over the years and will continue to do so. A constant thread through the history of the field is the dynamic interaction between science and practice—in most cases for the betterment of organizations and their employees.

SUGGESTED ADDITIONAL READINGS

Dunnett, M. D. (1998). Emerging trends and vexing issues in industrial and organizational psychology. *Applied Psychology: An International Review, 47,* 129–153.

Koppes, L. L. (1997). American female pioneers of industrial and organizational psychology during the early years. *Journal of Applied Psychology, 82,* 500–515.

Landy, F. J. (1997). Early influences on the field of industrial and organizational psychology. *Journal of Applied Psychology, 82,* 467–477.

Hyatt, D., Cropanzano, R., Finfer, L. A., Levy, P., Ruddy, T.M., Vandaveer, V., & Walker, S. (1997). Bridging the gap between academics and practice: Suggestions from the field. *Industrial-Organizational Psychologist, 35,* 29–32.

Kanigel, R. (1997). *The one best way: Frederick Winslow Taylor and the enigma of efficiency.* New York: Viking.

Zickar, M. J. (2003). Remembering Arthur Kornhauser: Industrial psychology's advocate for worker well-being. *Journal of Applied Psychology, 88,* 363–369.

Chapter Two

Research Methods and Statistics

Organizational psychologists often design scientific investigations to answer a variety of research questions about behavior in organizational settings; in some cases research is designed to test theories. In order to conduct research, one must make use of research designs, as well as a variety of statistical analyses. As will be shown in this chapter, research methods may range from simple observation of behavior to more elaborate designs. Likewise, statistical methods may range from very simple descriptive measures, to very elaborate model testing.

Research methodology and statistical analysis are also crucial to the practice of organizational psychology. For example, organizational psychologists often use systematic research methods to provide organizational decision makers with information regarding employees' attitudes. In other cases, research methodology and statistical analysis may be used to evaluate some intervention designed to enhance organizational effectiveness. An organization may want to know, for example, whether a team development intervention will enhance the functioning of work groups. This question, and others like it, can also be answered with the aid of typical research methods and statistical analyses used in organizational psychology.

In addition to facilitating the science and practice of organizational psychology, research methodology and statistical analysis have both emerged as legitimate fields of study within organizational psychology. Some organizational psychologists study topics such as job satisfaction, motivation, and organizational change; others have devoted their attention to methodological and statistical issues. For example, there are organizational psychologists who investigate the validity of self-report measures (e.g., Spector, 1994), as well as the analysis of data from multiple organizational levels (Bliese & Jex, 2002). Both topics will be discussed later in the chapter.

This chapter is designed to provide an introduction to the methods organizational psychologists use to collect data, as well as the statistical techniques used to analyze that data. From the student's perspective, research methodology and statistics are often viewed with some degree of apprehension. Even at the graduate level, courses in research methodology and statistics are often the most feared. Despite these negative perceptions, research methodology and statistics courses are probably the most valuable part of graduate training. Students who are well grounded in research methodology and statistics are in the best position to read and critically evaluate the research literature. They also possess a set of skills that are quite valuable, regardless of the setting in which they choose to work.

METHODS OF DATA COLLECTION

There are literally thousands of research questions that have been, and continue to be, explored by organizational psychologists. Are employees who perceive a high level of autonomy in their work likely to be highly satisfied with their jobs? Does a high level of conflict between work and family responsibilities lead to poor health? Does job performance remain consistent over time? Regardless of the research question being asked, there is a need for relevant data to be collected if the question is ever to be answered. In this section, four data-collection methods will be discussed. These include observational methods, survey research, experimentation, and quasi-experimentation.

Observational Methods

Observational methods actually encompass a variety of strategies that may be used to study behavior in organizations (Bouchard, 1976). *Simple observation*, the most basic of these strategies, involves observing and systematically recording behavior. If one wishes, for example, to investigate decision-making processes used by corporate boards of directors, one might observe these individuals during quarterly meetings and record relevant observations. These observations may reveal that the chairperson has more input into decisions than other board members, or perhaps that younger board members have less input into decisions than their more experienced counterparts.

The primary advantage of simple observation is that it allows behavior to be captured in its natural context. This allows the researcher to avoid the problem of *reactivity*, or changing the phenomenon of interest in the process of measuring it. This is only a potential advantage, however, because the presence of an observer could cause research participants to act differently than they normally would. One way to address this issue is to establish rapport with research participants to the point where they are comfortable enough with the researcher to act naturally. Another option would be to observe the behavior of interest without being detected. For example, if one were interested in the emotions displayed by service employees toward customers, one might sit in a coffee shop and observe as customers' orders are taken. This technique is also used by many retail stores; they send *mystery shoppers* to stores in order to measure the quality of customer service. Observing behavior in this way raises ethical concerns, however, because when it is used, research participants typically are not able to make an informed choice as to whether they wish to participate in the research.

Despite potential advantages, a primary disadvantage of simple observation is that it is a very labor-intensive activity. Observing and recording behavior takes a great deal of time and effort. Also, once observations are recorded, making sense out this information can be very time consuming as well. Another potential disadvantage is that observations are often subjective and may be impacted by the observer's biases. Nevertheless, simple observation can often be quite useful, particularly in the very early stages of a research program. Also, from a practical perspective, managers may find the information generated from observational studies easier to understand, and therefore more useful, than numerical data.

A form of simple observation that may be useful in some cases is *participant observation*. Participant observation is essentially the same as simple observation, except that the observer is also a participant in the event he or she is studying. In the previous example of a researcher studying corporate boards of directors, this could be participant observation if the researcher were also a member of the board. Participant observation can be highly useful, particularly when being a participant in an event provides the researcher with information that may not be obtained otherwise. This point is illustrated well by Van Maanen's (1975) investigation of police recruits as they made the transition from the training academy to regular police work. In conducting this study, Van Maanen participated in the police academy training as a recruit, and thus became a participant in the event he was studying. By doing this, he undoubtedly was able to gather information that would have been unavailable through the use of other methods (see Comment 2.1).

Despite the potential advantages of participant observation, this method also carries some risks. The biggest of these is that by taking on the role of participant, a researcher may change the phenomenon under investigation. This is somewhat ironic, considering that the general advantage of observational methods is that they reduce the risk of reactivity. Being a participant may also lead the researcher to lose his or her objectivity. As previously stated, all observations are subject to distortion, but assuming the role of a participant may compound this problem. In Van Maanen's (1975) study, this problem was dealt with by supplementing his observations with survey data from other police recruits.

Archival Data

A second method for studying behavior in organizations is through the use of *archival data* sources. Archival data represent any form of data or other records that are compiled for purposes that are independent of the research being conducted (Webb, Campbell, Schwartz, Sechrest, & Grove, 1981). Compared to other observational methods, the use of archival data is more prevalent in organizational psychology, largely because of the sheer abundance of archival data sources. Within organizations, records are typically kept on many employee behaviors such as job performance, absenteeism, turnover, and safety, to name a few. In addition, the governments of many countries maintain databases that may be relevant to the study of behavior in organizations. In the United States, for example, the Department of Labor produces the Dictionary of Occupational Titles (DOT), which contains information on the working conditions of a vast number of occupations. This database has been used in several investigations of behavior in organizations (e.g., Schaubroeck, Ganster, & Kemmerer, 1994; Spector & Jex, 1991). Recently, the DOT has been supplemented by a more extensive database in the form of the Occupational Information Network (O*NET). This represents an improvement over the DOT because the occupations that comprise the O*NET are more up to date, and the dimensions on which these occupations are described are more extensive. To date, only a few studies have used O*NET as an archival data source in the same manner

COMMENT 2.1

THE PROS AND CONS OF QUALITATIVE RESEARCH METHODS

Within the general field of psychology, and organizational psychology in particular, qualitative data collection methods such as observation are not widely used. In other fields such as sociology and anthropology, qualitative methods are used quite frequently. In psychology, we make much greater use of surveys and, to a lesser extent, experimentation and quasi-experimentation (Sackett & Larsen, 1990). In talks with colleagues over the years, the typical disadvantages associated with qualitative methods have been that they are too labor-intensive and too many biases are associated with the observational process.

Unfortunately, because of these disadvantages, many in the field psychology fail to see many of the positive features of qualitative data-collection methods. Chief among these is that observation typically provides a much richer description of whatever one is trying to study than questionnaire data do. For example, observing a group working together for a week is probably more meaningful than knowing that group members rate the group's cohesiveness as 4.3 on a scale of 1 to 6. Another advantage of most qualitative data-collection methods is that they do not require research participants to provide assessments of either themselves or the work environment. For example, we may be able to determine, through observations, that an employee has a great deal of autonomy built into his or her job. If we were to ask the employee several questions about job autonomy via a questionnaire, the employee's responses might be biased because of a temporary mood state or overall job satisfaction.

In reality, researchers do not have to make either/or decisions in choosing between qualitative and quantitative research methods. For example, in conducting employee opinion surveys, I typically use closed-ended questionnaire items, but I also include space at the end of the survey for employees to write comments that are then analyzed for content. This allows for quantitative analysis of the closed-ended survey items, but employees can express their opinions in their own words. Written comments may also reveal very useful suggestions to organizational decision makers.

Source: P. R. Sackett and J. R. Larsen, Jr. (1990). Research strategies and tactics in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1, pp. 419–490). Palo Alto, CA: Consulting Psychologists Press.

as the DOT (e.g., Liu, Spector, & Jex, 2004), but it is likely that more will follow.

In addition to these common archival data sources, organizational psychologists have also made use of other less common sources to study organizational processes. Sports statistics, in particular, are widely available and can be used to examine (albeit indirectly) organizational processes. Organizational psychologists, for example, have used the performance of professional baseball players to study equity theory (Lord & Hohenfeld, 1979), and professional hockey players to study leadership processes (Day, Sin, & Chen, 2004).

The use of archival data offers several advantages to researchers. First, many archival databases are readily available to the public and can be accessed quite easily—in many cases, via the Internet. Second, archival data are nonreactive. Archival data typically are not collected for the researcher's purpose, so there is no chance that participants will distort responses in a way that would impact the validity of the research. Finally, when archival data are used to measure employee behaviors, such records are usually less subject to distortion than self-reports of the same behavior.

Despite these advantages, the use of archival data may present several problems. One is that archival databases contain only indirect measures of the phenomenon of interest to the researcher. Use of databases such as the DOT or O*NET to measure characteristics of employees' jobs illustrates this problem quite well. Information contained in both of these databases is collected at the occupation level, so using it may mask important differences between individuals who may have the same occupation but perform substantially different work, or who perform under very different conditions. For example, a nurse employed in a rural health clinic may have very different job duties than one employed in a large urban hospital, even though they are part of the same occupation.

This issue becomes even more problematic when researchers use sports statistics to study organizational processes. Lord and Hohenfeld (1979), for example, examined the performance of baseball free agents in what is termed the *arbitration* year, or the year prior to going on the free-agent market. Based on the players' performance in this year, and the first year of their new contract, these researchers made inferences about how these players resolved their feelings of underpayment. What these researchers *didn't* do, however, was ask these players *directly* about whether they felt underpaid or how they planned to resolve feelings of underpayment.

Another potential problem with archival data is accuracy. Organizations differ widely in the precision of their record-keeping practices. Furthermore, there may be instances when an organization has some incentive to distort records. For example, organizations may underreport accidents or other negative incidents in order to avoid negative publicity or increases in insurance costs. Accuracy is probably less of an issue when archival data are obtained from government agencies and well-known academic research institutes. Nevertheless, when using any form of archival data it is always a good idea to ask for some evidence supporting the accuracy of the information.

