CHAPTER 3

VALIDITY, CONTROL, AND GENERALIZING

IN CHAPTER 2, you learned that a crucial feature of experimentation is the need to control for unwanted sources of variance. If secondary variance is not controlled and error variance is not minimized, then the validity of the experiment is compromised. In 1957, Campbell distinguished between internal and external validity. Ideally, researchers should aim for both. However, as you will see, one paradox of experimentation is that by maximizing internal validity, the researcher may be reducing external validity and vice versa. On the other hand, depending on the experimenter's intent, this tradeoff between the two types of validity may not be a serious concern.

INTERNAL AND EXTERNAL VALIDITY

Internal Validity

Internal validity refers to the researcher's ability to state that the relationship she or he predicted between the independent and dependent variables does indeed exist, and that this relationship is due to the effects of the independent variable and not to extraneous, uncontrolled variables. Simply put, if the study is well designed and well controlled, and if alternative explanations for the results can be ruled out, it is said to have internal validity. It is crucial to understand the concept of alternate explanations. In every experiment the researcher forms a hypothesis, or a prediction for the outcome of the experiment. When the study is finished, that hypothesis is either supported by the data or it is not supported. If the data support the hypothesis, the researcher needs to be able to say that the results are due to the manipulation of the independent variable, and not to outside factors such as extraneous variables. In other words, alternative or rival explanations need to be ruled out.

Let us return for a moment to the memory experiment example presented in Chapter 2, where you learned that if you tested all the participants in the concrete condition in the morning and all the abstract condition participants in the evening, you would have confounded your study. The presence of the confounding variable—the time of day when participants were tested—would not allow you to rule out alternative explanations for your findings. The superior performance in the concrete condition may be due to participants being more alert and refreshed in the morning, and not to the effects of the independent variable.

Extraneous, uncontrolled variables lead to alternate explanations for the outcome of the experiment, which is why control procedures are so crucial. The more alternative explanations exist, the less validity the study has. Thoughtful researchers deal with this problem right from the start, when they first design the experiment. Once the independent variable has been selected and its levels have been determined, potential alternative explanations are considered. In other words, the researcher asks, "What other factors could affect my dependent variable? What other explanations would be possible for the experiment's outcome?" Once these potential threats to internal validity have been identified, the researcher makes every effort to eliminate or control for them, even if it means having to redesign the experiment.

External Validity

External validity refers to the *generalizability* of the study. Recall that in Chapter 1 you read about samples and populations. The researcher selects a sample of participants from the population of interest, conducts the study, and then generalizes the findings from the sample back to the population. The extent to which the experimenter can make generalizations about the results is external validity. If the researcher can extend the findings beyond the experimental setting and the sample of participants tested, then the study is said to have external validity. In sum, external validity is the extent to which the observed relationship between the independent and dependent variables in the experiment can be generalized to other settings, other people, and other circumstances.

For example, let's say that a researcher was interested in developing a new method for teaching vocabulary words to elementary school children. She hypothesized that children taught with vivid imagery would learn the words faster than would children taught without imagery. She randomly assigned 30 children to the imagery condition and 30 children to the no imagery condition. In the imagery condition the children saw a list of words with a picture next to each word, while in the no imagery condition the children learned the same list of words but without the pictures. As predicted, the children learned significantly more words in the imagery condition. At this point, if the researcher can generalize the findings from her *sample* of 60 participants in the experiment to elementary school children in *general,* then the study has external validity. In other words, if imagery truly enhances learning, then elementary school children from a wide variety of schools should benefit from learning vocabulary words accompanied by imagery.

