researcher cannot measure in some way. In some cases, observable physical entities are measured. These are **substantial** measurements; that is, the things being measured have physical substance, an obvious basis in the physical world. An astronomer measures patterns and luminosity of light in the night sky; a neurologist measures intensity and location of activity in the brain; a chemist measures the mass of a compound both before and after transforming it in some way. All of these are attempts to measure substantial phenomena. Some devices designed to measure substantial phenomena, such as high-powered telescopes and MRI machines, are highly specialized and used only in particular disciplines. Others, such as balance scales and tape measures, are applicable to many fields of inquiry.

We can also measure those things—if "things" they be—that are **insubstantial,** that exist only as concepts, ideas, opinions, feelings, or other intangible entities. For example, we might attempt to measure the economic "health" of business, the degree to which students have "learned," or the extent to which people "value" physical exercise. We seek to measure these intangibles, not with tape measures or scales, but with the Dow-Jones index, achievement tests, questionnaires, or interviews.[2]

We continue with the next seven words of our definition of measurement: *so that those data may be interpreted.* We cannot emphasize this point enough: Research involves not only the collection but also the interpretation of data—the transformation of data into new discoveries, revelations, and enlightenments.

Now we finish our definition: *and, ultimately, compared to a particular qualitative or quantitative standard.* A researcher must have a goalpost, a true north, a point of orientation. In research, we call these standards *norms, averages, conformity to expected statistical distributions, goodness of fit, accuracy of description,* and the like.

Measurement is ultimately a comparison: a thing or concept measured against a point of limitation. We compare the length of an object with the scale of a ruler or a measuring tape. We "measure" an ideology against the meaning of it as articulated by its originator. For example, the essence of a philosophy arises from the writings and teachings of its founder: Platonism from Plato, Marxism from Karl Marx, and romanticism, perhaps, from Jean Jacques Rousseau. The essence of a religious belief lies in its sacred writings, in the precepts of its great teachers, and in its creed. The meaning of freedom is articulated in many political documents—for instance, in the Declaration of Independence and the Constitution of the United States. Against these original sources, it is possible to measure the thoughts and ideas of others and to approximate their similarity to or deviance from those sources.

Data examined statistically are constantly being interpreted in comparison with statistical norms: the normal curve, probability tables, and other accepted statistical standards. Data analyzed in a nonnumerical, qualitative manner are compared across data sources, across methods, and across time.

As you can see, then, our definition of measurement implies much more than an everyday understanding of measurement might suggest. Measurement provides an important tool with which data may be inspected, analyzed, and interpreted so that the researcher may probe the meaning that lies below the surface.

Measuring Insubstantial Phenomena: An Example

Measuring insubstantial phenomena—those phenomena that have no obvious, concrete basis in the physical world—can sometimes involve considerable creativity on the part of the researcher. For example, imagine that we want to examine—and hence also to *measure*—the interpersonal dynamics among a small group of people. Let's take a group of nine people, shown in Figure 4.3, who work together in the human resources department of a large corporation. They are to attend a recognition dinner at an exclusive hotel.

They arrive in four cars. They enter the hotel in the following order: Terri, Sara, Greg, Tim, Gretchen, Matt, Peter, Jeff, and Joe. They greet one another and have time for a brief conversation before dinner. Most of them position themselves in conversation groups, as shown in Figure 4.4.

To the perceptive observer, the interpersonal dynamics within the group soon become apparent. Who greets whom with enthusiasm or with indifference? Who joins in conversation with whom? Who seems to be a relative outsider? However, *to merely observe the behavior of individuals in a particular situation is not to measure it.*

One possible approach to measuring the group's interpersonal dynamics is to give each person in the group a slip of paper on which to record three choices: (a) one or more individuals in

[2] You may sometimes see the substantial–insubstantial distinction referred to as *manifest variables* (which can be directly observed and measured) versus *latent variables* (which lie below the surface and can be measured only indirectly through their effects on another, observable entity; e.g., see Bartholomew, 2004).

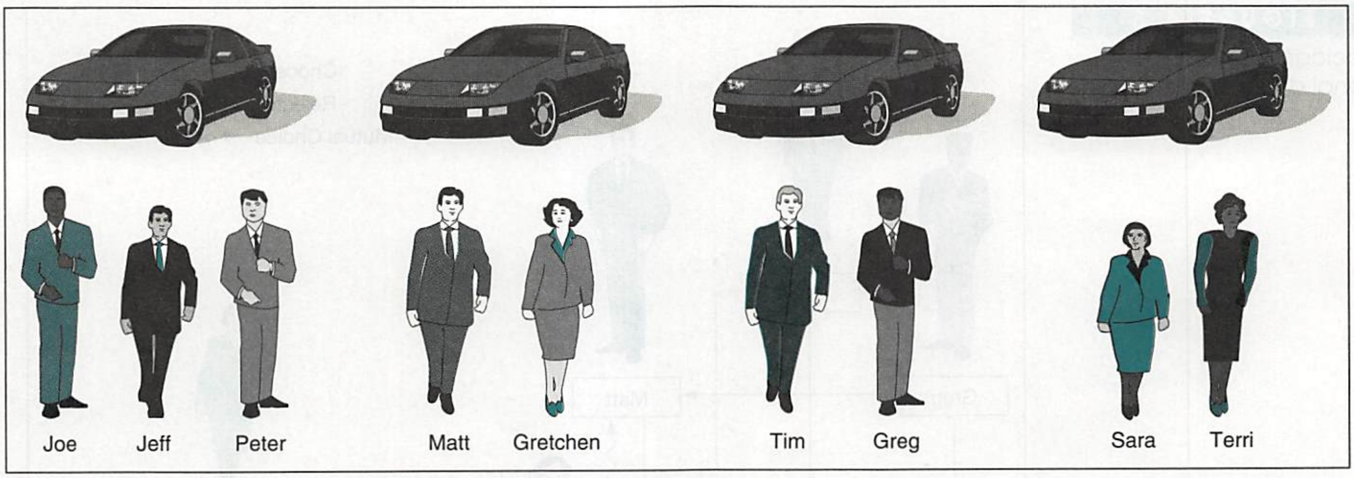


FIGURE 4.3
Recognition dinner participants

the group whom the person likes most, (b) one or more individuals whom the person likes least, and (c) one or more individuals for whom the person has no strong feeling one way or the other. When using this method, we should poll each person in the group individually and guarantee that every response will be kept confidential.