Survey Research

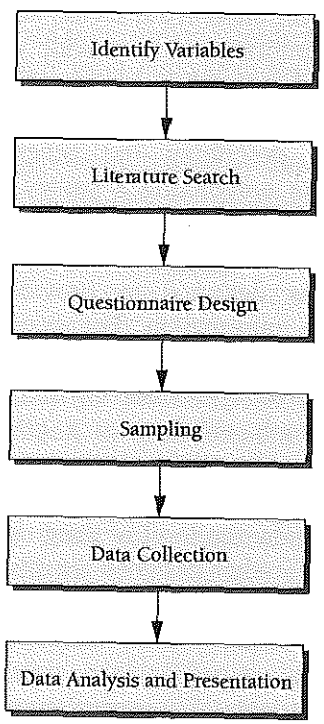
By far the most widely used form of data collection in organizational psychology is survey research (Scandura & Williams, 2000). Survey research simply involves asking research participants to report about their attitudes and/or behaviors, either in writing or verbally. This form of research is extremely common in our society and is used to gather information for a wide variety of purposes. Most readers have probably participated in some form of survey research.

Before describing the general steps involved in conducting a survey research project, it is useful to consider the purposes of survey research. In some cases, survey research is designed to provide purely descriptive information. For example, the top management team in an organization may wish to know the current level of employee job satisfaction, a government agency may want to assess the income level of working adults, or a research institute may want to know the level of drug use among teenagers. Studies designed for this purpose are often referred to as *prevalence studies*.

Survey research is also conducted to test hypotheses regarding the relationships between variables. For example, a researcher may want to assess whether employees who perceive a great deal of autonomy in their jobs also report a high level of job satisfaction. The researcher in this case is not really concerned with the actual levels of autonomy or job satisfaction, but rather is interested in whether these two variables are related.

As shown in Figure 2.1, the first step in conducting a survey research project is to identify the variables that one will be measuring. For theoretically based research projects, the variables will be directly linked to the research question one is examining. A researcher studying the relationship between interpersonal conflict on the job and employees' satisfaction with their jobs would obviously measure both of these variables. The choice of variables to be measured in

FIGURE 2.1
Steps Involved in Conducting a Survey Research Project



more applied research is often based on the concerns of upper-management personnel or, in some cases, input from employees from all levels of the organization. Determining what to measure in surveys is often achieved through the use of *focus groups* consisting of either top managers or groups of other employees. A focus group is a qualitative data-gathering technique that is often used to generate ideas during the preliminary stages of a research project. For example, to determine what to measure on a survey, a researcher might conduct a focus group with the top management of an organization. The researcher might begin the focus group session by posing a question: "What are the biggest concerns of employees in this organization?" This would be followed by an open-ended discussion, during which the researcher would take note of major issues that come up.

Once the researcher has decided which variables to measure, the next step is to conduct an extensive search of relevant literature on these variables. This is done to determine whether acceptable measures of the variables exist. For many variables of interest to organizational psychologists, several acceptable measures are available. Using previously developed measures saves a researcher considerable time because there is no need to develop new measures. While this can usually be done in theoretically based research projects, using established measures in applied projects such as employee opinion surveys is often more difficult to do. This is because many of the variables measured in employee opinion surveys may be unique to a particular organization. In the present authors' experience, organizations often want survey items *customized* in order to enhance the relevance of the information.

Once a researcher has decided on the variables to be measured and identified

acceptable measures, the next step is to design the questionnaire or survey instrument. This step is extremely important because the quality of the questionnaire will strongly impact the integrity of the data generated. Designing a high-quality survey instrument is a time-consuming, painstaking process. Fortunately, there are excellent sources of information one can refer to for assistance in the questionnaire design process (e.g., Dillman, 2000). One general rule should guide the development of any questionnaire: *It should be easy for the respondent to complete.* That is, instructions should be easy to understand, response categories should be well defined, and the items should be clearly written. It is probably for this reason that more and more researchers have been making use of Internet-based collection of survey data. Respondents need only click the appropriate response categories when completing an Internet-based survey, and simply click a send button when they're done. A potential disadvantage of Internet-based data collection, of course, is that the researcher has little control over who actually completes the survey instrument. It is also possible that samples generated via the Internet may systematically differ (e.g., they may be more educated) from samples generated through other methods.

Another important step in the questionnaire design process is to conduct some form of pilot testing, even if this involves simply asking a colleague to read through the questionnaire. Careful pilot testing may reveal unclear instructions, poorly worded items, or even misspellings.

After the questionnaire is designed and pilot tested, the next step is to determine specifically who the respondents will be. When research is conducted within organizations, this may simply involve asking all employees to complete the survey. In other cases, it may be necessary to narrow the pool of responding employees. For example, if a researcher were studying customer service behavior among employees, he or she would have to restrict the pool of respondents to those employees who have at least some contact with customers.

In some cases the number of potential respondents may be so large that it is impractical for the researcher to include everyone (e.g., a multi-national corporation with 50,000 employees). If this is the case, some form of *probability sampling* may be utilized. The idea behind probability sampling is that the researcher selects a sample from a larger group (or population) in order to generalize the results to that larger group, within some margin of error (Fowler, 1984). The most basic form of probability sampling is *simple random sampling*. This involves selecting members of a population such that all have an equal and nonzero probability of being included in the sample. As an example, a researcher could randomly select 200 employees from a current employee directory to participate in an organizational survey.

Another form of probability sampling sometimes used is *stratified random sampling*. This essentially represents the application of simple random sampling within identifiable groups or *strata*. Stratified random sampling is often used to increase the precision of estimates (Fowler, 1984); the logic is that if estimates are made within strata and pooled, these will be more precise than applying simple random sampling within an entire population. Stratified random sampling can also be used to increase the representativeness of samples. If, for example, an organization consists of five different employee groups that are represented in equal proportions, proportional stratified random sampling can be used to increase the chances that the sample will reflect this.

A third form of probability sampling that may be useful in some cases is *cluster sampling*. What distinguishes this from the other two forms of sampling previously described is that the unit of sampling is no longer the individual but, instead, some larger unit or *cluster*. An illustration of how cluster sampling can be used comes from a research project one of the authors conducted several years ago for the U.S. Army Recruiting Command (USAREC). This organization is very geographically dispersed and consists of multiple levels (brigades, battalions, companies, and stations). In the initial stages of the project, it was decided that approximately 50 face-to-face interviews needed to be conducted with personnel at brigade, battalion, and company levels. Rather than randomly selecting *individuals* from these three levels, it was decided to first randomly select two battalions within each brigade. Two individuals were interviewed in each battalion, as well as in the company located closest to each battalion.

A major advantage of cluster sampling is that it allows a researcher to cut down on travel time and expense. In the previously described project, imagine if simple random sampling had been used instead of cluster sampling. The 50 individuals selected to be interviewed may have been so geographically spread out that a separate trip would have been required to conduct each interview. Of course, the risk one runs in using cluster sampling is that the sample may not be as representative as it would be if simple random sampling were used. In most cases, though, researchers will accept the risk of decreased representativeness in order to cut down on costs (see Comment 2.2).

Once the researcher determines who the participants will be, the next step is to actually collect data. In collecting survey data, several options are available, and each option has advantages and disadvantages. With written organizational surveys, the ideal way to collect data is to have groups of employees complete the questionnaire in a centralized location and return the completed questionnaire to the researcher upon completion. This is ideal because it provides the best chance for a favorable response rate. A very low response rate is undesirable because it raises concerns about whether the survey results truly represent the target group. For example, in an organization where one of the authors once worked, an employee opinion survey was conducted and the response rate was approximately 10%! This low response rate was revealing in and of itself, but it also raised questions about the validity of the information.

In some cases, centralized data collection is not possible because of employees' schedules or concerns about confidentiality. Other options that are used in some cases are mailing questionnaires to employees' homes, administering a questionnaire verbally by telephone, or e-mailing questionnaires via the Internet or Intranet (i.e., internal system). Although these methods are somewhat less desirable than centralized on-site data collection, there are actually many ways that researchers can use them and achieve very favorable response rates (e.g., Dillman et al., 2000).

The final step in conducting a survey research project is the analysis and presentation of the data. The analysis of survey data is dictated by the purpose of the survey. If the purpose is description (which is usually the case when organizations initiate survey research projects), analyses are relatively simple and straightforward. Descriptive statistics (e.g., means, ranges, percentages) will usually suffice in such situations. In cases where survey data are used for theoretically

COMMENT 2.2

THE COST OF SURVEY RESEARCH

Survey research is by far the most commonly used data-collection method in organizational psychology. While survey research has many advantages, it is also true that surveys can be quite costly. Even if a researcher conducts a relatively simple, paper-and-pencil, self-administered survey of 200 employees in an organization, there are monetary costs associated with photocopying, incentives for respondents, and in many cases postage for return envelopes. This type of research project may also require that a researcher spend time personally contacting respondents and reminding them to complete the survey. Internet-based surveys reduce photocopying costs, but the time required to properly design a web-based data-collection tool is often considerable.

As the size and scope of a survey research project grows, researchers often must hire survey research firms to handle the data collection. This increases efficiency for the researchers, but unfortunately is very costly. For example, it is not unusual for survey research firms to charge several thousand dollars to collect survey data even when sample sizes are relatively modest. Typically when researchers wish to sample very large numbers of respondents, they must seek financial support from government agencies or private foundations in order to cover the costs of these projects.

based hypothesis testing, analyses are conducted to test hypothesized relations between variables. More detailed information on statistical analyses used in hypothesis testing will be discussed later in the chapter. However, it is worth noting here that survey data are typically best for assessing covariation among variables; that is, a change in one variable is associated with a change in another. Assessing whether one variable *causes* a change in another variable is difficult to do with survey data because such data are usually collected at one point in time.

One way to make survey data more amenable to the assessment of causality is to use a *longitudinal* design. An example of this might be measuring employees' job perceptions at one point in time, and then measuring job attitudes 6 months later. Compared to cross-sectional designs, longitudinal research is better because it at least gives the researcher a temporal basis on which to make causal statements. In the previous example, since job perceptions were measured *prior to* job attitudes it is certainly plausible that job perceptions might have a causal impact on job attitudes. As another example, Britt, Castro, and Adler (2005) examined whether being personally engaged in one's job could buffer soldiers from the adverse effects of working long hours and being away a large number of days for training. These authors found support for this hypothesis by showing that work hours and days training assessed at one point in time predicted health symptoms 3 months later only for soldiers who reported low levels of job engagement. However, an obvious downside to longitudinal research is that it is often impractical because researchers have to keep track of respondents. In recent years, many researchers have begun using very short-term intensive longitudinal designs in which participants provide a large number of observations over a short time period (e.g., Fuller et al., 2003).

Experimentation

Another common form of data collection in organizational psychology is *experimentation*. An experiment is a highly controlled situation that provides a researcher the best opportunity to assess cause-and-effect relationships. This is important because the hallmark of any science is to detect and explain causal relationships.

Because the term *experiment* is very commonly used, students are often unclear about what constitutes a *true* experiment. According to Cook and Campbell (1979), three characteristics distinguish a true experiment from other methods. These are (1) manipulation of an independent variable and measurement of a dependent variable; (2) random assignment to experimental treatment conditions; and (3) maximum control by the experimenter. Let's examine each of these characteristics.

The term *independent variable* is used to designate the variable that is proposed to have some effect on other variables, and hence is typically of primary interest to the researcher. When the independent variable is *manipulated*, this means that research participants experience different levels of this variable. If a researcher were interested in the impact of feedback on performance, for example, the independent variable would be feedback. This variable could be manipulated by providing one group of research participants with feedback after performing a task, while providing no feedback to a second group.