Internal vs. External Validity

Earlier in the chapter you read about the possibility of a tradeoff between internal and external validity. By now you know that the more carefully you control the experiment, the more internal validity the study has. However, the stricter you control conditions in a laboratory setting, the less the laboratory setting may resemble real life and the less you may be able to generalize your findings. Should this be a concern? The answer depends on the type of research that is being conducted: **basic research** or **applied research.**

Basic research involves the testing of hypotheses generated by a particular theoretical framework. The purpose of basic research is to gather a body of "evidence" or knowledge that supports a given theoretical position. Although it would be an added bonus for the scientist to be able to apply his or her findings to real-life situations, this ability to go beyond the laboratory setting is not the primary goal of the researcher. For example, let's say that you are interested in the Levels of Processing (LOP) theory of human memory (Craik & Lockhart, 1972). According to Craik and Lockhart, the more deeply the information is encoded, or processed, during study time, the greater the ability to retrieve the information during test time. Given this condition, you may decide to test this theoretical position by hypothesizing that participants who were given a deep level of encoding instructions during the study phase would recall more words than participants who were given a shallow level of encoding instructions. At this point, you may not be concerned with the applied value of the LOP theory, but only with finding support for it in the laboratory. Therefore, you would be less concerned with trading off external validity for strong internal validity.

Applied research, on the other hand, is conducted in order to predict behavior in the real world (as opposed to the laboratory) or to solve real-life problems. If the intent of the researcher is to predict behaviors in their natural setting, or to solve problems in the real world, then the researcher should be concerned with external validity. In general, applied research strives for external validity: the ability to apply the findings to real-life situations and settings.

The fields of basic and applied research are not necessarily two separate and distinct dichotomies. Frequently they overlap and, even more frequently, applied research is founded on basic research. An example of this overlap, or interrelationship, can be seen in an area of psychology known as *human factors.* Human factors research is the application of knowledge regarding human abilities such as memory, learning, perception, attention, and so on, to the design of tools, machines, or manuals used by people in the real world. In other words, scientists engaged in human factors use the knowledge derived from basic research on human abilities in order to design, among other things, more user-friendly computers, software, aircraft cockpits, telephones, and answering machines.

This is not to imply that basic research has nothing to do with the real world or that scientists engaged in basic research have no interest in how their findings may apply to real life. It is merely to point out that whether the researcher is willing to compromise external validity for the sake of internal validity, or vice versa, depends on the primary goal of the research project. In fact, the long-held belief that external validity is compromised by strengthened internal validity has been questioned recently (Anderson, Lindsay, & Bushman, 1999).

Anderson, Lindsay, and Bushman conducted a massive study that compared the findings of laboratory research and field research articles. They analyzed 288 articles published in psychological journals, and their question of interest was whether the effects of the same independent variables on the same dependent variables were consistent in laboratory and field settings. They were interested in seeing whether the results found in a laboratory were comparable to results found in real-world or field research settings. Their results showed a high consistency between laboratory and field settings, indicating that the artificial controls applied in the laboratory did not compromise external validity. Anderson et al. (1999) offer the following statement regarding the concern over the internal and external validity tradeoff:

> *As long as scholars in both settings keep in mind the complementary pitfalls of too little control over extraneous variables (leading to low internal validity) and of overgeneralizing from the specific features of a specific study (leading to low external validity), we believe the psychological research enterprise will continue to succeed. (p. 8)*

SECONDARY VARIANCE AND CONTROL

Secondary variables can affect both the internal and external validity of the experiment, depending on their source. In the previous chapter you were introduced to the concept of secondary variance and to Kerlinger's Principles of Control. According to Kerlinger, secondary variables should be eliminated when possible or, if that is not possible, held constant. In addition, secondary variables can also be randomized and in certain instances even be made into or considered an additional independent variable. The following section lists some potential sources of unwanted variance. As you study these sources, keep in mind that whether the experimenter should worry about a particular source of secondary variance depends on the research design.

Sources of Secondary Variance

Factors that affect internal validity include the following:

- **Nonrandom assignment** of participants into the conditions. If the participants are not randomly assigned into the conditions, all the "competent" participants may end up in the experimental condition, while the "less competent" ones may end up in the control condition. In this instance, the scores of the experimental group will be superior to the scores received by the control group, regardless of how the independent variable was manipulated.