We can then draw a chart, or **sociogram**, of these interpersonal reactions, perhaps in the manner depicted in Figure 4.5. We might also assign "weights" that place the data into three numerical categories: +1 for a positive choice, 0 for indifference, and –1 for a negative reaction. Categorizing the data in this way, we can then construct a sociometric matrix. To create a matrix, we arrange the names of each person twice: vertically down the left side of a grid and horizontally across the top of the grid. The result is shown in Table 4.1. The dashes in the grid reflect the fact that the people can choose other individuals but cannot choose *themselves.*

Certain relationships begin to emerge. As we represent group dynamics in multiple forms, certain clusters of facts suggest the following conclusions:

- Jeff is the informal or popular leader (sometimes called the "star") of the group. He received five choices and only one rejection (see the "Jeff" column in Table 4.1). The sociogram confirms Jeff's popularity with his colleagues.
- Probably some factions and possible tension are present in this group. Notice that Peter, Sara, and Terri form a subclique, or "island," that is separated from the larger clique that Jeff leads. The apparent liaison between these two groups is Joe, who has mutual choices with both Jeff and Peter.
- Friendship pairs may lend cohesion to the group. Notice the mutual choices: Matt and Gretchen, Gretchen and Jeff, Jeff and Joe, Joe and Peter, Peter and Terri, Terri and Sara. The sociogram clearly reveals these alliances.

FIGURE 4.4
Conversation groups

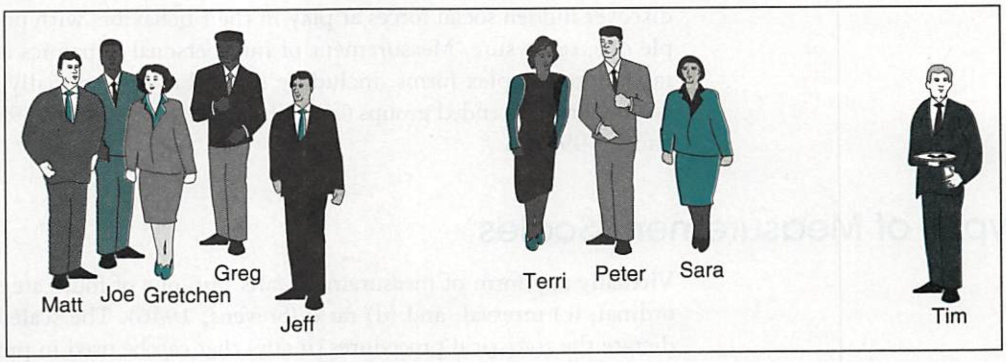
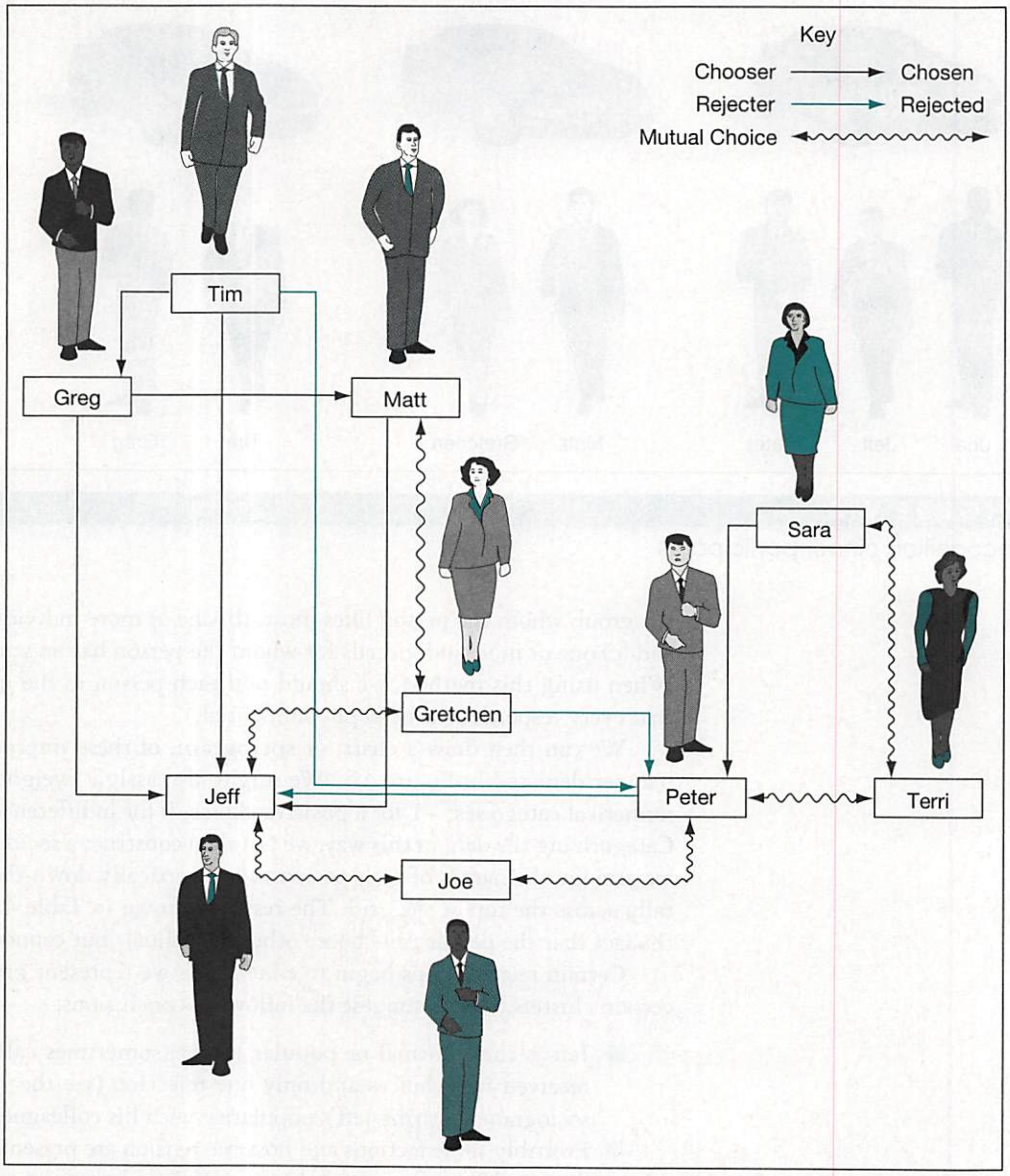