The measurement of the *dependent variable* simply involves some systematic record of the research participants' behaviors or attitudes that may be impacted by the independent variable. Choice of dependent measures is often based on prior research, or accepted practice. In organizational psychology, for example, it is common practice to measure attitudinal-dependent variables with surveys. It is always important, however, to keep in mind that the dependent measure being used is really just an operational definition of a concept. For example, job satisfaction represents whether a person has a positive or negative feeling about his or her job or a job situation. If a five-item scale is used to *assess* job satisfaction, this measure is really being used to represent this conceptual definition.

The second defining characteristic of experimentation, *random assignment*, means that research participants are assigned to groups receiving different levels of the independent variable (also called *treatment conditions*) in a random or nonsystematic fashion. Randomly assigning research participants can be done quite easily—for example, by flipping a coin. The logic behind random assignment is very simple—if research participants are assigned in a truly random fashion, it is likely that the different treatment groups will be similar in all ways except for the independent variable. This allows the researcher to isolate the independent variable as the cause of any differences between treatment conditions on the dependent variable.

The third defining characteristic of an experiment, *maximum control*, means that manipulation of the independent variable and the measurement of the dependent variable are done under controlled conditions. The researcher tries to make sure that all variables other than the independent variable are held constant. Like random assignment, this is done to isolate the independent variable as the cause of any differences among the treatment groups. When experiments are conducted in laboratory settings, researchers can usually achieve a desirable level of control. This is a much greater challenge when experiments are conducted in field settings, though not impossible. Eden (1985), for example, conducted a field experiment in the Israeli Defenses Forces to evaluate the impact of a team development intervention.

Quasi-Experimentation

According to Cook, Campbell, and Perrachio (1990), a quasi-experiment is similar to a true experiment except that it lacks one or more of the essential features previously described. In organizational settings, the independent variable of interest often cannot be manipulated because it is under the control of the organization, or may even be a naturally occurring event. Examples of independent variables that are usually under organizational control would include training programs or the redesign of jobs. Naturally occurring events that could be used as independent variables may include computer shutdowns, changes in government regulations, or mergers. In all of these cases, the researcher has no control over which research participants receive which *treatments*.

Quasi-experimental designs are also used in organizational settings because research participants usually cannot be randomly assigned to treatment conditions. Assignment to training programs provides a good example of nonrandom assignment. Employees typically participate in training programs, either voluntarily or on the basis of an identified training need (Goldstein, 1993). Thus, in most cases, if a researcher were to compare training-program participants to nonparticipants, these two groups could possibly differ in important ways.

Given the constraints that accompany quasi-experimentation, how do researchers set about proving that an independent variable has a causal impact on a dependent measure? One way is to measure and statistically control variables that may obscure the relationship between the independent and dependent variables. For example, if the average age of a group of employees receiving one level of the independent variable is higher than the age of groups receiving other levels, age can be measured and statistically controlled when comparing the groups. This would be using age as a *covariate*.

Other than statistical control, quasi-experimentation typically requires that researchers systematically identify and rule out alternatives to the independent variable when differences between treatment groups are found. According to Cook and Campbell (1979), there are a variety of explanations, other than the independent variable, that may lead to a difference between a treatment group(s) and a control group in quasi-experimental designs. For example, participants in different groups may be exposed to different historical events, participants may change at different rates, or participants may have differing views about participating in the research.

A researcher conducting a quasi-experiment can never know for sure whether any number of alternative explanations are impacting his or her findings. However, it is often possible to assess the *plausibility* of different alternative explanations. For example, let's say a researcher conducted a quasi-experiment in which the job of bank teller was redesigned at one branch of a bank, but remained the same at another. After 3 months it is found that customer satisfaction is much higher at the branch where the job redesign took place compared to the branch where the job was not changed. The job redesign may have caused the increase in customer satisfaction, but since this was not a true experiment, there may be explanations other than the job redesign. To rule out these alternative explanations, the researcher could begin by comparing these two branches to see whether any preexisting differences between employees in the two branches could have caused the difference in customer satisfaction. If the employees at these two branches were similar in terms of tenure and overall job performance, these could be ruled out as alternative explanations for the findings. The researcher could also gather information on the nature of the customers who frequent each of the two branches. If customers at the two branches tend to be demographically similar, and have similar income levels, this could also be ruled out as an alternative explanation of the difference in customer satisfaction. The researcher, in effect, plays detective in order to identify and rule out alternative explanations for his or her findings. Note that it is never possible to identify *every* possible alternative explanation, so researchers typically attempt to rule out only the most plausible.

Choosing Among Data-Collection Methods

Given the information presented about each method of data collection, readers may wonder how to choose which method to use. Unfortunately, there is no concrete formula for making this choice. Perhaps the best approach is to weigh the advantages and disadvantages of each method. As is illustrated in Figure 2.2, the primary advantage of observational methods is that they provide the researcher with an opportunity to study behavior in its natural context. Unfortunately, observational techniques tend to be highly labor intensive.

Archival data may allow researchers to avoid potential problems associated with self-report measures. An additional advantage of archival data is that they are often widely available. The primary disadvantage of archival data is that the researcher usually has little control over how such data were collected. That is, one must take on faith that such data have been properly collected and are accurate.

FIGURE 2.2

Summary of the Primary Advantages and Disadvantages Associated with Each of the Four Data Collection Methods

Observational Methods

Advantages	Disadvantages
• Behavior is captured in its natural context.	• May be highly labor intensive.
• Avoids the problem of "reactivity."	• Observations may be subject to bias.
• Some forms of observational data are readily available.	• Some forms of observational data only measure behavior indirectly.

Archival Data

Advantages	Disadvantages
• Easy to obtain	• Measures behavior indirectly.
• Non-reactive	• Not always accurate

Survey Research

Advantages	Disadvantages
• Allows the collection of data from large numbers of participants at low cost.	•Difficult to draw causal interferences from survey data
• Survey data can typically be analyzed with very powerful statistical methods.	• Response rates for some forms of survey data are low.
	• Survey design is a difficult, time-consuming process.

Experimentation

Advantages	Disadvantages
• Best way to assess causal relationships.	• Generalizability of findings may be questionable.
• Best way to isolate the impact of a specific variable.	•Examining a variable in isolation may be unrealistic.
• Gaining compliance of participants is easier compared to survey research.	• Participants may not take the experimental situation serious.

Quasi-Experimentation

Advantages	Disadvantages
• Allows the researcher a way to access causality in naturalistic settings.	•Organizations may be reluctant to allow these to be conducted.
• An excellent way to evaluate the impact of organizational interventions.	• Researchers have very little control.

Survey methodology allows the researcher to obtain data from a large number of participants at a relatively low cost. However, it is typically difficult to draw causal inferences from survey data, especially when the data are cross sectional. Experimentation provides the researcher with the best way to assess causal relationships. In some cases, however, the generalizability of experimental findings may be questionable. Finally, quasi-experimentation, in many cases, offers the researcher a way to assess causal relationships in naturalistic settings. However, quasi-experiments may be difficult to conduct because researchers typically have little control in most field settings.

Given the advantages and disadvantages summarized in Figure 2.2, the choice of a data-collection method depends largely on a researcher's objectives. If establishing causality is of primary importance, then experimentation is likely to be the method of choice. On the other hand, if capturing behavior in its natural context is the primary concern, then observation or quasi-experimentation may be preferred. Ideally, the best course of action is to use multiple methods of data collection (see Comment 2.3).

COMMENT 2.3

THE CASE FOR MULTIPLE DATA-COLLECTION METHODS

UNFORTUNATELY, A SIGNIFICANT portion of research in organizational psychology suffers from what has been termed *mono-operation* bias. This means that, in many studies, all of the variables are measured using only one form of data collection. Often, this one form of data collection is a self-report questionnaire, although it does not have to be. For example, a study would suffer just as much from this form of bias if all variables were measured using simple observation.

Why is it a problem to measure all variables in a study with only one form of data collection? One obvious reason is that the relationships among variables may be inflated because they share a common method (e.g., common-method bias). Another way to view this issue is to think about the positive impact of using multiple forms of data collection in a single study. Let's say a researcher is interested in whether job autonomy is positively related to job satisfaction. Further assume that, in this study, job autonomy is measured through a self-report measure completed by employees, and through archival information collected during a job analysis. Job satisfaction could be measured through a self-report measure and thorough observation of employees through their workday.

After these data are collected, we would likely find that the self-report autonomy measure would be positively related to the self-report job-satisfaction measure. However, what if the archival measure of job autonomy is also related to the self-report job-satisfaction measure? What if the self-report job autonomy measure is positively related to the observational measure of job satisfaction? If both of these results occur, this would most certainly strengthen the conclusion that job autonomy really does positively relate to job satisfaction. Thus, the real benefit of using multiple data-collection methods is that it allows us to show relationships between variables in multiple ways.

SPECIAL ISSUES IN DATA COLLECTION

Now that the most common methods of data collection have been described, we will explore, in this section, some important contemporary issues related to these methods. Contemporary issues include validity of self-report measures, generalizing laboratory findings to field settings, gaining access to organizations for data collection, and conducting research in different cultures.

Validity of Self-Reports

Self-report measures are used very frequently in organizational psychology. For example, employees are asked to report how much they like their jobs, how much variety they perceive in their work, how committed they are to their employing organization, and how anxious they feel about their jobs—just to cite a few examples. Because self-reports are used so frequently, we often don't give much thought to the assumptions we are making when we use such measures, or whether or not they are valid. Both issues are examined in this section.

Self-report measurement is really based on two implicit assumptions. First, we assume that respondents *know* the information we are asking for in self-report measures. Many of the questions asked in organizational surveys are subjective (i.e., there is no right or wrong answer), so it is pretty reasonable to assume that respondents know this information. Most people know whether they like their job, for example. In other cases, lack of knowledge may compromise the validity of self-report measures. For example, one of the authors once worked in a university system that conducts an annual survey of the job-related activities of faculty. Faculty were asked on this survey to indicate the number of hours in a typical week they devote to course preparation, teaching, research, and university service. While some university faculty may keep detailed logs of what they do each day, most probably have only a very vague idea of the number of hours spent on each of the activities on the survey.

A second assumption underlying self-report measurement is that respondents will be truthful in their responses. Compared to researchers interested in some forms of behavior (e.g., drug use, criminal activity), organizational psychologists are relatively fortunate in this regard. Because most of the items on organizational surveys are not highly sensitive or invasive, employees probably respond truthfully to such items, provided they believe their responses will be held in confidence. In reality, however, employees' comfort levels with surveys vary greatly. For example, when organizational researchers use self-reports to measure things such as absenteeism, turnover intentions, or various forms of counterproductive behavior (e.g., theft, sabotage), employees may not answer truthfully. In such cases, all a researcher can really do is take great care to reassure employees and conduct the survey in such a way that supports the promise of confidentiality. This might include providing employees with stamped envelopes to mail completed surveys to the researcher offsite, or perhaps making sure there is no identifying information contained on the survey instrument itself.