- **Maturation** of participants, which refers to how participants change over time. As the experiment progresses, especially if it is a lengthy one, people grow tired, frustrated, hungry, or bored. As such, the dependent variable measure may indicate maturation effects rather than effects of the independent variable. In addition, as you will see in a later chapter, when the experimental design entails pretest and posttest measures, the possibility that maturation is confounded with the independent variables is a possibility, if the experimenter does not control for it.
- **Order or pretest effects,** which occur when performance on a second task is affected by the first task regardless of the independent variable. For example, students may do better on the Scholastic Assessment Test (SAT) the second time due to practice effects, not because they enrolled in a training course. In a similar vein, participants may make more errors on the second task not because it is more difficult than the first task, but because of fatigue or boredom.
- **Regression toward the mean,** which is a statistical phenomenon that tends to occur when participants are selected on the basis of extreme scores on some measure. For example, let's say you are interested in studying low- and high-sensation seekers. If you select participants on the basis of their extremely low- or extremely high-sensation seeking score on a personality inventory, over time the high scorers will regress toward the mean or average sensation-seeking score. Therefore, if you were to assess their sensation seeking a second time, they would tend to have lower sensation-seeking scores even if they were not exposed to any treatment between test 1 and test 2.
- **History,** which refers to the passing of time between multiple observations. Studies that entail a before and after measure—meaning before and after the introduction of the independent variable—may have history as a potential confounding factor. Events that occur between the multiple measures (test 1, test 2, and so on) may affect the dependent variable and mask the true effects of the independent variable. For example, let us say that an investigator was interested in seeing whether people's attitudes regarding the treatment of prisoners could be manipulated by showing a film. On Friday, he measured the participants' initial attitude toward the treatment of prisoners, and on Monday, he showed a film that was contrary to the initial attitude. Following the film, the participants' attitude was once again measured. Unfortunately for the experimenter, on Saturday there was a major prison riot, which was covered extensively by the media. As you can imagine, any change (or for that matter, no change) in the participants' initial attitude might be due to this intervening event rather than to the effect of the film.
- **Subject mortality,** which affects studies that are either longitudinal, meaning long term, or require several experimental sessions. Participants may drop out of the study altogether or may not show up for a subsequent session. This tends to affect the more tiresome, difficult, or frustrating experimental condition, resulting in differential loss of participants. For example, if the independent variable is stress, and the experimenter is measuring the number of errors made as a function of stress, it

is highly probable that it will be the high-stress condition participants who will not return to complete the study.

- **Different time and place** of testing, which should be held constant for all participants.
- **Differential treatment** by the experimenter which could refer to any factor that is not held constant. For example, all participants should receive the same instructions, the same number of practice trials, if any, and the same amount of time to learn the task.

Factors that can affect the external validity of a study include:

- **Pretest sensitization,** which occurs when the pretest alerts the participants to the hypothesis and/or treatment variable. This sensitization then leads the participants to respond differently on the posttest than they might have if there had been no pretest. For example, let's say that you are given a pretest to assess your attitude toward abortion, and then you are exposed to the independent variable, which is a film on abortion. It wouldn't take you long to put two and two together and think that the experimenter is interested in seeing whether the film could change your initial attitude. When it's time to take the posttest attitude measure, you might respond differently than you would have, had you not linked the pretest to the independent variable. As you can see, in pretest sensitization the pretest interacts with the treatment.
- **Reactive effects,** which refer to the participants' awareness that, since they are taking part in an experiment, their behaviors are being observed and measured. As a result, the participants' behaviors or responses may not be "natural," or true to real-life situations. For example, if you are being observed as you complete a difficult task, you may take more care in avoiding errors than you would in a more natural, real-life situation. In addition, the strictly controlled, artificial setting of the research laboratory may affect the experimenter's ability to generalize the findings. In other words, the external validity of the experiment may be reduced. Reactive effects are covered in detail in the next chapter.
- **Multiple-treatment interaction,** which affects experiments in which each participant experiences every level of the independent variable, or variables. If having participated in one condition changes the way a participant responds in the second condition, then we can say that there is multiple-treatment interaction. For example, if the researcher is studying the effects of alcohol and marijuana on reaction time, then a participant's consumption of alcohol in the first condition may interact with the marijuana in the second condition. If the participant had both alcohol and marijuana, the researcher would not be able to tell whether the participant's reaction time reflected the effects of alcohol, marijuana, or an interaction of the two.