FIGURE 4.5
Sociogram of interpersonal dynamics



- Tim apparently is the isolate of the group. He received no choices; he is neither liked nor disliked. In such a position, he is probably the least influential member of the group.

With this example we have illustrated what it means to interpret data by measuring an insubstantial phenomenon and analyzing the resulting data. Notice that we did not just observe the behaviors of nine individuals at a social occasion, but we also looked below the surface to discover hidden social forces at play in their behaviors with one another. Our example is a simple one, to be sure. Measurement of interpersonal dynamics and social networks can certainly take more complex forms, including some that are especially helpful in studying social forces within large, extended groups (e.g., Chatterjee & Srivastava, 1982; Freeman, 2004; Wasserman & Faust, 1994).

Types of Measurement Scales

Virtually any form of measurement falls into one of four categories, or **scales**: (a) nominal, (b) ordinal, (c) interval, and (d) ratio (Stevens, 1946). The scale of measurement will ultimately dictate the statistical procedures (if any) that can be used in processing the data.

TABLE 4.1 Data from Figure 4.5 presented as a sociometric matrix

How Each Person Rated the Others	How Each Person Was Rated by the Others								
	Gretchen	Joe	Greg	Sara	Peter	Jeff	Tim	Matt	Terri
Gretchen	—	0	0	0	-1	+1	0	+1	0
Joe	0	—	0	0	+1	+1	0	0	0
Greg	0	0	—	0	0	+1	0	+1	0
Sara	0	0	0	—	+1	0	0	0	+1
Peter	0	+1	0	0	—	-1	0	0	+1
Jeff	+1	+1	0	0	0	—	0	0	0
Tim	0	0	+1	0	-1	+1	—	0	0
Matt	+1	0	0	0	0	+1	0	—	0
Terri	0	0	0	+1	+1	0	0	0	—
Totals	2	2	1	1	1	4	0	2	2

Nominal Scales

The word *nominal* comes from the Latin *nomen,* meaning "name." Hence we can "measure" data to some degree by assigning names to them. Recall that the definition of measurement presented earlier includes the phrase *limiting the data.* That is what a **nominal scale** does—it limits the data—and just about all that it does. Assign a specific name to anything, and you have restricted that thing to the meaning of its name. For example, we can measure a group of children by dividing it into two groups: girls and boys. Each subgroup is thereby measured—restricted—by virtue of gender as belonging to a particular category.

Things can be measured nominally in an infinite number of ways. We can further measure girls and boys according to the home site of each child. Imagine that the town in which the children live is divided into two sections by Main Street, which runs from east to west. Those children who live north of Main Street are "the Northerners"; those who live south of it are "the Southerners." In one period of U.S. history, we measured the population of the entire nation in just such a manner.

Nominal measurement is quite simplistic, but it does divide data into discrete categories that can be compared with one another. Let's take an example. Imagine that we have six children: Zahra, Paul, Kathy, Binh, Ginger, and Nicky. They can be divided into six units of one child each. They can also form two groups: Zahra, Kathy, and Ginger (the girls) in one group and Paul, Binh, and Nicky (the boys) in the other. Perhaps all six children are students in a class that meets in Room 12 at Thompson's Corner School. By assigning a room number, we have provided the class with a name, even though that "name" is a number. In this case, the number has no quantitative meaning: Room 12 is not necessarily bigger or better than Room 11, nor is it inferior to Room 13.

Only a few statistical procedures are appropriate for analyzing nominal data. We can use the *mode* as an indicator of the most frequently occurring category within our data set; for example, we might determine that there are more boys than girls in Room 12 at Thompson's Corner School. We can find the *percentage* of people in various subgroups within the total group; for example, we could calculate the percentage of boys in each classroom. We can use a *chi-square test* to compare the relative frequencies of people in various categories; for example, we might discover that more boys than girls live north of Main Street but that more girls than boys live south of Main Street. (We discuss these statistics, as well as the statistics listed in the following discussions of the other three scales, in Chapter 11.)

Ordinal Scales

With an **ordinal scale** of measurement, we can think in terms of the symbols > (greater than) and < (less than). We can compare various pieces of data in terms of one being greater or higher than another. In essence, this scale allows us to *rank-order* our data—hence its name *ordinal.*

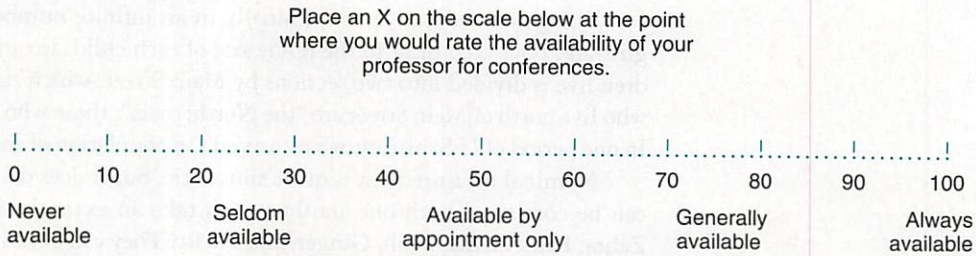
As an example, we can roughly measure level of education on an ordinal scale by classifying people as being unschooled or having completed an elementary, high school, college, or graduate education. Likewise, we can measure members of the workforce by grades of proficiency: unskilled, semiskilled, or skilled.