The situation that has generated the most controversy surrounding the use of self-reports is when such measures are used to rate job and organizational conditions. A researcher, for example, may ask respondents about the level of time pressure in their jobs. According to Spector (1994), self-reports often do not correlate well with more

objective measures of the work environment, such as ratings by job analysts or by others familiar with the same job (Liu, Spector, & Jex, 2004; Spector, Dwyer, & Jex, 1988; Spector & Jex, 1991). Use of self-report measures is also controversial when such measures are correlated with other self-report variables. When this is the case, the correlations between such variables may possibly be inflated due to *common method variance*—a term that is used quite frequently but is rarely explicitly defined. Common method variance represents shared sources of measurement bias between two variables that can be directly tied to the method of measurement being used (Spector, 1987b). As an example, let's say that a researcher is measuring two variables via self-report. Further assume that both of these measures, for some reason, are impacted by social desirability responding (Crowne & Marlowe, 1964)—that is, responses to items in both measures differ in their levels of social desirability. This shared source of measurement bias may lead these two variables to be correlated, even if there is little or no underlying conceptual relationship between the two variables. In cases in which these measures are conceptually related, the presence of common method variance may inflate the magnitude of the relationship between the two variables.

Should researchers be concerned about common method variance? The consensus in the literature seems to be "Yes" (e.g., Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). However, empirical efforts to actually demonstrate the effects of common method bias on relationships between variables have provided only mixed results. Spector (1987b), for example, empirically investigated the prevalence of common method variance in the measurement of job characteristics and job satisfaction. Based on an analysis of several data sets, he concluded that there was no strong evidence that correlations were inflated due to common method variance.

Spector's (1987) investigation prompted several attempts to replicate his findings; most of these attempts utilized more complex statistical techniques (e.g., Bagozzi & Yi, 1990; Williams & Anderson, 1994; Williams, Cote, & Buckley, 1989). A complete discussion of the findings of these studies is beyond the scope of this chapter, but the general conclusion of these studies was that the impact of common method variance is greater than Spector had estimated. However, as Brannick and Spector (1990) pointed out, there are problems in the use of complex statistical methodology to test for the effects of common method variance.

Perhaps the best way to empirically assess the impact of common method variance is to compare correlations that contain a shared method with those that do not share a method. Crampton and Wagner (1994) conducted a meta-analysis in which they summarized 42,934 correlations from studies using single and multiple methods. Overall, they found that correlations between variables that were both measured via self-report were not appreciably larger than other correlations. In the measurement of some variables, however, correlations based on a single source were larger than others. This suggests that the impact of common method variance is real; however, the magnitude of this effect varies widely, depending on the nature of the variables being measured.

The best conclusion one can draw about the validity of self-report measures is that it depends primarily on the variable being measured, and the research question being asked. For example, if one were interested in measuring employees' feelings about their jobs, then a self-report measure would be quite appropriate. On the other hand, if one were interested in measuring employees' levels of job autonomy, levels of discretion in decision making, or (perhaps) workload, then measuring these variables only with self-report measures is not really appropriate. This is because in all of these examples the researcher is interested in characteristics of the *environment*, not those of the individual employee. When researchers wish to measure characteristics of the work environment, the best course of action is to use multiple measurement methods (e.g., Glick, Jenkins, & Gupta, 1986). Given the reliance of much organizational research on self-report measurement, it is likely that the pros and cons of self-report measurement are likely to be debated for quite some time (see Comment 2.4).

COMMENT 2.4

THE SELF-REPORT CONTROVERSY

Self-report measurement is undoubtedly the most common form of data collection in organizational psychology. It is also a form of data collection that has evoked a great deal of controversy, particularly when self-reports are used to measure all of the variables in a study. Dr. Steve Jex has followed this issue for over a decade, primarily because it has a great deal of relevance to his own research program in occupational stress, since self-report measures tend to predominate.

On the positive side, self-reports allow us to measure something that is important in determining human behavior—namely, individuals' perceptions of their environments, their emotional states, and, in some cases, their views of other people. Self-report measurement is also very economical. In the time it might take to collect meaningful observations of 20 people, a self-report measure could be distributed to 100 times that many people.

The primary drawback to self-report measurement is that humans are not analytical instruments; thus, self-reports may not always produce accurate information. For example, when we ask employees to provide self-reports of characteristics of their jobs, these ratings may be biased by internal mood states, social influences of coworkers, or stable internal dispositions (Spector, 1994). These same biases may also influence self-reports of emotional and affective states.

What is the most reasonable conclusion one can draw about self-report measures? In my opinion, it is that self-report measurement, like any other data collection method, has both advantages and disadvantages. Whether one uses self-report measurement should be dictated primarily by the variables one is trying to measure, which are ultimately dictated by the research question one is trying to answer. As a general rule, if one is primarily interested in perceptions, then self-report measurement is a logical choice. However, if one is interested in actual environmental conditions, then self-reports should be supplemented with other forms of data collection.

Source: P. E. Spector. (1994). Using self-report questionnaires in OB research: A comment on the use of a controversial method. *Journal of Organizational Behavior, 15,* 385–392.

Generalizing Laboratory Findings

A common criticism of psychology is that it is a science based largely on laboratory studies that investigate the behavior of white rats and college students. Research in organizational psychology tends not to be conducted as much in laboratory settings in comparison to other areas of psychology (e.g., physiological, cognitive). Laboratory studies do, however, still account for a substantial portion of the research in both organizational psychology and I/O psychology in general (Locke, 1986; Sackett & Larsen, 1990; Scandura & Williams, 2000). The purpose of this section is to examine the issue of whether findings from laboratory investigations can be generalized to *real* organizational settings.

The strongest argument made against laboratory findings' generalizing to field settings is that laboratory situations lack realism. University laboratories are not real organizations; thus, laboratory settings lack what is referred to as *mundane realism*. Realism, however, must also be considered from the perspective of the research participant. It is certainly possible to place a research participant in a situation that lacks mundane realism, yet manipulate variables in such as way that participants react genuinely to the situation. For example, one can be in a completely contrived situation yet still feel pressure to perform well or conform to group norms. When this is the case, it can be said that there is a high degree of *experimental realism* for research participants. Many classic laboratory studies conducted over the years, particularly in social psychology (e.g., Asch, 1957; Milgram, 1974), have lacked mundane realism yet have retained a very high degree of experimental realism.

Another reason laboratory investigations may generalize has to do with the research participants. At the beginning of this section, it was remarked, somewhat facetiously, that laboratory investigations often utilize college students as participants. This often leads to the criticism that, because college students are different from the general population, research findings cannot be generalized. This criticism certainly does have some merit—college students are more intelligent and typically come from higher income levels than the general population (Sears, 1986). However, for the study of many organizational issues, use of college students as research participants probably does not compromise generalizability a great deal. College students, for the most part, represent the cadre of individuals who will hold many of the white-collar jobs in the future. Thus, they may be quite similar to such employees, both in terms of attitudes and abilities, even though they are obviously lacking in relevant organizational experience. By contrast, college students are probably a poor research sample if the aim is to generalize to employees holding blue-collar and manual labor jobs.

Despite these arguments for the generalizability of laboratory experiments, there are clearly important differences between laboratory and field settings. In particular, the high level of experimental control in laboratory settings allows the researcher to isolate the impact of a variable in a way that is impossible in field settings because so many things are occurring that the impact of any single variable may be greatly diluted. Also, when variables are investigated in laboratory settings, they are taken out of their natural context. By taking a variable out of context, the researcher runs the risk of changing the substantive nature of that variable. A good illustration of this point is laboratory research on the effects of ambient temperature on aggression (e.g., Baron & Bell, 1976). In a laboratory setting, it is possible to completely isolate the impact of temperature. In natural settings, however, temperature increases often occur in conjunction with other variables such as loud noise and crowding.

Another important difference to consider is that laboratory settings are typically short term (Runkel & McGrath, 1972). As a result, participants in laboratory investigations have very little time invested and have no reason to form any social ties with others. In contrast, employees in organizations invest a considerable amount of time in their jobs, and often develop important social ties with fellow employees. These differences between laboratory research participants and actual employees may lead to very different reactions to the same situations.

A final important difference between laboratory and field settings is the nature of the tasks performed by research participants. Since laboratory investigations are short term, it is very difficult to match the complexity of the tasks performed by employees in real organizations. Thus, many laboratory studies ask participants to perform relatively simple tasks such as assembling tinker toys, solving anagrams, and putting together puzzles. In contrast, employees in organizational settings perform much more complex tasks.

After examining the pros and cons of laboratory investigations, we are still left with the question of whether laboratory findings generalize to field settings such as organizations. Unfortunately, there is no definitive answer to this question, although it has been examined extensively (e.g., Berkowitz & Donnerstein, 1982; Dipboye & Flanagan, 1979). The most comprehensive analysis of this issue, relevant to organizational psychology, is contained in *Generalizing from Laboratory to Field Settings*, a book edited by Edwin Locke in 1986. The general conclusion one can draw from this book is that well-designed laboratory investigations often do generalize to field settings. A well-designed laboratory investigation is one in which participants are highly engaged in the task being performed and variable(s) of interest are well simulated. The reader should be cautioned, however, against concluding that all findings do or do not generalize. In the end, generalizability is an empirical question, and the best course of action is to replicate laboratory findings in field settings whenever possible.

Gaining Access to Organizations

One of the biggest challenges from field research is simply gaining access to an organization for data collection. The authors have known, over the years, many colleagues and students who have come up with interesting research questions, but could find no organization in which to collect data. Unfortunately, there is very little in the organizational literature to help guide researchers in their efforts to gain access to organizations. Thus, most of this section is based on both the authors' experience as researchers, and the experiences of fellow organizational researchers.

Before exploring ways to gain access to organizations, consider reasons why organizations would *not* let a researcher gather data. Based on past experience, there are two primary reasons: (1) data collection usually requires employees' time, and (2) organizations are concerned that employees may divulge sensitive or proprietary information about the organization. Organizations that operate in very competitive industries (e.g., consumer products, high technology) are often very concerned with divulging any information that might put them at a competitive disadvantage. In such organizations, the secrecy surrounding activities such as product development often carries over to other activities, regardless of whether the concerns are warranted.

Given these potential objections to the collection of research data, how can organizational researchers still gain access? Perhaps the most fundamental suggestion that can be made in this regard is: Ask. Many researchers who complain about lack of access have actually asked relatively few organizations for their cooperation. They simply assume that they will be unable to collect data. One way to enlist an organization is to contact several organizations by telephone and try to make contact with someone in the human resources department. Another approach is to *mass mail* to organizations, asking for cooperation. T. E. Becker (1992), for example, mailed letters to the presidents of 30 organizations asking for permission to collect data and eventually collected data in one of these.

General appeals or *cold calling* may result in a data-collection opportunity, but it is often much more efficient to use established connections in organizations. Most people have family and/or friends who work in organizations, and such people may be in a position to either authorize the collection of data or put the researcher into contact with someone who has the authority to do so. This suggests that researchers should not be afraid to use established connections in organizations. Researchers should also invest time and energy to develop connections with people who can help with data collection in their organizations in the future. This often takes time and energy but, in the long run, the contact may result in excellent data-collection opportunities (see Comment 2.5).

Let's now assume that a researcher has persuaded an organization to at least consider the possibility of data collection. How can a researcher convince an organization to actually go ahead with data collection? The most useful suggestion that can be made in this regard is: The researcher should offer the organization something in return for its cooperation. For example, researchers often provide a summary of the research findings to the organization, in return for its cooperation. Other researchers may offer to perform some consulting service at no cost to an organization. Organizations typically do not provide a researcher with access to their employees unless the access will provide some tangible benefit in return.