As mentioned earlier, it is the research design that determines whether any of the above-mentioned sources of unwanted variance should be of concern to the experimenter. For example, pretest sensitization is not a problem if the design does not include a pretest. Similarly, if the study is not long term (longitudinal) or does not involve multiple sessions, subject mortality is not a potential source of unwanted variance. In later chapters, you will revisit many of these potential sources when the particular type of design that might be affected is discussed.

CONTROLLING SECONDARY VARIANCE

Eliminating Secondary Variables

Whenever possible, eliminating unwanted variables is an excellent way of dealing with secondary variance. For example, one potential source of unwanted variance arises when the participant is aware of, or even thinks she is aware of, the hypothesis and/or the desired treatment effect. Suppose that a group of psychologists is clinically testing the effects of a new drug on depression. They give the drug to 20 participants who have been diagnosed with depression, while a control group of participants receives no treatment. Following the treatment phase, both groups are assessed for depression and it is found that the treatment group has improved but the control group has not. At this point, can the researchers conclude that the drug is effective in reducing depression? No, they cannot because alternative explanations for the improvement cannot be ruled out.

For example, it is possible that the participants improved simply because they *expected* to improve. In other words, if the participants believed that the medication would improve their mood, the belief and the accompanying expectation might have been what caused the improvement, and not the drug at all. In a similar vein, it is also possible that the control group did not improve simply because they knew that there was no reason for them to improve. Since they did not receive any medication, they had not expected to improve. In addition, if the psychologists themselves were aware of the treatment conditions and the identity of the participants in those conditions, then they, too, had expectations. For example, they may have erroneously observed more depressive symptoms in the control group and fewer symptoms in the treatment condition. The cliché, "We see what we want to see," can become a real source of secondary variance. However, such *expectancy effects* can be eliminated.

Think for a moment: How would you eliminate both subject and experimenter expectancy effects in the preceding example? The best way to eliminate the problem would be to use a **double-blind** placebo design. In such a design, a placebo pill, usually made of sugar, is used in order to control for expectancies. The **single-blind** version means that the participants are not aware of whether they are receiving the actual drug, or the placebo. In the double-blind version, neither the participants nor the administrators of the drug know who receives the drug and who receives the placebo. Therefore, both subject and experimenter expectancies are eliminated.

In general, the best way to eliminate subject expectancies is to keep the participants from being aware of the hypothesis and to make sure they remain ignorant of the treatment effects you are expecting in terms of the experimental conditions. For example, don't say to the participant, "You will see both pleasant and unpleasant words, and I am predicting that you will remember more of the pleasant ones."

Holding Secondary Variables Constant

When the elimination of secondary variables is not possible, then hold these variables constant. Treat all participants in the same way, and test them under the same conditions. For example, if the room where you are conducting your study is unpleasantly warm and you fear that the temperature will adversely affect the participants' ability to remember, test everyone under the same temperature condition. This way you can assume that the temperature affects everyone's score equally, and thus the effect of this secondary variable is held constant.

Random Assignment

Random assignment of participants into conditions is one of the most common ways of dealing with a source of unwanted variance: subject factors or individual differences. Participants are individuals and therefore come with their own unique "history," their own unique characteristics, traits, and abilities, and their own unique ways of seeing and analyzing information. In other words, individual differences will always be a source of secondary variance. Are all sources of individual differences a problem? No, because the particular characteristic you should worry about depends on your topic under investigation. For example, individual differences in visual acuity are a problem if you are studying visual discrimination of colored lights but not if you are studying auditory discrimination between different pitches of sounds. Similarly, individual differences in verbal abilities may be a problem in a memory study but not in a study examining spatial skills.