An ordinal scale expands the range of statistical techniques we can apply to our data. In addition to the statistics we can use with nominal data, we can also determine the *median,* or halfway point, in a set of data. We can use a *percentile rank* to identify the relative position of any item or individual in a group. We can determine the extent of the relationship between two characteristics by means of Spearman's *rank order correlation.*

Interval Scales

An **interval scale** of measurement is characterized by two features: (a) It has equal units of measurement and (b) its zero point has been established arbitrarily. The Fahrenheit (F) and Celsius (C) scales for measuring temperature are examples of interval scales: The intervals between any two successive numbers of degrees reflect equal changes in temperature, but the zero point is not equivalent to a total absence of heat. For instance, when Gabriel Fahrenheit was developing his Fahrenheit scale, he first took as his zero point the coldest temperature he observed in Iceland. Later, he made it the lowest temperature obtainable with a mixture of salt and ice. This was purely an arbitrary decision. It placed the freezing point of water at 32° and the boiling point at 212° above zero.

The rating scales employed by many survey groups, businesses, and professional organizations are often assumed to be on interval scales. For instance, many universities ask students to use rating scales to evaluate the teaching effectiveness of various professors. Following is an example of an item from one university's teaching evaluation form:



Notice that the scale has 11 equidistant points ranging from 0 to 100. The equidistance creates what is presumed to be an interval scale for the measure. At five points along the scale are descriptive labels that can help students determine how they should rate their professor's availability. We might place descriptors at more places along the scale (perhaps at 10-point distances), thus potentially making the scale more sensitive or more accurate. For indicating the availability of a professor, such fineness of discrimination may not be either possible or desirable, but one can conceive of situations in which such a degree of discrimination may be necessary and appropriate.

Interval scales of measurement allow statistical analyses that are not possible with nominal or ordinal data. Because an interval scale reflects equal distances among adjacent points, any statistics that are calculated using addition or subtraction—for instance, *means, standard deviations,* and *Pearson product moment correlations*—can now be used.

Ratio Scales

Two commonly used measurement instruments—a thermometer and a yardstick—might help you understand the difference between the interval and ratio scales. If we have a thermometer that measures temperature on the Fahrenheit scale, we *cannot* say that 80°F is twice as warm as

40°F. Why? Because this scale does not originate from a point of absolute zero; a substance may have some degree of heat even though its measured temperature falls *below* zero. With a yardstick, however, the beginning of linear measurement is absolutely the beginning. If we measure a desk from the left edge to the right edge, that's it. There is no more desk in either direction beyond those limits. A measurement of "zero" means there is no desk at all, and a "minus" desk width isn't even possible.

More generally, a **ratio scale** has two characteristics: (a) equal measurement units (similar to an interval scale) and (b) *an absolute zero point,* such that 0 on the scale reflects a total absence of the entity being measured.

Let's consider once again the "availability" scale presented earlier for measuring professor effectiveness. This scale could never be considered a ratio scale. Why? Because there is only one condition in which the professor would be absolutely unavailable—if the professor were dead!—in which case we would not be asking students to evaluate this individual.

What distinguishes the ratio scale from the other three scales is that *the ratio scale can express values in terms of multiples and fractional parts,* and the ratios are *true* ratios. A yardstick can do that: A yard is a *multiple* (by 36) of a 1-inch distance; an inch is one-twelfth (a *fractional part*) of a foot. The ratios are 36:1 and 1:12, respectively.

Ratio scales outside the physical sciences are relatively rare. And whenever we cannot measure a phenomenon in terms of a ratio scale, we must refrain from making comparisons such as "this thing is three times as great as that" or "we have only half as much of one thing as another." Only ratio scales allow us to make comparisons that involve multiplication or division.

We can summarize our description of the four scales this way:

If you can say that
- One object is different from another, you have a *nominal scale;*
- One object is bigger or better or more of anything than another, you have an *ordinal scale;*
- One object is so many units (degrees, inches) more than another, you have an *interval scale;*
- One object is so many times as big or bright or tall or heavy as another, you have a *ratio scale.*

(Senders, 1958, p. 51)

Table 4.2 provides a quick reference for the various types of scales, their distinguishing characteristics, and the statistical analysis possibilities for each scale. When we consider the statistical interpretation of data in later chapters (especially in Chapter 11), you may want to refer to this table to determine whether the type of measurement instrument you have used will support the statistical operation you are contemplating.

TABLE 4.2 A summary of measurement scales, their characteristics, and their statistical implications

	Measurement Scale	*Characteristics of the Scale*	*Statistical Possibilities of the Scale*
Non-Interval Scales	Nominal scale	A scale that "measures" only in terms of names or designations of discrete units or categories	Enables one to determine the mode, percentage values, or chi-square
	Ordinal scale	A scale that measures in terms of such values as "more" or "less," "larger" or "smaller," but without specifying the size of the intervals	Enables one also to determine the median, percentile rank, and rank correlation
Interval Scales	Interval scale	A scale that measures in terms of equal intervals or degrees of difference, but with an arbitrarily established zero point that does not represent "nothing" of something	Enables one also to determine the mean, standard deviation, and product moment correlation; allows one to conduct most inferential statistical analyses
	Ratio scale	A scale that measures in terms of equal intervals and an absolute zero point	Enables one also to determine the geometric mean and the percentage variation; allows one to conduct virtually any inferential statistical analysis

CONCEPTUAL ANALYSIS EXERCISE Identifying Scales of Measurement

Each of the following scenarios involves measuring one or more variables. Decide whether the various measurements reflect nominal, ordinal, interval, or ratio scales, and justify your choices. Be careful, as the answers are not always as obvious as they might initially appear. The answers are provided after the "For Further Reading" list at the end of the chapter.