After an organization gives permission to collect data, there is often some negotiation between the researcher and the organization, regarding issues such as research design and measures. At this stage, researchers and organizations often clash, because of their differing goals and objectives. Researchers typically desire a high level of methodological rigor in their investigations because their ultimate goal is to publish their findings in peer-reviewed journals. Unfortunately, methodological rigor may be perceived by the organization as costly in a number of ways. For example, supplementing self-report measures with organizational records may be time consuming and require that employees reveal identifying information. It may also be impossible for an organization to allow a researcher the control needed for experimental or even quasi-experimental investigations. This is a tricky issue for researchers to navigate because just gaining access to organizations is such a challenge. The key is this: The researcher must be willing to accommodate the organization, but not to such an extent that it completely compromises the scientific integrity of the investigation. Unfortunately, researchers often severely compromise the methodological rigor of studies without attempting to persuade organizations of their value. In most cases, a well-designed, methodological, rigorous study will not only help the researcher but will also be more informative to the organization (Campion, 1996).

COMMENT 2.5

GAINING ACCESS TO ORGANIZATIONS: SOME EXAMPLES

STEVE JEX: AS I wrote the section on gaining access to organizations, I thought of the various ways I have gained access to organizations in order to collect data. Like many researchers, I have used family connections. For example, I was able to gain access to an insurance company in Tampa, Florida, to conduct my Master's thesis research while in graduate school. My wife was employed there at the time. To this day, I can't figure out whether my wife was trying to advance science, or just wanted me to get out of graduate school! Another study I conducted, which was ultimately published in the *Journal of Applied Psychology* (Jex, Beehr, & Roberts, 1992), was actually made possible through the efforts of my mother. This study was conducted at a hospital in Saginaw, Michigan (my hometown), where my mother was employed as a nurse. She introduced me to a person in the human resources department who was ultimately able to grant me access to all hospital employees. In this case, I think my mother's help was driven primarily by a desire to see her son get tenure. In addition to using family connections, I have gained access in many other ways. In some cases, current and former students have helped facilitate data-collection efforts. I have also, on occasion, relied on former graduate school classmates, or other colleagues, to provide either data-collection sites or useful contacts.

Is there any underlying theme when I think about the various ways in which I have gained access to organizations? The most obvious theme is that developing and maintaining relationships with people is important. This includes family, students, and professional colleagues. I'm not suggesting that relationships should be initiated only on the basis of what people can do for you. However, the fact is, it is much easier to ask someone for assistance if you've taken the time to maintain an ongoing relationship with him or her. The other important lesson I've learned over the years is simply to ask. We often assume incorrectly that family, friends, and colleagues do not want to be bothered helping with data collection. However, my experience has been that people often are very willing (and even flattered) to help if they're asked.

Thomas Britt: Getting access to samples is a critical issue when trying to do quality organizational research. When I was conducting primarily social-psychological research I took for granted how easy it was to obtain samples by using the subject pool of students taking "Introduction to Psychology." The more I got into organizational psychology, the more I realized the difficulty and tenacity required to gain access to samples. When I was in the U.S. Army as an organizational psychologist, I had a somewhat *captive audience* of soldiers to participate in research projects (although it should be noted that all soldiers provided informed consent to participate). However, even when conducting research with soldiers, it was necessary to convince unit commanders of the importance of the research and why it was worth the time of their soldiers when they could be spending more time training. You also had to provide the commanders with summaries of the results that told them something important, and provide recommendations for what they should do given the results.

Since arriving at Clemson I have continued my ties with the military but have also started collecting data from different applied samples. For example, I have recently begun an assessment of staff at Clemson University examining the influence of positive motivational states at work on well-being and performance. I have worked closely with the administration at all levels to convey the importance of the research (and why we need supervisor ratings of performance in addition to employee reports), and my graduate students and I have spent a great deal of time coordinating with individual units in the university to ensure a high response rate of participation. I have also been struck by the diversity of jobs people have in a university setting (e.g., firefighters, library personnel, campus recreation, facilities). When orienting new graduate students into our Industrial/Organizational Psychology program, I emphasize that getting access to a quality sample of employees who will be willing to work with you on a research question is often more difficult than coming up with the hypothesis you want to test!

Source: S. M. Jex, T. A. Beehr, and C. K. Roberts. (1992). The meaning of occupational "stress" items to survey respondents. *Journal of Applied Psychology, 77,* 623–628.

Conducting Research in Different Cultures

Given increasing globalization, it is more and more common for organizational psychologists to examine cross-cultural issues. Despite the value of cross-cultural research, data collection in such studies is often challenging for a number of reasons. For example, when self-report measures are used, these often must be translated from one language to another. This may seem rather simple; often, it is not. The typical procedure used to translate self-report measures into different languages is called *back translation.* This involves translating the items on a measure from one language to another (e.g., from English to Chinese), and then back to the original language. The researcher must assess whether the items have retained their meaning to respondents after being translated from a different language.

Another issue researchers must consider in conducting cross-cultural research is sampling. Researchers conducting cross-cultural research often want to compare employees in one culture to employees in another, so it is important to utilize samples that are similar in all aspects except culture (Arvey, Bhagat, & Salas, 1991). The ideal way to accomplish this would be to utilize employees from different cultures who work for the same organization (e.g., De La Rosa, 2006). If this could not be done, a researcher would typically want to select samples from different cultures that work in the same industry and perhaps have similar levels of work experience.

Researchers conducting cross-cultural research must be on the lookout for things that are specific to a given culture and may adversely affect data collection. For example, a researcher utilizing self-ratings of performance must be aware of the fact that, in Asian cultures, it is considered improper to rate oneself high in performance (Fahr, Dobbins, & Cheng, 1991). There may also be vast cultural differences in research participants' degree of comfort when they are asked to provide ratings of persons in positions of authority (Hofstede, 1980).

STATISTICAL METHODS IN ORGANIZATIONAL PSYCHOLOGY

Regardless of the data-collection method used, once data are collected, researchers must analyze those data to assess whether or not their hypotheses are supported. Fortunately for organizational researchers, many statistical methods are available to help make sense out of data. Because a comprehensive review of statistical methodology is beyond the scope of this chapter, we will review, in this section, the statistical methods that are used most frequently in analyzing research data.

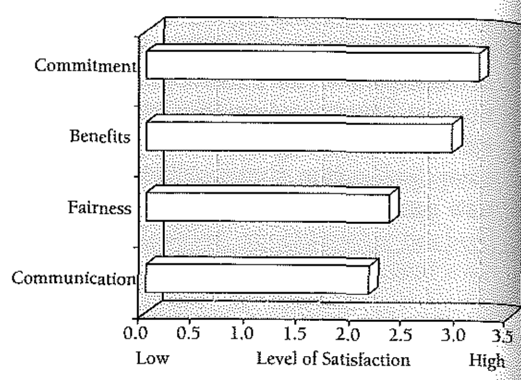
Descriptive Statistics

The first thing a researcher needs to do after obtaining a set of data is to get a feel for general trends. For example, if we were to collect data on job satisfaction within an organization, two relevant questions might be (1) what is the *overall level* of job satisfaction in the organization, and (2) are employees very similar in their levels of job satisfaction, or do they vary widely? To answer the first question, some descriptive measure of *central tendency* would be used. The most commonly used measure of central tendency is the mean (also called the *average*), which is calculated by simply adding up all of the scores on a variable and dividing by the total number of scores. Other common measures of central tendency include the median and mode. The *median* is the score on a variable that splits the distribution into two equal halves. Unlike the mean, the median is unaffected by the presence of extremely high or extremely low values. Because of this, the median is useful as a supplement to the mean, in cases in which a distribution contains extreme scores. The *mode* is simply the most frequently occurring score and is typically not very informative unless there is a very dramatic preference for one response over others.

Measures of central tendency are useful because they provide information about the manner in which variables are distributed. This is important because most statistical methods are based on assumptions about the manner in which variables are distributed. Measures of central tendency are also valuable when organizational policy makers

FIGURE 2.3

Graphical Representation of Mean Levels of Four Dimensions Measured in an Employee Opinion Survey.



Notes: Communication = Satisfaction with amount of communication in the organization; Fairness = Satisfaction with level of fairness in the organization; Benefits = Satisfaction with current fringe benefit package; Commitment = Organizational commitment. Mean values may range from 1 to 4.

are assessing survey results. Figure 2.3, for example, contains a graphical representation of employee opinion survey data collected by one of the authors. This figure graphically represents the mean values of four dimensions contained on the survey. A quick perusal of this figure indicates relatively low satisfaction with the levels of communication and fairness in this organization. On the other hand, employees in this organization appear to be committed to the organization and are reasonably satisfied with their fringe benefits package. While this is certainly not complicated information, it could be important to an organization. In this case, the organization used the information as the basis for interventions to enhance communication and fairness.

In addition to measures of central tendency, researchers often want to know whether responses are uniform or whether there is a great deal of dispersion. The most basic measure of dispersion is the *range*, which is the difference between the highest and lowest value for a particular variable. It is often useful to compare the observed range for a given variable to the possible range. For example, if a variable is scaled such that it may range from 10 to 50 and the observed range is 30 to 50, this indicates potential problems with range restriction.

While the range may be useful in identifying problems with range restriction, it is still a very crude measure of dispersion. More precise and more commonly used measures of dispersion include the *variance* and *standard deviation*. The variance represents the variability of scores around the mean. To calculate the variance, you simply subtract the mean from each score in a distribution, square each value, add up these squared values, and divide by the total number of scores. The standard deviation is simply the square root of the variance.

Given the way in which the variance and standard deviation are calculated, higher values indicate greater dispersion about the mean. The standard deviation is also useful because it can be used in converting raw scores to *standard scores*. A standard score is simply the score on a given variable, expressed (in terms of its distance from the mean) in standard deviation units. The simplest form of standard score is a *z-score*, which is calculated by subtracting the mean from a raw score and dividing the result by the standard deviation. Standard scores can be useful in cases in which the researcher wishes to compare a respondent's scores on different variables that may utilize different scales of measurement (see Comment 2.6).

A final type of descriptive measure that is used in the analysis of research data is *reliability*. Reliability is defined as the extent to which a variable is being measured without error (Nunnally & Bernstein, 1994). What is

COMMENT 2.6

CONFESSIONS OF A STATISTICAL MINIMALIST

In his initial statement as editor of *Journal of Applied Psychology* in 1995, Philip Bobko referred to himself as a "statistical minimalist" (Bobko, 1995, p. 4) in describing his views on statistical analysis. What is a statistical minimalist? Perhaps the best way to understand this is to consider more of Bobko's editorial statement. Specifically, he advised potential authors: "Please look at 'simple' statistics, such as means, standard deviations, correlations, effect sizes, and so forth. And do not just look at them; consider them when attempting to understand and explain what's going on. I believe that one can often (usually?) learn more by looking at these simple statistics with a critical and understanding eye than one can learn by computing the newest fashion in statistics with an amazed eye" (p. 4).

The important point that Bobko was trying to make in this editorial is that even relatively simple descriptive statistics are important if one's goal is to understand their data. A more subtle message here is that the choice of statistical methods to use should be driven by the question being asked, not by the latest fad. Although not always the case, it is often possible to answer important research questions without resorting to overly complex statistical analyses.

Source: P. Bobko. (1995). Editorial. *Journal of Applied Psychology, 80,* 3–5.

considered error, however, depends on the particular context in which a measure is being used. When multi-item measures are used, which is typically the case in organizational research, it is necessary to assess the *internal consistency reliability*. A measure of internal consistency reliability provides an estimation of the extent to which all items on a scale are measuring the same attribute. Suppose, for example, we constructed a five-item measure of job satisfaction. If internal consistency reliability were estimated to be very high, this would suggest that all five items were measuring the same thing.