In general, one of the best ways to handle individual differences is to randomly assign the participants into conditions. Random assignment theoretically *equates* your groups prior to the introduction of the independent variable. With random assignment, each participant has an equal chance of being assigned to any of the conditions. Therefore, it is unlikely that all the really "good" participants will end up in one condition and all the not so "good" participants in the other condition. The probability of equating the groups by random assignment increases as the sample size increases. For example, let's return to the earlier example of violent television and aggressiveness. One major individual difference between the children is their own level of aggression prior to exposure to the independent variable. Some children are simply more aggressive than others, but, by randomly assigning them, it is unlikely that you will end up with all the more aggressive children in the violent condition and with all the less aggressive or not aggressive children in the nonviolent condition. It is more likely that both conditions will have children who are aggressive and children who are not as aggressive.

Typically, researchers want an equal number of participants in each condition, but with true random assignment it is possible to end up with 40 participants in one condition and only 20 participants in the other condition. For example, if the researcher decides to randomly assign participants by flipping a coin, where "heads" go into one condition and "tails" into the other, it is possible to obtain mostly "heads" and very few flips with "tails." If the researcher desires the same number of participants in each group, then random assignment is carried out within limits of equal cell sizes. Cell sizes refer to the number of participants in each condition. Equal cell sizes means that once all the participants have been randomly assigned to one condition, the remaining participants automatically go into the other condition.

As you will see in later chapters, randomization is possible not only with participants, but also with stimulus materials, treatment conditions, and the order in which stimuli are presented.

Systematizing the Secondary Variable

When you systematize the secondary variable, you make it an independent variable. This way, you can systematically observe and measure its effects, not only by itself, but also in conjunction with your original independent variable. For example, let's say that you are interested in studying the effect of a training course on performance on a standardized test. Being aware of the literature on differences between male and female performance on standardized tests (Steele, 1997), you worry that gender might be a potential source of secondary variance. Well, you can always make gender an additional primary variable. As you can see in Table 3.1, you will need four groups of participants.

When the study is over, you can look for the effects of the training, for gender differences, and even for an interaction between gender and training. Interactions are covered in detail in later chapters, but briefly, an interaction would occur if the levels of the independent variable of training affected the genders differently, for example, facilitating performance for males but not for females.

MINIMIZING ERROR VARIANCE

Recall from Chapter 2 that error variance leads to inconsistent fluctuations in the dependent measure, and it may arise from individual differences, observation and measurement errors, or even during data entry and analysis. There will always be some error variance, or "noise," in an experiment, no matter how careful experimenters are with their designs and methodologies, and the best that a careful researcher can do is to take every step to minimize it.

Individual differences, as you saw earlier, can be minimized by randomly assigning the participants into conditions and by increasing the sample size. When the effects of individual differences are consistent, they produce secondary variance. When the effects are inconsistent, they produce error variance.

Measurement errors may occur if the researcher is careless or inconsistent with the observations and/or data gathering. Today, the use of computers in the laboratory to present stimuli and to record responses has greatly reduced this potential source of error variance. For example, many studies use reaction time as the dependent variable, and, when participants are tested on a computer, it is the computer that records the reaction times. This eliminates the worry that the timing may be inaccurate or that the experimenter may have missed a response from the participant due to distraction or inattention. In general, the more automated you make the experiment, the less you need to worry about variability and inaccuracy in stimulus presentation and response measurement.

TABLE 3.1 SYSTEMATIZING A SECONDARY VARIABLE

	Training	No Training
Males	Group I	Group II
Females	Group III	Group IV

Data analysis error may occur if the researcher enters the data incorrectly into the computer for analysis or selects the wrong statistical procedure. The availability of computerized statistical packages such as SPSS, MyStat, and Minitab, to name a few, has greatly reduced the chance for errors in calculations. However, to date no statistical package exists that enters the data and selects the appropriate procedure for the experimenter. As you will see in Chapter 8, different experimental designs and different dependent variable measures require different types of statistical analyses. The inappropriate statistical procedure can lead to erroneous conclusions. In addition, entering hundreds of data points (participant responses) into a computer is a lengthy and tedious procedure. Methodical researchers carefully check their data entry prior to analysis to ensure that no errors were made.

In summary, good research requires good control procedures. Secondary variance may be eliminated or held constant, and error variance may be minimized. However, since we don't live in a perfect world, there will always be some random "noise," or error variance in all research studies.