1. An environmental scientist collects water samples from streams and rivers near large industrial plants and saves exactly one liter of water from each sample. Then, back at the lab, the researcher determines the amounts of certain health-jeopardizing bacteria in each sample. What measurement scale does the measurement of bacteria content reflect?
2. A tourism researcher is studying the relationship between (a) a country's average annual temperature and (b) the amount of tourist dollars that the country brings in every year. What scales underlie the two variables in this study?
3. A political science researcher wants to determine whether people's political party membership is correlated with the frequency with which they have voted in local elections in the past five years. The researcher can easily obtain information about people's party membership and voting records from the town clerks in several communities. To simplify data collection, the researcher uses the following coding scheme for party membership: 1 = Registered as Democrat, 2 = Registered as Republican, 3 = Registered as member of another party, 0 = No declared party affiliation. What measurement scale(s) underlie (a) political party membership and (b) voting frequency?
4. A marketing researcher in the United States wants to determine whether a certain product is more widely used in some parts of the country than others. The researcher separates the country into ten regions based on zip code; zip codes below 10000 are northeastern states, zip codes of 90000 and above are western states, and so on. What measurement scale does the researcher's coding scheme for the regions represent?
5. An economist is studying the home-buying behaviors of people of different income levels. The researcher puts people into four categories: Group A includes those earning up to $20,000 per year, Group B includes those earning between $20,001 and $50,000 per year, Group C includes those earning between $50,001 and $100,000 per year, and Group D includes those earning more than $100,000 per year. In this study, what kind of scale is income level?
6. A geographer is studying traffic patterns on four different types of roads that vary in quality: superhighways (i.e., roads accessible only by relatively infrequent on–off ramps), highways (i.e., roads that allow relatively high speeds for long distances but may have an occasional traffic light), secondary roads (i.e., well-paved two-lane roads), and tertiary roads (narrow, infrequently traveled roads; some may consist only of gravel). The type of road in this study reflects which type of measurement scale?
7. A psychologist is developing an instrument designed to measure college students' test anxiety. The instrument includes 25 statements—for example, "My heart starts to pound when I see the word *test* on a course syllabus" and "My palms get sweaty while I'm taking a multiple-choice test." Students must rate each of these statements on a 5-point scale, as follows:

   0 This is never true for me.
   1 This is rarely true for me.
   2 This is sometimes true for me.
   3 This is often true for me.
   4 This is always true for me.

   Students who answer "never" to each of the 25 questions get the lowest possible score of 0 on the instrument. Students who answer "always" to each of the 25 questions get the highest possible score of 100 on the instrument. Thus, scores on the instrument range from 0 to 100. What kind of scale do the scores represent?

Validity and Reliability in Measurement

Regardless of the type of scale a measurement instrument involves, the instrument must have both validity and reliability for its purpose. The validity and reliability of measurement instruments influence the extent to which a researcher can learn something about the phenomenon under investigation, the probability that the researcher will obtain statistical significance in any data analysis, and the extent to which the researcher can draw meaningful conclusions from the data.

In your research proposal and final research report, you should provide evidence that the instruments you use have a reasonable degree of validity and reliability for your purposes. But validity and reliability take different forms, depending on the nature of the research problem, the methodology being used to address the problem, and the nature of the data that are collected. In the following sections, we briefly discuss the kinds of validity and reliability you may need to consider and how you might determine and enhance them.

Validity of Measurement Instruments

The validity of a measurement instrument is the extent to which the instrument measures what it is intended to measure. Certainly no one would question the premise that a yardstick is a valid means of measuring length. Nor would most people doubt that a thermometer measures temperature; for instance, in a mercury thermometer, the level to which the mercury rises is a function of how much it expands, which is a function of the degree to which it is hot or cold.

But to what extent does an intelligence test actually measure a person's intelligence? How accurately do people's annual incomes reflect their social class? And how well does a sociogram capture the interpersonal dynamics in a group of nine people? Especially when we are measuring *insubstantial* phenomena—phenomena without a direct basis in the physical world—our measurement instruments may be somewhat suspect in terms of validity.

Let's return to the rating-scale item presented earlier to assess a professor's availability to students (see p. 86) and consider its validity as such a measure. Notice how fuzzy some of the labels are. The professor is "always available." What does *always* mean? Twenty-four hours a day? Could you call the professor at 3:00 a.m. any day of the week, or only whenever the professor is on campus? If the latter is the case, could you call your professor out of a faculty meeting or out of a conference with the college president? We might have similar problems in interpreting "generally available," "seldom available," and "never available." On careful inspection, what seems at first glance to be a scale that anyone could understand has limitations *as a measurement instrument* for research purposes.

A paper-and-pencil test may be intended to measure a certain characteristic, and it may be *called* a measure of that characteristic, but these facts don't necessarily mean that the test actually measures what its creator says it does. For example, consider a paper-and-pencil test of personality traits in which, with a series of check marks, a person indicates his or her most representative characteristics or behaviors in given situations. The person's responses on the test are presumed to reveal relatively stable personality traits. The question that validity asks is: Does such a test, in fact, measure the person's personality traits, or does it measure something else altogether? The answer depends, at least in part, on the extent to which the person is or *can* be truthful in responding. If the person responds in terms of characteristics and behaviors that he or she believes to be socially desirable, the test results may reveal not the person's actual personality, but rather an idealized portrait of how he or she would like to be perceived by others.

The validity of a measurement instrument can take several different forms, each of which is important in different situations:

- **Face validity** is the extent to which, on the surface, an instrument *looks like* it is measuring a particular characteristic. Face validity is often useful for ensuring the cooperation of people who are participating in a research study. But because it relies entirely on subjective judgment, it is not, in and of itself, a terribly dependable indicator that an instrument is truly measuring what the researcher wants to measure.
- **Content validity** is the extent to which a measurement instrument is a representative sample of the content area (domain) being measured. Content validity is often a consideration when

a researcher wants to assess people's *achievement* in some area—for instance, the knowledge students have learned during classroom instruction or the job skills workers have acquired in a training program. A measurement instrument has high content validity if its items or questions reflect the various parts of the content domain in appropriate proportions and if it requires the particular behaviors and skills that are central to that domain.