In other cases, researchers must provide other reliability estimates. For example, if a variable is going to be assessed at multiple points in time, it is important for the researcher to show that the measure of the variable is not strongly impacted by random fluctuations over time. In this case, an appropriate form of reliability assessment would be *test-retest reliability*, which simply involves administering a measure at two different points in time and calculating the correlation between these scores. If this correlation is high, it suggests that the measure is not strongly impacted by random temporal fluctuations.

Another form of reliability assessment, *interrater reliability*, may be necessary in cases in which multiple raters are utilized to assess some attribute of a person (e.g., performance) or the environment (e.g., job characteristics). There are many ways to assess interrater reliability, but they all basically allow the researcher to assess whether the ratings provided by different raters are similarly ordered. The researcher can also assess whether raters agree on the absolute value of the ratings. This issue will be discussed in greater detail in the final section of the chapter, which deals with aggregation and levels of analysis issues.

Why do researchers need to be concerned about reliability? The answer to that question has to do with the nature of measurement error. Measurement error, by definition, represents sources of influence on the measure other than the hypothesized construct. A *constant* or systematic error would be the tendency of a respondent to answer all items in a way that he or she feels is socially desirable. A *random* error would be a momentary distraction causing as respondent to respond to an item "Strongly Agree" when he or she really meant to respond "Strongly Disagree." When a measure is reliable it is relatively free of random error, though it may still contain considerable constant error. When a measure is unreliable, however, it contains a great deal of measurement error. This is problematic because random error, by definition, is unrelated to other variables; thus, reliability sets an upper bound on the magnitude of relationships between a measure and other variables.

Tests of Mean Differences

After assessing descriptive measures, researchers should hopefully be able to conclude that there are no major distributional problems, and that all variables are measured with a minimal amount of error. If this is indeed the case, the next step is to perform some analysis to test whatever hypotheses are being proposed. There are many different types of hypotheses; a common type of hypothesis involves testing differences in the mean level of a given variable. For example, a researcher may hypothesize that employees in white-collar jobs have higher organizational commitment than blue-collar employees, or that the performance of groups that participate in team-building activities is higher than that of groups not participating. In this section, we cover the two most common statistical tests of mean differences.

Before describing these statistical tests, it is useful to provide a brief overview of the logic behind tests of statistical significance. Regardless of the statistical test being used, a test of statistical significance essentially involves establishing a rule for distinguishing chance from nonchance outcomes. All statistical significance tests begin with the assumption of what is termed the *null hypothesis*, which is another way of saying there is no effect or no relationship between variables. Assuming that the null hypothesis is true, it is possible for a variety of research outcomes to occur simply on the basis of chance. Thus, the researcher needs some decisive rule for determining whether a given result represents a chance occurrence or a legitimate scientific finding. The standard used most often for distinguishing chance from nonchance—the one that has come to be adopted in the behavioral sciences over the years—is 5%. Assuming that the null hypothesis is true, if the probability of a research outcome occurring by chance is 5% or less, scientists typically conclude that it is a legitimate scientific finding (e.g., they reject the null hypothesis). Thus, when the statement is made that a finding is "significant beyond the .05 level," the researcher is saying that it is very unlikely that the finding observed is a chance occurrence.

When testing mean differences, the simplest scenario is testing the difference between two groups. For example, a researcher may wish to test whether the average age of those who participate in training and development activities differs from those who choose not to participate. The statistic most commonly used in this situation would be a *t-test*. The magnitude of the t statistic depends on the absolute difference between means relative to the level of variation within the groups being compared. Thus, even if the absolute difference between the means is substantial, a high degree of variation within the different groups will keep the t value at a relatively low level, and lead the researcher to conclude that there is no meaningful difference between the groups.

There are other instances in organizational research in which the means of more than two groups must be compared. For example, a researcher might want to compare the mean level of job satisfaction in several different work groups that have and have not participated in team development activities. In this case, the statistical procedure used would be *analysis of variance*. The general purpose of analysis of variance is to assess the variation between different groups, relative to the variation within groups. To perform an analysis of variance, it is necessary to calculate several different variance estimates or *mean-squares*. These are used to estimate the variance between groups and the variance within groups. The actual test of statistical significance employed in analysis of variance is the *F-test*, which is simply a ratio of the variance between groups to the variance within groups. When an F is statistically significant, this indicates that the ratio of variance between groups to the variation within groups is very unlikely to have occurred by chance, given the null hypothesis. Recall that the same basic logic is employed with the t-test. If a statistically significant F is found in analysis of variance, this indicates that there is some difference among the means in the groups of interest, although it does not tell the researcher which means are different. To figure this out, follow-up tests would be used to assess the difference within each possible pair of group means.

Given the basic logic behind analysis of variance, this statistical procedure can be used a variety of ways. For example, different forms of analysis of variance can be used to assess (1) the impact of multiple independent variables, (2) repeated measures of dependent variables, and (3) the impact of multiple dependent variables. Readers interested in more detailed information on analysis of variance procedures should consult Keppel and Zedeck (1989).

Correlation and Regression Analysis

Given the prevalence of cross-sectional field surveys in organizational research, hypotheses are often tested by assessing the covariation among the variables of interest. The most commonly used statistical index of covariation is the *Pearson product-moment correlation coefficient*. The correlation coefficient can range from +1.00 to −1.00, but typically falls in between these values. The larger the absolute value of a correlation coefficient, the greater the degree of covariation. This degree is often expressed by squaring the correlation coefficient to obtain the amount of shared variation between two variables. For example, if the correlation between two variables is .30, they share 9% of their variance in common (e.g., $[.30]^2$). When the sign of a correlation is positive, this simply means that two variables covary in the same direction. A negative sign, by contrast, indicates that two variables covary in opposite directions.

The correlation coefficient is useful in testing many hypotheses in organizational research, but it provides very limited information about causal relationships. For example, if job satisfaction were found to be correlated with job autonomy, this could mean that high job autonomy causes employees to be more satisfied with their jobs. On the other hand, it could also mean that a high level of job satisfaction causes employees to see greater levels of autonomy in their jobs. It is also possible that two variables may be correlated primarily because of the influence of a third variable; for example, employees who have higher salaries may be more satisfied and tend to hold jobs with high autonomy. If this is the case, it is said that the relationship is *spurious*.

Correlational analysis is also limited by the fact that only two variables may be examined at a time. In many instances, a researcher may be interested in the extent to which several variables are related to some other variable of interest. For example, a researcher may be interested in the degree to which pay, length of service, level of performance, age, and job type all contribute to employees' overall satisfaction with their employing organization. One way to address this question would be to examine the correlation between job satisfaction and each of these variables individually. Unfortunately, such an analysis does not provide the researcher with information about the extent to which this entire set of variables is related.

The statistical procedure that is used to assess the relation of a set of variables (called predictors) to another variable (called the criterion) is *multiple linear regression* or, simply, multiple regression. Multiple regression is useful because it provides a quantitative estimate of the amount of covariation between a set of predictors and a criterion variable. This is assessed by the multiple R statistic, which is analogous to the correlation coefficient. In most instances, however, researchers report the squared value of multiple R, which serves as a measure of the amount of variance in the criterion variable that is explained by a set of predictors.

Multiple regression is also useful because it allows the researcher to assess the relative

impact of each predictor in explaining the criterion variable. When a set of predictors is used to estimate a criterion variable, the criterion is estimated to be a linear function of the predictor set. The general form of this equation is:

$$Y = A + B_1X_1 + B_2X_2 + \ldots B_kX_k$$

where Y is the criterion variable that is being predicted, the Xs represent the predictor variables, A is a constant, and each B-value represents the weighting of a given predictor or the extent to which it contributes to the prediction of the criterion. The advantage of using these statistical weights, as opposed to correlations, is that they are calculated in a way that takes into account the intercorrelations among the other predictor variables in the set. Thus, B-values in multiple regression represent the *unique* contribution of a given variable to the prediction of some criterion measure.

Beyond correlation and regression analysis, many other related methods can be employed for data analysis. Most of these fall under the general category of *multivariate methods* (e.g., Tabachnick & Fidell, 1996) and, due to their complexity, are not covered in this chapter. These methods are quite useful to the researcher, particularly in field investigations. Like all statistical methods, they should be used only if necessary to test a given hypothesis.

Meta-Analysis

A final form of statistical analysis that is being used increasingly in organizational research is *meta-analysis*. Meta-analysis involves the quantitative summary of research findings and is typically used in research domains where a considerable number of studies have been conducted (Rosenthal, 1991). For example, meta-analyses have been conducted on the relation between job satisfaction and job performance (Podsakoff & Williams, 1986), the impact of unemployment on well-being (McKee-Ryan, Song, Wanberg, & Kinicki, 2005), and the impact of different leadership styles (Judge, Piccolo, & Illies, 2004). In all three cases, so many studies have been conducted that it would be difficult to provide an accurate qualitative summary of the findings.

Statistically, meta-analysis essentially involves averaging effect sizes (e.g., from correlation coefficients or differences between two means). Before these effects sizes are averaged, however, researchers typically control for a number of *statistical artifacts*—factors that may lead to differences in the findings between studies. The most basic statistical artifact is sample size. Studies with larger sample sizes need to be weighted more heavily when averaging correlations compared to those with smaller sample sizes. Another common statistical artifact controlled in meta-analyses is measurement unreliability. Earlier in the chapter, *reliability* was defined as the degree to which a variable is measured without error. When measurement procedures are unreliable, this means that they contain considerable error. As was stated earlier, this is important because it sets a lower boundary on the degree to which a variable can be correlated with other variables. Controlling for unreliability puts all variables on a level playing field in terms of measurement error.

The other common statistical artifact controlled in meta-analyses is *range restriction*. In some studies, correlations between variables may be reduced because the values do not cover the entire possible range. This may occur because of a variety of factors (e.g., Johns, 1991), but it always serves to limit the magnitude of correlations. When researchers control for range restriction, they are estimating what the correlations would be if the variables of interest were measured without any range-restriction problems.

Once all relevant statistical artifacts are controlled, two important statistics are typically calculated in meta-analysis. Most researchers calculate some overall estimate of effect size between two variables. This estimate represents the effect size after controlling for the impact of important statistical artifacts, and it provides a good estimate of the *true* relationship between variables. The other statistic typically calculated in meta-analysis is the amount of variation in effect sizes that remains after important statistical artifacts are controlled. Usually, after important statistical artifacts are controlled, there is a relatively small amount of variation between studies' findings. However, if there is still a substantial amount of variation, factors other than statistical artifacts may be contributing to the differences in findings between the studies. Such factors are called *moderator variables*. Some of the more typical moderator variables examined in meta-analyses include aspects of the study design (laboratory experiment versus field study), characteristics of the research samples (employees versus college students), and specific measures used to assess key variables (well-established measures versus measures developed for one study).

SPECIAL ISSUES IN STATISTICAL ANALYSIS

At this point, readers should have a basic understanding of the more typical statistical methods used in organizational psychology. The purpose of statistical methodology is to help researchers answer questions (i.e., it is a means to an end), but it has also become a vibrant field of inquiry in and of itself. In fact, within organizational psychology, many researchers focus on statistical and methodology issues. Because of this focus, several issues in statistical methodology have surfaced over the years and have been the subject of inquiry and debate. In this section, we briefly review four important contemporary issues in the use of statistical methodology in organizational research.

Statistical Power in Organizational Research

Statistical power refers to the sensitivity of statistical tests to detect meaningful treatment effects. To use an analogy, one might think of the statistical power of different tests in the same way as differences between types of microscopes. An inexpensive microscope purchased from a toy store provides some magnification, but extremely small objects (e.g., viruses) cannot be detected. In contrast, an expensive electron microscope provides a much higher level of magnification that allows for the detection of even extremely small particles.