SAMPLING TECHNIQUES

Earlier in this chapter you learned about external validity, which is the ability to generalize from the experimental sample to the population of interest. One important factor to keep in mind is that this ability to generalize is greatly affected by the sample of participants in the study. If the sample is not representative of the population, then the ability to generalize, or external validity, is compromised. A **representative sample** is one in which the characteristics of the participants reflect the characteristics of the population. For example, if you were interested in the attitude of North American women toward abortion, you would need a sample of North American women representative of the population. A sample of women chosen from a religious organization would be as nonrepresentative and biased as a sample of women chosen from a political feminist organization.

> *One issue of concern is that, given that the majority of research studies are conducted on college students (Sears, 1986), to what extent can we generalize the findings of research studies? Can we assume that these participants are representative of the population? College students may differ from the population in terms of age, level of education, socioeconomic status, cognitive skills, and intelligence, to name a few potential variables. In addition, as our population becomes increasingly diverse, gender, race, and ethnicity are important variables and, as some investigators point out, not taking these variables into account can also result in a non-representative sample. (Denmark, Russo, Frieze, & Sechzer, 1988)*

Just how concerned should the researcher be, then? Once again, the answer is that it depends on the purpose of the research study. If the purpose is to accurately describe population characteristics, then the answer is, very concerned. On the other hand, if the aim of the research is simply to establish a relationship between the independent and dependent variables, or to test hypotheses derived from various theories, then the concern to be able to apply the findings to all people, everywhere, may not be so great. However, keep in mind that, regardless of the purpose of the study, having both male and female participants, as well as participants from diverse racial and ethnic backgrounds, can only improve the external validity of the research project.

Sampling techniques fall into two major categories: **probability sampling,** and **nonprobability sampling.** Depending on the purpose of the research project, an experimenter may opt for either type of sampling technique.

Probability Sampling

When the aim of the study is to describe or make inferences about population characteristics, then the researcher must use *probability sampling techniques.* In probability sampling, we can calculate the probability that a specific member of the population will be selected to participate in the study. Survey research in particular demands that a probability sampling technique be used, since the researcher's aim is to accurately describe population characteristics from the basis of the sample. Opinion polls, surveys about religious or moral beliefs, and questionnaires regarding political and social attitudes, for example, must be administered to a representative sample if the researcher wishes to accurately describe how the population at large thinks or feels about the issues addressed in the survey. Two commonly used probability sampling techniques are **random sampling** and **stratified random sampling.**

In *random sampling* every member of the population has an equal chance of being selected for participation. For example, if the population of a hypothetical city is 2 million people, every member has a probability of 1 in 2 million of being selected. Random sampling is ideal in some instances, for it is the most unbiased way of selecting participants. Random sampling may be attempted by having a computer select people by their social security numbers. Since all documented residents in the United States must have a social security card, it would be considered random selection. Additional ways of selecting randomly would be to rely on a random numbers table, where you would select a number, let's say 124,003, then go to the telephone book, turn to page 124, and select the third name on the page. You would continue this until all participants had been selected.

But random sampling does not always guarantee that the sample you selected is truly representative. For example, suppose that you are interested in studying the population of the hypothetical city mentioned earlier, and you know that the ethnic makeup of the city is 75 percent White, 10 percent African-American, 10 percent Hispanic, and 5 percent Asian. Notice that by using true random sampling, Asians have the least chance of being selected for the study, while Whites have the greatest chance of being selected. Therefore, if you wanted a sample of 200 people, there is a good chance that your sample would contain mostly White participants. So, what is a researcher to do?

Stratified random sampling ensures that the sample matches the population on certain characteristics such as gender, race, ethnicity, religion, or any other factor that is known about the population. Given the example above, knowing the makeup of the city you would randomly draw participants from each ethnic *stratum* until you had a sample that mirrored the population. For example, you would randomly select 150 Whites, 20 African-Americans, 20 Hispanics, and 10 Asians, and you would end up with a sample of 200 people, with each ethnic group represented in proportion to the population.