- **Criterion validity** is the extent to which the results of an assessment instrument correlate with another, presumably related measure (the latter measure is, in this case, the *criterion*). For example, a personality test designed to assess a person's shyness or outgoingness has criterion validity if its scores correlate with other measures of a person's general sociability. An instrument designed to measure a salesperson's effectiveness on the job should correlate with the number of sales the individual actually makes during the course of a business week.
- **Construct validity** is the extent to which an instrument measures a characteristic that cannot be directly observed but is assumed to exist based on patterns in people's behavior (such a characteristic is a *construct*). Motivation, creativity, racial prejudice, love—all of these are constructs, in that none of them can be *directly* observed and measured. When researchers ask questions, present tasks, or observe behaviors as a way of assessing an underlying construct, they should obtain some kind of evidence that their approach does, in fact, measure the construct in question.

Sometimes there is universal agreement that a particular instrument provides a valid instrument for measuring a particular characteristic; such is the case for yardsticks, thermometers, barometers, oscilloscopes, and so on. But whenever we do *not* have such widespread agreement, we must provide evidence that an instrument we are using has validity for our purpose.

It is critical to note that the validity of any measurement instrument can vary considerably depending on the purpose for which it is being used. In other words, *the validity of an instrument is specific to the situation.* For example, a tape measure wrapped horizontally around a person's head is a valid measure of the person's head circumference but *not* a valid measure of the person's intelligence. Likewise, a widely used intelligence test might provide a reasonable estimation of children's general cognitive development but is *not* suitable for determining how well the children can perform in, say, a geometry class or interpersonal conflict.

Determining the validity of a measurement instrument. An in-depth discussion of how to determine validity is beyond the scope of this book; measurement textbooks, such as those listed in the "For Further Reading" section at the end of the chapter, provide more detailed information. But here we offer three examples of what researchers sometimes do to demonstrate that their measurement instruments have validity for their purposes:

- **Table of specifications.** To construct a measurement instrument that provides a representative sample of a particular content domain—in other words, to establish content validity—the researcher often constructs a two-dimensional grid (*table of specifications*) listing the specific topics and behaviors that reflect achievement in the domain. In each cell of the grid, the researcher indicates the relative importance of each topic–behavior combination. He or she then develops a series of tasks or test items that reflects the various topics and behaviors in appropriate proportions.
- **Multitrait–multimethod approach.** Two or more different characteristics are each measured using two or more different approaches (Campbell & Fiske, 1959). The different measures of the same characteristic should be highly correlated. The same ways of measuring different characteristics should *not* be highly correlated. For example, in a classroom situation, the constructs *achievement motivation* and *social motivation* might each be measured by both self-report questionnaires and teacher observation checklists. Statistical analyses should reveal that the two measures of achievement motivation are highly correlated and that the two measures of social motivation are also highly correlated. Results from the two self-report questionnaires—because they are intended to assess different and presumably unrelated characteristics—should *not* be highly correlated, nor should results from the two teacher checklists.

- **Judgment by a panel of experts.** Several experts in a particular area are asked to scrutinize an instrument and give an informed opinion about its validity for measuring the characteristic in question.

Although none of the approaches just described guarantees the validity of a measurement instrument, each one increases the likelihood of such validity.

Reliability of Measurement Instruments

Imagine that you are concerned about your growing waistline and decide to go on a diet. Every day you put a tape measure around your waist and pull the two ends together snugly to get a measurement. But just how tight is "snug"? Quite possibly, the level of snugness might be different from one day to the next. In fact, you might even measure your waist with different degrees of snugness from one *minute* to the next. To the extent that you are not measuring your waist in a consistent fashion—even though you always use the same tape measure—you have a problem with reliability.

More generally, **reliability** is the consistency with which a measuring instrument yields a certain, consistent result when the entity being measured hasn't changed. As we have just seen in our waist-measuring situation, instruments that measure physical phenomena aren't necessarily completely reliable. As another example, think of a measuring cup that a baker might use while making a cake. When measuring a half-cup of flour, the baker won't always measure *exactly* the same amount of flour each time.

Instruments designed to measure social and psychological characteristics (insubstantial phenomena) tend to be even less reliable than those designed to measure physical (substantial) phenomena. For example, a student using the rating-scale item presented earlier for measuring professor availability might easily rate the professor as "70" one day and "90" the next, not because the professor's availability has changed overnight but because the student's interpretations of the phrases "generally available" and "always available" *have* changed. Similarly, if we asked the nine people portrayed in Figure 4.3 (Gretchen, Joe, Greg, etc.) to indicate the people they liked best and least among their colleagues, they wouldn't necessarily always give us the same answers they gave us previously, even if the interpersonal dynamics within the group have remained constant.

Determining the reliability of a measurement instrument. Like validity, reliability takes different forms in different situations. But in the case of reliability, its particular form is essentially equivalent to the procedure used to determine it. Following are four forms of reliability that are frequently of interest in research studies:

- **Interrater reliability** is the extent to which two or more individuals evaluating the same product or performance give identical judgments.
- **Test-retest reliability** is the extent to which a single instrument yields the same results for the same people on two different occasions.
- **Equivalent forms reliability** is the extent to which two different versions of the same instrument (e.g., "Form A" and "Form B" of a scholastic aptitude test) yield similar results.
- **Internal consistency reliability** is the extent to which all of the items within a single instrument yield similar results.

For each of these forms, determining reliability involves two steps:

1. Getting two measures for each individual in a reasonably large group of individuals—in particular by doing one of the following:
   a. Having two different raters evaluate the same performance for each individual (interrater reliability)
   b. Administering the same instrument to the individuals at two different points in time—perhaps a day, a week, or a month apart (test-retest reliability)

c. Giving each individual two parallel versions of the same instrument (equivalent forms reliability)

d. Administering only one instrument but calculating two subscores for the instrument—for instance, calculating one score for odd-numbered items and another score for even-numbered items (internal consistency reliability)

2. Calculating a correlation coefficient that expresses the degree to which the two measures are similar (see Chapter 11 for a discussion of correlation coefficients)

You can find more in-depth discussions about determining reliability in almost any general measurement textbook.