Several factors contribute to statistical power (Cohen, 1992). One is *sample size*. All things being equal, larger sample sizes provide higher levels of statistical power. This is one reason why survey researchers are concerned about nonresponse, and laboratory researchers are concerned about participants not showing up. A second factor impacting power is *effect size*, or the relative strength of the effect a researcher is trying to detect. There are actually several ways to express effect size, but the easiest way to explain it is based on the size of correlations. Generally speaking, if the true correlation between two variables is very small, this effect is much harder to detect than a much larger effect. Smaller-effect sizes require a more powerful "microscope" for detection.

A third factor that impacts statistical power is the *alpha level* chosen in statistical significance testing. The alpha level represents the cutoff for distinguishing chance from nonchance findings. Recall, from the previous discussion of statistical significance testing, that 5% has become the conventional rule in the behavioral sciences. The reason that the alpha level is set so low is to reduce the probability of committing a *Type I error*, or falsely concluding that one has uncovered a legitimate scientific finding. In an organizational setting, an example of committing a Type I error would be falsely concluding that a training program had a positive effect on employee performance. In contrast, a *Type II error* is committed when a researcher fails to detect a legitimate effect when it is present. In the previous example, this would involve conducting a statistical test and falsely concluding that a useful training program had no impact on employee performance (see Comment 2.7).

As the alpha level becomes more stringent (e.g., smaller than 5%), this reduces the chance of committing a Type I error, but also tends to reduce power and hence increases the chances of committing a Type II error. In contrast, a more liberal alpha level (e.g., 10%) tends to increase power, although this comes at the cost of an increase in the probability of committing a Type I error.

A final factor impacting power is measurement error. Specifically, higher levels of measurement error are associated with low levels of power. This is simply due to the unsystematic nature of measurement error, which was discussed earlier.

COMMENT 2.7

TYPE I VERSUS TYPE II ERROR: WHICH IS THE GREATEST SIN?

GIVEN THE FACT that the alpha level is typically set at .05 or, in some cases, even .01, one would assume that committing a Type I error is a bad thing. Recall that when a Type I error is made, a researcher concludes that a finding is scientifically meaningful when it really is not. Why is this bad? From a scientific point of view, Type I errors are bad because they lead us down *blind alleys*, and ultimately may lead to faulty theories. From a practical point of view, a Type I error may lead an organization to spend a considerable amount of money on a training program that ultimately is not effective. Given these negative effects of a Type I error, we want to minimize the chance that one will occur, so we set alpha at a very low level.

Unfortunately, in minimizing the chances of Type I error, we increase the chances of Type II error. As you recall, Type II error is committed when a researcher fails to uncover a legitimate scientific effect. Is it better to make a Type II than a Type I error? It really depends on the situation. Let's say, for example, that a researcher is testing a drug that could potentially neutralize the HIV virus. It would obviously be bad if this researcher were to falsely conclude that this drug was effective (e.g., commit a Type I error). However, consider the implications of committing a Type II error in this case. If this drug is effective, and research does not show it, a great chance to reduce human suffering has been missed.

Ultimately, research should be designed to balance the risks of both Type I and Type II errors. To minimize the risk of Type I error, alpha levels should be set sufficiently low, and proper statistical procedures should be used. On the other hand, Type II error can be minimized primarily by employing adequate sample sizes and reliable measures.

Given the previously described determinants of statistical power, let us now consider the level of statistical power in organizational research. Mone, Mueller, and Mauland (1996) examined this issue in a meta-analysis of the level of power in 26,471 statistical tests from 210 research studies conducted between 1992 and 1994. These authors also explored common practices with respect to the assessment of power prior to conducting research.

The results of the meta-analysis were revealing—and, in fact, somewhat troubling. Given that an acceptable level of statistical power is considered to be 80% (e.g., there is an 80% chance of detecting a true effect; Cohen, 1992), the authors found that across all effect sizes, an acceptable level of power was achieved only 50% of the time. What this means is that across all studies in this meta-analysis, researchers assume a 50% chance of failing to detect a true effect when it is present. This suggests that many studies conducted in organizational research are underpowered.

Low statistical power is especially problematic when researchers are attempting to detect small effect sizes. When Mone et al. (1996) calculated the level of statistical power for small effect sizes, it was found that the percentage of studies achieving an acceptable level of power was only 10%! That is, the vast majority of studies attempting to detect small effects are grossly underpowered. This is unfortunate because small effects are very common in organizational research, due to the vast number of variables impacting employees in organizations.

The results of the survey of authors were also revealing. Perhaps the most important finding was that 64% of the authors surveyed reported that they do not perform any type of power analysis prior to conducting a study. One reason frequently cited for this was that, in many cases, researchers have little or no control over sample sizes in field research. Thus, even if a power analysis indicated that a larger sample size would be desirable, it would not be possible to increase. Many authors in this survey also noted that scholarly journals do not insist on power analysis during the review process, although there are some exceptions (e.g., Campion, 1993). This is unfortunate because scholarly journals serve an important *gate-keeping* function, and insistence on power analysis would serve to heighten awareness of this issue. As it stands right now, there are probably many meaningful effects in organizational psychology that go undetected due to low statistical power.

Detection of Moderator Variables

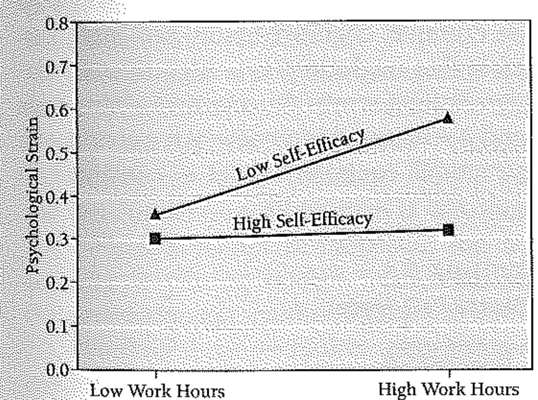
Recall from the section on meta-analysis that a moderator variable changes the relationship between two other variables (James & Brett, 1984). More specifically, the relationship between two variables differs at different levels of the moderator variable. In organizational psychology, many theoretical models contain moderator variables; thus, it is important to understand the statistical procedures used for assessing whether or not moderated relationships exist.

There are actually several ways to test moderator effects (e.g., see James & Brett, 1984), but the most commonly used procedure is through the use of multiple regression analysis (Cohen & Cohen, 1983). In this procedure, which is known as *cross-product regression*, the independent variable is first entered into the regression equation. In the next step, the moderator variable is entered. In the final step, the cross-product of the independent variable and moderator is entered. The cross-product term is created by multiplying the independent variable by the moderator for each respondent. If the variation explained by the cross-product term is statistically significant, a moderated relationship is present. This means that the relationship between the independent variable and the dependent variable differs as a function of the moderator. This is usually shown visually by plotting the relationship at high (one standard deviation above the mean) and low (one standard deviation below the mean) levels of the moderator. Figure 2.4 illustrates how this is done. In this case, self-efficacy moderates the relationship between work hours and psychological strain. Notice that, when self-efficacy is low, there is a positive relationship between work hours and psychological strain. In contrast, when self-efficacy is high, there is essentially no relationship between these two variables.

The procedure for detecting moderator variables is rather straightforward, but, in practice, the actual detection of moderators is difficult, primarily because of low statistical

FIGURE 2.4

Graphical Representation of a Moderated Relationship



Source: S. M. Jex and P. D. Bliese. (1999). Efficacy beliefs as a moderator of the impact of work-related stressors: A multilevel study. *Journal of Applied Psychology, 84*, 349–361. Copyright © 1999 by the American Psychological Association. Reprinted with permission.

power. This is because moderator effects are typically small, since the variance explained by a moderator effect is that which is left over after the effects of the independent variable and moderator are taken into account. Power is also reduced when the independent variable and the moderator are strongly correlated and, in the case of dichotomous variables (e.g., race, gender), when the proportion differs greatly from 50/50 (Aguinis & Stone-Romero, 1997).

What can be done to increase the power of moderator tests? Given the previous general discussion of statistical power, researchers testing moderator effects should try to employ large samples and highly reliable measures. A somewhat more controversial way to increase power is to increase the alpha level beyond the conventional .05. Recall that the alpha level represents the researcher's decision rule for distinguishing chance from nonchance findings. If a less stringent alpha level of .10 is adopted, for example, this means that results with a 10% or lower probability of occurring by chance are considered legitimate treatment effects.

Given the low power associated with moderator tests, the decision to adopt a less stringent alpha level would appear to be logical. It is not extremely unusual to find researchers using alpha levels of .10 in moderator tests (e.g., Jex & Elacqua, 1999), but the practice is not widespread. This is likely due to the fact that the .05 level is highly ingrained in our thinking. Most students are taught that an alpha level beyond .05 is "cheating," and they are extremely reluctant to raise it.

Beyond statistical considerations in the detection of moderator effects, it is always good practice to have a solid theoretical rationale before searching for moderators. Often, moderator variables that are very intuitively appealing may not be theoretically

COMMENT 2.8

THE ELUSIVE MODERATOR EFFECT

As will become evident as readers make their way through this book, many theories in organizational psychology propose moderator hypotheses; that is, certain relationships may hold under certain conditions, but not under others. Moderator variables are important in theory development because they allow us to specify the precise conditions under which some phenomenon may occur. They also may have a great deal of practical value, for example, by providing an organization with guidance about whether there are certain conditions under which interventions such as job redesign may or may not work.

Despite the theoretical and practical value of moderator variables, they are very difficult to demonstrate empirically. The is primarily due to the fact that moderator variables almost always explain a small portion of the variance-dependent measures and, as a result, statistical power to detect these effects is often very low. Thus, in many cases, researchers propose theoretically sound moderator hypotheses yet come up empty when they test for these effects. What can researchers do to avoid this fate? The most logical steps one can take to increase statistical power of moderator tests are: employ large sample sizes, utilize reliable measures, adopt a reasonable alpha level, and try to cut down on extraneous sources of variation.

justified. Statistical methodology will never compensate for poor theory development (see Comment 2.8).