Nonprobability Sampling

You read in the previous section that, depending on the purpose of the research study, an experimenter may choose either a probability or a *nonprobability sampling technique.*

Although probability sampling may result in the most unbiased and representative sample, a much more common technique is nonprobability sampling. In nonprobability sampling the probability of a given member of the population being selected for participation is not known because the sampling is based on convenience. Therefore, there is always a possibility that the sample is not truly representative of the population, and external validity may be reduced. A large percentage of experimental research in psychology takes place on university or college campuses, where students majoring in psychology are required to participate in experiments. This ready-made *subject pool* is the reason why it was mentioned earlier that the typical participant in a research study tends to be a college student.

If, as noted earlier, the purpose of the research project is to demonstrate a relationship between the independent and dependent variables, the researcher might not be interested in generalizing the findings to all members of the population of interest. Therefore, the expensive and time-consuming process of probability sampling may be undesirable or unattainable. Furthermore, should the experimenter find a relationship between the two variables under investigation in the laboratory, an interesting question would be: Is there a reason to assume that this relationship would not exist in the real world? For example, suppose you find that the participants in your study recalled more concrete words than abstract words. Is there a compelling reason to assume that, just because your participants were college students, the population at large would not recall more concrete words as well? As you read earlier in the chapter, there is evidence that a strong relationship exists between laboratory and real-world findings, which supports the validity of experimental samples in general.

Nonprobability sampling techniques include the **available sample** and the **quota sample.**

The *available sample* is also called a *haphazard sample,* or a *convenience sample.* As Cozby (1997) called it, it is a "take-them-where-you-find-them" approach to gathering participants. This approach may take advantage of subject pools, such as psychology students at colleges and universities who are required to participate for course credit, or it may depend on *volunteers* who are willing to participate in the study. In addition, some experimenters pay individuals for their participation. Keep in mind that college students who are required to participate are still, in a sense, volunteers; the students must give informed consent to participate, and they typically have a series of experiments from which to choose. For example, most colleges and universities have sign-up sheets posted on a board outside the research laboratories. Students can read over the various studies and choose one over the other. Therefore, some students may opt for a social psychology research project, while others may choose experiments in cognitive psychology. In addition, participation is not an absolute requirement in most institutions in that the students are typically given the option to write a research paper instead. However, given that students have a choice of experiments, it is important to keep in mind that certain topics will attract certain participants. As such, the sample may not be as representative as the researcher hopes.

Volunteer characteristics have been studied extensively (Rosenthal & Rosnow, 1975), and several factors have emerged that may make these participants different from the population in general. For example, regardless of where the volunteers are found, they tend to be female, more educated, have a higher socioeconomic status, higher intelligence in general, and a higher need for approval. In addition, they tend to be more curious, unconventional, and less authoritarian than nonvolunteers. In terms of religion, Catholics are the least likely to volunteer, and Jews are the most likely to volunteer.

In general, as our population becomes more diverse, there is hope that college students and volunteers will become more diverse as well. Also, as you read in Chapter 1, an important hallmark of experimentation is the ability to replicate the findings. If the findings can be replicated at other colleges and universities, with other samples, then the concern about limited generalizability is mitigated.

Quota sampling is similar to stratified sampling, except that the members of the subgroups of interest are not randomly selected from the various strata. To return to the example given under stratified sampling, if you wanted to ensure that your sample accurately reflected the proportion of various ethnic groups in the city, you would still need 150 Whites, 20 African-Americans, 20 Hispanics, and 10 Asians. However, these participants are chosen not randomly, but on the basis of their availability and/or willingness to participate. In other words, the participants will be chosen haphazardly rather than randomly, and, as such, this method will have the same limitations as a haphazard or available sample.