Enhancing the Reliability and Validity of a Measurement Instrument

Both validity and reliability reflect the degree to which we may have *error* in our measurements. In many instances—and especially when we are measuring insubstantial phenomena—a measurement instrument may allow us to measure a characteristic only indirectly and so may be subject to a variety of biasing factors (e.g., people's responses on a rating scale are apt to be influenced by their interpretations, prejudices, memory lapses, etc.). In such cases, we have error due to the imperfect *validity* of the measurement instrument. Yet typically—even when we are measuring substantial phenomena—we may get slightly different measures from one time to the next simply because our measurement tool is imprecise (e.g., the waist or head size we measure may depend on how snugly we pull the tape measure). In such cases, we have error due to the imperfect *reliability* of the measure. Generally speaking, validity errors reflect biases in the instrument itself and are relatively constant sources of error. In contrast, reliability errors reflect *use* of the instrument and are apt to vary unpredictably from one occasion to the next.

We can measure something accurately only when we can also measure it consistently. Hence, by increasing the reliability of a measurement instrument, we might also increase its validity. A researcher can enhance the reliability of a measurement instrument in several ways. First, the instrument should always be administered in a consistent fashion. In other words, there should be **standardization** in use of the instrument from one situation or individual to the next. Second, to the extent that subjective judgments are required, specific *criteria* should be established that dictate the kinds of judgments the researcher makes. And third, any research assistants who are using the instrument should be *well trained* so that they obtain similar results for any single individual or situation being measured.

Yet even if we enhance the reliability of our measurements, we don't necessarily increase their accuracy. In other words, *reliability is a necessary but insufficient condition for validity.* For example, we could use a tape measure to measure a person's head circumference and claim that the result is a good reflection of intelligence. In this situation, we might have reasonable reliability—we are apt to get similar measures of an individual's head circumference on different occasions—but absolutely no validity. Head size is *not* a good indication of intelligence level.

Creative researchers use a wide variety of strategies to enhance the validity of their measurement instruments. One important strategy is to consult the literature in search of measurement techniques that other researchers have effectively used. Another is to show a first draft of an instrument to experienced colleagues and ask for their feedback and suggestions. Still another strategy is to conduct one or more *pilot studies* specifically to try out a particular instrument, carefully scrutinizing it for obvious or possible weaknesses, and then modify it in minor or major ways.

We cannot overemphasize the importance of determining and maximizing the validity and reliability of your measurement instruments. Without reasonably valid and reliable measures of the characteristics and phenomena under investigation, you cannot possibly obtain informative and useful data for addressing and solving your research problem.

As you plan your research project, you should clearly identify the nature of the measurement instruments you will use and carefully scrutinize them with respect to their potential validity and reliability. Furthermore, in your research proposal and final research report, you should describe any instrument in explicit, concrete terms. For example, if you are using a particular piece of equipment to measure a certain physical characteristic or phenomenon, you should describe the equipment's specific nature (e.g., its manufacturer, model number, and level of precision). And if you are assessing some aspect of human thought or behavior, you should describe the questions asked or tasks administered, the overall length of the instrument (e.g., number of items, time required for administration), and the method of scoring responses.

CONCEPTUAL ANALYSIS EXERCISE Identifying Problems with Validity and Reliability in Measurement

In each of the scenarios in this exercise, a researcher encounters a measurement problem. Some of the scenarios reflect a problem with the validity of a measure. Others reflect a problem with a measure's reliability—a problem that indirectly also affects the measure's validity. For each scenario, choose the most obvious problem from among the following alternatives:

- Face validity
- Content validity
- Criterion validity
- Construct validity
- Interrater reliability
- Test-retest reliability
- Equivalent forms reliability
- Internal consistency reliability

The answers appear after the "For Further Reading" list at the end of this chapter.

1. After using two different methods for teaching basic tennis skills to non-tennis-playing adults, a researcher assesses the effectiveness of the two methods by administering a true–false test regarding the rules of the game (e.g., faults and double-faults, scoring procedures).
2. A researcher writes 120 multiple-choice questions to assess middle school students' general knowledge of basic world geography (e.g., what the equator is, where Africa is located). To minimize the likelihood that students will cheat on the test by copying one another's answers, the researcher divides the questions into three different sets to create three 40-item tests. In collecting data, the researcher distributes the three tests randomly to students in any single classroom. After administering the tests to students at many different middle schools, the researcher computes the students' test scores and discovers that students who answered one particular set of 40 questions scored an average of 3 points higher than students who answered either of the other two 40-question sets.
3. In order to determine what kinds of situations provoke aggression in gorillas, two researchers observe mountain gorillas in the Virunga Mountains of northwestern Rwanda. As they watch a particular gorilla family and take notes about family members' behaviors, the researchers often disagree about whether certain behaviors constitute "aggression" or, instead, reflect more peaceful "assertiveness."
4. A researcher uses a blood test to determine people's overall energy level after drinking or not drinking a can of a high-caffeine cola drink. Unfortunately, when two research assistants independently rate people's behaviors for energy level for a 4-hour period after drinking the cola, their results don't seem to have any correlation with the blood-test results.
5. In a two-week period during the semester, a researcher gains entry into several college classrooms in order to administer a short survey regarding college students' beliefs about climate change. The survey consists of 20 statements about climate change (e.g., "Devastating floods in recent years are partly the result of the earth's gradually rising overall temperature"), to which students must respond "Strongly disagree," "Disagree," "Agree," or "Strongly agree." Many of the students voluntarily put their names on their surveys. Thanks to the names on many survey forms, the researcher discovers that a few

students were in two of the classes surveyed and thus completed the survey twice. Curiously, however, these students sometimes gave different responses to particular statements on the two different occasions, and hence their overall scores were also different.