Use of Causal Modeling

Over the past 20 years, a statistical technique that has become increasingly popular in organizational psychology—and many other fields—is *causal modeling* (James, Mulaik, & Brett, 1982). The logic behind causal modeling is that the researcher derives a set of predictions about how a set of variables relate to one another, and tests all of these relations simultaneously. In practice, this is typically done through the use of either *path analysis* or *structural equation modeling*. With path analysis, the variables that constitute a causal model are the actual variables that are measured. This is illustrated in the simple path model depicted in Figure 2.5. This model proposes that high levels of cognitive ability and work experience lead to high levels of job knowledge, which in turn leads to high levels of job performance. Structural equation modeling is similar to path analysis except that the variables comprising the causal model are *latent* rather than measured variables. A latent variable is a hypothetical variable that is purported to cause the interrelationships among measured variables. As an example, cognitive ability is a latent variable that might lead to a high intercorrelation among scores on cognitive ability tests and a structured interview. An example of a structural equation model is contained in

FIGURE 2.5
Simple Path Model

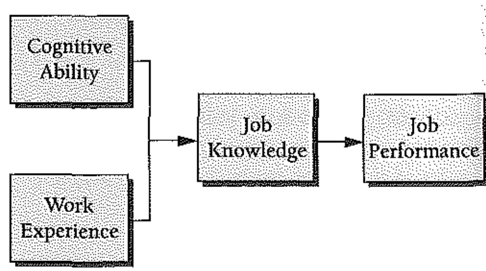


FIGURE 2.6
Simple Structural Equation Model

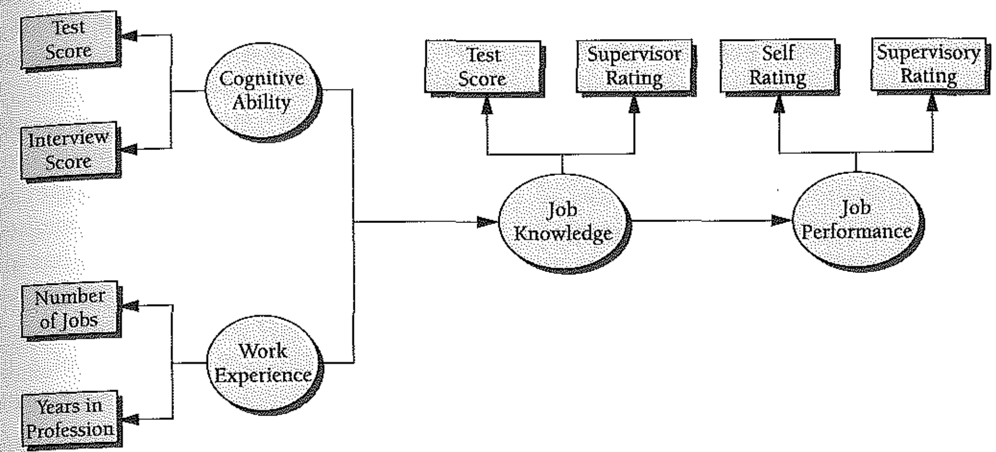


Figure 2.6. The circles are meant to denote latent variables, and the boxes represent measured variables. Continuing with the example from Figure 2.5, scores on a cognitive ability test and a structured interview are used as indicators of the latent variable cognitive ability, and so on. Notice that this is essentially the same model depicted in Figure 2.5. The only difference is that the proposed relationships are among latent rather than measured variables.

Once a model is proposed, the researcher seeks to assess whether the model *fits* the actual data. There are actually several indexes of model fit (Bentler, 1990), but the logic underlying all of them is very similar. When a model is proposed, the researcher is placing certain restrictions on the covariation among the variables of interest. Based on these restrictions, an expected covariance matrix of relations among variables in the causal model can be calculated. This expected covariance matrix is then compared to the actual covariation among the variables in the proposed model. When a model is said to "fit the data well," this means that the actual covariation among the variables closely matches that which would be expected, based on the proposed relations among the variables.

Causal modeling is a powerful technique because it allows the researcher to simultaneously test all the relations comprising an entire theoretical model. With correlation and regression analysis, it is usually possible to test only *parts* or individual segments of a theoretical model. The use of causal modeling, however, has been somewhat controversial. Some of these controversies are technically beyond the scope of this chapter and are related to things such as parameter estimation methods and the assessing model fit. Some, however, have questioned whether this technique has been overused, and whether model tests have been too data driven and not grounded enough in theory.

Like any statistical technique, causal modeling is neither good nor bad. If applied properly, it can be a very useful part of an organizational psychologist's statistical tool kit. Generally speaking, causal modeling is most powerful when the model being tested has a strong theoretical base, and there is a fairly large sample available. It is only at this point that a researcher has enough insight to propose the complex set of interrelations among variables that comprises most causal models. Thus, it is usually not appropriate to use causal modeling early in a theoretically based research program.

Aggregation and Levels of Analysis

A recent trend in organizational psychology is the exploration of variables at multiple levels of analysis; that is, researchers have increasingly become interested in the impact of variables that are conceptualized not only at the individual level but also at group and even organizational levels. Researchers have also become interested in how variables at different levels of analysis impact each other. This latter type of investigation is known as *cross-level analysis*.

Exploring multiple levels of analysis obviously presents researchers with some important theoretical issues (e.g., Bliese & Jex, 2002; Chan, 1998; Klein, Dansereau, & Hall, 1994). However, with these theoretical considerations come methodological and statistical considerations as well. Let us first consider the issue of *aggregation*. When data are aggregated, this simply means that one value is used to represent the unit of aggregation. An example of this would be using the mean level of job satisfaction within a work group to represent "group-level satisfaction." Note that when a variable is aggregated, all individual differences within the unit of aggregation are suppressed.

When is it appropriate to aggregate individual responses? Generally speaking, researchers must be prepared to justify aggregation on three different levels. First, there must be theoretical justification. The issue here is whether the variable created through aggregation is theoretically meaningful. In the example provided in the previous paragraph, the researcher would need to make the case that the average level of job satisfaction within a work group is a theoretically meaningful variable.

If aggregation is theoretically justified, the researcher must also provide some methodological justification for the decision to aggregate. The most basic methodological question a researcher faces has to do with choosing the unit of aggregation. Most researchers would probably find it acceptable to aggregate the responses of a five-person work group whose members interact frequently. There would likely be less agreement on this issue if one were to aggregate the responses of one division of an organization consisting of 100 members. Unfortunately, there is no hard and fast rule regarding what is and what is not an appropriate unit of aggregation; ultimately it comes down to the variable one is measuring (Bliese & Jex, 2002).

A second methodological issue has to do with the measurement of variables. In many cases, individual responses are aggregated because items make reference to respondent perceptions of the unit of aggregation. For instance, if a researcher were to measure organizational climate (James & Jones, 1974), the items should make reference to the organization and not the individuals responding. This suggests that researchers should make the decision to aggregate before data are collected so items can be reworded appropriately.

Assuming that aggregating is justified theoretically and methodologically, researchers must also be prepared to justify aggregation statistically. In most instances in which individual responses are aggregated, the researcher is doing so in order to measure some attribute of the unit of aggregation. For example, a researcher may want to measure the level of cohesiveness in a group, or the level of trust within an organization. In such cases, it is incumbent upon the researcher to justify aggregation by showing some statistical evidence of agreement in responses within the unit of aggregation. If respondents within a group do not agree on the level of cohesiveness within the group, it usually makes little sense to average their responses. There are several ways to measure interrater agreement, but the most frequently used method has become the r_{wg} statistic (James et al., 1984).

Besides aggregation, the other major issue confronting researchers exploring multiple-level issues is statistical analysis. In any research investigation, the choice of statistical analysis is driven by the research question being asked. Thus, in some cases, the analysis of multilevel data is relatively straightforward. For example, if a researcher were interested in the relation between group cohesiveness and group performance, it would make sense to examine the correlation between aggregate-level measures of both of these variables. The only drawback to this approach, of course, is that it greatly reduces sample size and, hence, statistical power.

In other instances, the analysis of multilevel data is more complex because researchers wish to examine the effects of multiple levels within the same analysis. For example, a researcher may be interested in estimating the relative contribution of individual-level versus group-level effects. In other cases, researchers may be interested in exploring the impact of group- or organizational-level variables on the relation between individual-level variables. In still other cases, researchers may wish to examine the behavior of a small number of research participants over many occasions (e.g., diary studies); such data are *multi-level* because individuals are nested within time periods or measurement occasions. Fortunately, statistical procedures are available to researchers in order to allow the analysis of data at multiple levels.

To examine cross-level relations, a statistical technique that has become increasingly popular is *random coefficient modeling* (Bliese & Jex, 2002; Byrk & Raudenbush, 1992). Random coefficient modeling can be used, for example, to test whether the magnitude of relations between individual-level variables (represented by regression coefficients) differs as a function of some aggregate-level variable. While both of these techniques are very useful, they are also very complex, and they require the use of special computer software. However, if used appropriately, both can help researchers untangle the complexity of multilevel data.

CHAPTER SUMMARY

This chapter explored the methodological and statistical foundations of organizational psychology. As was shown, organizational psychologists have several options when collecting data about behavior in organizations. These range from simple observation methods to highly complex quasi-experimental investigations. The most frequently used technique, however, is survey research.

In the collection of data in organizations, several important issues must be considered. For instance, researchers need to be cognizant of the limitations of self-report measures and aware of limits on the generalizability of research findings across research settings. When cross-cultural research is attempted, researchers must be attuned to issues of language and sampling. A more practical issue is simply gaining access to organizations to collect research data.

A variety of statistical methods were discussed that can be used to analyze data once they are collected. These range from simple descriptive statistics to more complex correlation and regression analysis. The choice of any statistical technique is dictated by the nature of the question the researcher is attempting to answer.

In the statistical analysis of data, a number of important issues must be considered. Researchers should be aware of the importance of statistical power and attempt to maximize it whenever possible. This is particularly true when researchers are interested

PEOPLE BEHIND THE RESEARCH

PAUL D. BLIESE AND THE IMPORTANCE OF STATISTICS

I've always been attracted to the basic tenets of statistics—the idea that probability can be used to detect patterns in complex data. What fascinated me about the area of multilevel statistics was the idea that researchers and practitioners could make accurate predictions about central tendencies of groups even in cases where the seemingly same data failed to make accurate predictions for individuals. In the Army, this meant that we might be able to predict the average health and well-being of groups of soldiers (platoons, companies) even if we could not necessarily predict the well-being of individual soldiers in the groups.

I became interested in this topic area early in my career when I observed a strong correlation between the average number of hours a group worked and the average well-being of the group members. When I analyzed the same data at the level of individual work hours and individual well-being, I observed a much weaker relationship. Over the years, I noted that patterns such as those involving work hours and well-being were common in organizational psychology, and I began to explore the conditions that caused correlations to differ across levels (as had many others).

Eventually, this work led me to become interested in the idea that a variable like work hours might actually change meaning when it was averaged within groups. Up to this time, the dominant idea had been that variables maintained the same meaning in both their individual-level form and their aggregated form. For an example of a change in meaning, consider that individual reports of work hours vary for numerous reasons to include work requirements, desire to get ahead and work ethic. When the variable is aggregated by averaging across groups, however, individual differences wash out, and one is left with a group mean that reflects work requirements imposed on the group. In this way, the two variables (average group work hours and individual work hours) have a subtle but important change in meaning across levels.

What have I learned from all this? I guess the answer is that taking the time to track down answers to questions such as "why do two correlations differ?" can lead to an entire career's worth of work.

LTC Paul D. Bliese
Commander, U.S. Army Medical Research Unit-Europe.

in demonstrating the effect of moderator variables. Complex statistical techniques, such as causal modeling, can be useful tools to organizational researchers, provided they are used judiciously and are based on sound theory. The exploration of multilevel data has become increasingly popular in organizational psychology in recent years. Researchers conducting multilevel analysis must be prepared to justify aggregation, and must choose the analytical technique that best represents the substantive issue of interest.

SUGGESTED ADDITIONAL READINGS

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregatrion and analysis. In K. J. Klein & S. W. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 349–381). San Francisco: Jossey-Bass.

Dane, F. C. (1990). *Research methods*. Pacific Grove, CA: Brooks-Cole.

De LaRosa, G. M. (2006). *Towards an understanding of individual ratings of cohesion within work units: A multilevel study*. Unpublished master's thesis. Bowling Green State University, Bowling Green: OH.

Hays, W. L. (1988). *Statistics* (4th ed.) Fort Worth, TX: Harcourt Brace Jovanovich.

McCall, M. W., Jr., & Bobko, P. (1990). Research methods in the service of discovery. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1 pp. 381–418). Palo Alto, CA: Consulting Psychologists Press.

Schaffer, B. S., & Riordan, C. M. (2003). A review of cross-cultural methodologies for organizational research: A best practices approach. *Organizational Research Methods*, *6*, 169–215.

Schaubroeck, J., & Kuehn, K. (1992). Research design in industrial and organizational psychology. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology, 1992* (pp. 99–121). Chichester, England: Wiley.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.) New York: HarperCollins.