SUMMARY

- Internal validity refers to the researcher's ability to state that the relationship she or he predicted between the independent and dependent variables does indeed exist and that this relationship is due to the effects of the independent variable and not to extraneous, uncontrolled variables. If the study is well controlled and alternative explanations for the results can be ruled out, it is said to have internal validity.
- External validity refers to the extent to which the results of an experiment can be generalized to other settings, other people, and other circumstances.
- A major paradox of research is that by strengthening internal validity, external validity may be compromised. Depending on the purpose of the research study, an investigator may be more concerned with internal validity than external validity and vice versa.
- Basic research is typically driven by theory-generated hypotheses, and therefore the researcher may be less concerned with external validity. The aim of applied research, on the other hand, is to solve real-life problems outside of the laboratory. As such, applied researchers need to be concerned with external validity.
- Secondary variables can affect both the internal and external validity of a study, depending on their sources. To control secondary variance these unwanted variables should be eliminated, held constant, or systematized.
- Secondary variables that affect internal validity include nonrandom assignment into conditions, maturation, order or pretest effects, regression toward the mean, history, subject mortality, and differential treatment of participants.
- Secondary variables that affect external validity include pretest sensitization, reactive effects, and multiple-treatment interaction.
- Error variance may arise from individual differences or measuring error. Individual differences can be controlled by randomly assigning the participants into conditions. Measurement error can be minimized by using care in data collection and data analysis procedures.
- Random assignment theoretically equates the groups prior to introducing the independent variable. In random assignment, every participant has an equal chance of being assigned to any of the conditions.
- The external validity of a study depends on the sample. If the sample is biased, and/or nonrepresentative of the population, the external validity of the study is compromised. In a representative sample, the characteristics of the sample accurately reflect population characteristics.

- Sampling techniques fall into two major categories: probability or nonprobability sampling. In probability sampling the probability of a given member of the population being selected can be determined, whereas in nonprobability sampling the probability of being selected is not known.
- Random and stratified sampling belong to the category of probability sampling, whereas available samples and quota samples represent nonprobability sampling.
- Attempts should be made to ensure that the sample represents the diversity of the population in terms of gender, race, and ethnicity.

KEY CONCEPTS

Applied Research
Available Sample
Basic Research
Differential Treatment
Double-Blind
External Validity
History
Individual Differences
Internal Validity
Maturation
Multiple-Treatment Interaction
Nonrandom Assignment
Nonprobability Sampling
Order or Pretest Effects
Pretest Sensitization
Probability Sampling
Quota Sample
Random Assignment
Random Sampling
Reactive Effects
Regression Toward the Mean
Representative Sample
Single-Blind
Stratified Random Sampling
Subject Mortality

QUESTIONS

Short answers can be found in Appendix A.

1. Identify whether the following factors affect the internal or external validity of the study.
   a. nonrepresentative sample of participants
   b. pretest sensitization
   c. practice effects
   d. subject mortality
   e. reactive effects
   f. maturation

2. A teacher gave her fifth grade class a reading test in September, introduced a new teaching method for reading, and gave the class another reading test in June. Since the reading scores were higher in June, she concluded that the new teaching method was effective. Which of the following is a possible explanation for the improved scores? Check as many as apply.
   a. effective teaching method
   b. maturation
   c. history
   d. practice effects
   e. pretest sensitization

3. Explain basic and applied research. Give two examples for each.

4. Explain how you would control for expectancy effects (both subject and experimenter) in a study on the effect of alcohol on reaction time.

5. What are reactive effects? Define and give an example.

6. In a study on verbal learning, participants were randomly assigned to a recall or recognition condition. Both groups saw the same words, which were presented by a computer at the rate of one word per every two seconds. Each group was given five minutes to recall or recognize as many words as possible, depending on the experimental condition. In the recall condition, the participants entered their responses into a computer, and in the recognition condition, the participants were asked to highlight as many words on the computer screen as they could recognize as having been seen before.
   a. Explain how the experimenter controlled secondary variance.
   b. Explain how error variance was minimized.

7. Identify the type of sampling method used in the following:
   a. Asking people to fill out a survey as they are coming out of a department store
   b. Asking for volunteers by posting fliers around a college campus

c. Contacting all the people whose phone numbers were selected by a computer-generated program

d. Identifying the proportion of Catholics, Protestants, Jews, and Muslims in a city, and then randomly selecting participants from each category so that the sample represents the religious makeup of the city

e. Asking for volunteers from the African-American, Asian, and Hispanic communities so that the sample will represent the proportion of each group in the population

f. Asking every fifth person leaving a high school cafeteria to fill out a questionnaire.