6. In order to get a sense of how harmonious most long-term marriages are, a researcher administers a questionnaire to married couples who have been married for at least 20 years. The questionnaire consists of 60 statements to which both spouses must individually respond either "This describes my marriage" or "This doesn't describe my marriage." All 60 statements describe a possible characteristic of a nonharmonious marriage (e.g., "We fight all the time," "We rarely agree about how to spend our money"), and the researcher has sequenced them in a random order on the questionnaire. Even so, the researcher discovers that respondents more frequently agree with the first 30 items than with the last 30 items. If one were to look only at responses to the first 30 items, then, one would think that married couples fight a lot. But if one were to look only at responses to the last 30 items, one would conclude that most long-term couples live in relative peace and harmony. (*Note:* We recommend that questionnaires *not* be slanted in a one-way direction, as this one is; see the "Constructing a Questionnaire" guidelines in Chapter 8).
7. A researcher develops and uses a questionnaire intended to measure the extent to which college students display tolerance toward a particular religious group. However, several experts in the researcher's field of study suggest that the questionnaire measures not how tolerant students actually *are,* but what students would like to *believe* about their tolerance for people of a particular religion.
8. Students in an introductory college psychology course must satisfy their "research methods" requirement in one of several ways; one option is to participate in a research study called "Intelligence and Motor Skill Learning." When students choosing this option report to the laboratory, one of their tasks is to respond as quickly as possible to a series of simple computer-generated questions. Afterward, the researcher debriefs the students about the nature of the study and tells them that the reaction-time measure was designed to be a simple measure of intelligence. Some of the students object, saying, "That's not a measure of intelligence! Intelligence isn't how quickly you can do something, it's how *well* you can do it."

Linking Data and Research Methodology

Earlier in the chapter we described data as a link between Absolute Truth and the researcher's inquiring mind. Data are like ore: They contain pieces of the truth but are in a rather unrefined state. To extract meaning from the data, we employ what is commonly called *research methodology.*

Data and methodology are inextricably intertwined. For this reason, the methodology to be used for a particular research problem must always take into account the nature of the data that will be collected in the resolution of the problem.

An example may help clarify this point. Imagine that a man from a remote village decides to travel to the big city. While he is there, he takes his first ride on a commercial jetliner. No one in his village has ever flown before, so naturally when he returns home, they are curious about his trip. One afternoon two friends ask him about his experience, yet each one asks very different questions. The first friend asks, "How fast did you move?" "How far did you go?" and "How high did you fly?" The second friend asks, "How did you feel when you were moving so fast?" "What was it like being above the clouds?" and "What did the city look like from so high?" Both friends are asking questions that can help them learn more about the experience of flying in an airplane. Yet they differ considerably in what they want to know. Because they ask different kinds of questions, they obtain different kinds of information. If they were to recount what they had learned about "flying in a jetliner," they would most likely describe very different things. Although neither of them has the "wrong" story, neither does each one have the whole story.

In research, too, different questions yield different types of information. Different research problems lead to different research designs and methods, which in turn result in the collection of different types of data and different interpretations of those data.

Furthermore, many kinds of data may be suitable only for a particular methodology. To some extent, *the data dictate the research method.* As an example, consider historical data, those pieces of information gleaned from written records of past events. You cannot extract much meaning from historical documents by using a laboratory experiment. An experiment is simply not suited to the nature of the data.

Over the years, numerous research methodologies have emerged to accommodate the many different forms that data are likely to take. Accordingly, we must take a broad view of the approaches the term *research methodology* encompasses. Above all, we must not limit ourselves to the belief that only a true experiment constitutes "research." Such an attitude prohibits us from agreeing that we can better understand Coleridge's poetry by reading the scholarly research of John Livingston Lowes (1927, 1955) or from appreciating Western civilization more because of the historiography of Arnold Toynbee (1939–1961).

No single highway leads us exclusively toward a better understanding of the unknown. Many highways can take us in that direction. They may traverse different terrain, but they all converge on the same destination: the enhancement of human knowledge.

In Chapters 6 through 10 of this book we zero in on various research methodologies. But many researchers tend to categorize research studies into two broad categories: quantitative research and qualitative research. We now look at this distinction.

Comparing Quantitative and Qualitative Approaches

As you might guess, **quantitative research** involves looking at amounts, or *quantities,* of one or more variables of interest. A quantitative researcher typically tries to measure variables in some numerical way, perhaps by using commonly accepted measures of the physical world (e.g., rulers, thermometers, oscilloscopes) or carefully designed measures of psychological characteristics or behaviors (e.g., tests, questionnaires, rating scales).

In contrast, **qualitative research** involves looking at characteristics, or *qualities,* that cannot be entirely reduced to numerical values. A qualitative researcher typically aims to examine the many nuances and complexities of a particular phenomenon. You are most likely to see qualitative research in studies of complex human situations (e.g., people's in-depth perspectives about a particular issue, the behaviors and values of a particular cultural group) or complex human creations (e.g., television commercials, works of art). Qualitative research is not limited to research problems involving human beings, however. For instance, some biologists study, in a distinctly qualitative manner, the complex social behaviors of other animal species; Dian Fossey's work with gorillas and Jane Goodall's studies of chimpanzees are two well-known examples (e.g., see Fossey, 1983; Goodall, 1986).

Quantitative and qualitative approaches involve similar processes—for instance, they both entail identifying a research problem, reviewing related literature, and collecting and analyzing data. Yet these processes are often combined and carried out in different ways, leading to distinctly different research methods. For instance, quantitative researchers often start with one or more specific hypotheses to be tested. They isolate the variables they want to study, use a standardized procedure to collect some form of numerical data, and use statistical procedures to analyze and draw conclusions from the data. In contrast, qualitative researchers often start with general research questions rather than specific hypotheses, collect an extensive amount of verbal data and/or nonverbal artifacts, organize those data and artifacts into some form that gives them coherence, and use verbal descriptions to portray the situation they have studied.

To some extent, quantitative and qualitative research designs are appropriate for answering different kinds of questions. As a result, we learn more about the world when we have both