

CHAPTER 7

Experimental Design

Independent Groups Designs

OBJECTIVES

After reading this chapter, students should be able to

- List the steps in conducting an experiment
- Describe the difference between a controlled experiment and a field experiment
- Describe the advantages and disadvantages of controlled experiments and field experiments
- Describe the nature of an independent groups design
- Describe the difference between a completely randomized groups design and a randomized factorial design
- Interpret the statistical outcome of each design, including main effects and interactions in factorial designs
- Design an experiment to assess a given hypothesis

A few years ago, one of the authors of your book read an article in a local newspaper with the headline “Don’t take engineering, young ladies, if you hope to marry!” The writer described some data obtained from a major university that indicated that female engineering graduates were less likely to marry than female graduates from other faculties. The reader was left with the impression that women were somehow dooming themselves to singlehood by enrolling in the faculty of engineering. We continue to be amazed to read

134 METHODS IN PSYCHOLOGICAL RESEARCH

about the many causal interpretations that are made about data that simply do not permit such interpretations.

We can confidently conclude a cause-and-effect relationship between variables *if and only if* the appropriate study has been conducted. To conclude that being educated as an engineer causes a decrease in marriageability *could* be made if there was a significant difference between the postgraduate marriageability of two groups of women who were *initially* equivalent in marriageability and were then randomly assigned to an engineering group and a nonengineering group. And, of course, it would have to be established that during the course of their education, variables extraneous to the education experience did not differentially affect the groups. Do you think this was the case? We don't either.

WHY WE DO EXPERIMENTS

The **experiment** is the cornerstone of scientific research. The goal when we conduct experiments is to show that an **independent variable** (IV) causes a change in the **dependent variable** (DV). In psychological research, the dependent variable is usually some measure of behavior. Perhaps we would like to know whether technologically enhanced courses compared to traditional course delivery techniques (IV) improve student performance in courses (DV). Or perhaps we are interested in comparing psychotherapy with medical therapy (IV) in the treatment of anorexia symptoms (DV). Conducting an experiment is considered to be the best way to provide us with the answers to these kinds of questions. Not all problems lend themselves to experimental study, but those that do are best approached this way. We will begin this chapter by discussing the basics of the experimental approach.

To be a true experiment, as opposed to a quasi-experiment or a nonexperiment, the independent variable must be *under the control of the researcher*. In other words, the researcher must assign participants to the levels of the independent variable. Consider the example about psychotherapy versus medical therapy in the treatment of anorexia. If our goal is to compare psychotherapy with medical therapy for the treatment of anorexia, we could study people who had been treated with either therapy in the past. We could determine how well they have progressed with each type of treatment. This approach, however, is not an experimental approach because *we* did not assign the patients to the type of treatment (IV). Rather, we compared those who *themselves* had chosen either psychotherapy or medical therapy. This type of study is called a *quasi-experimental design*, and such designs are treated in Chapter 10. For us to conduct a true experiment to study this problem, we, the researchers, must be able to assign the participants to each condition or group in the experiment. We must be the ones to decide who gets psychotherapy and who receives medical therapy. This is crucial because we, the researchers, can then take steps to ensure that there are no systematic differences between the groups before the experiment begins. As a result, we are able to conclude that if we find a difference between the groups at the end of the experiment, that difference is due to the way the people were treated. Imagine that only the most severely affected individuals got medical treatment,

Conceptual Exercise 7A

For each of the following hypotheses, decide if a true or a quasi-experiment is indicated.

1. Young offenders have poorer impulse control than nonoffending youth.
2. Children who view a film promoting helping behavior show more altruistic behaviors than children who view a neutral film.
3. Pigeons on an intermittent schedule of reinforcement emit pecking behavior that is more resistant to extinction than pigeons on a continuous schedule.
4. Women rate pornography as less interesting than men do.

and those with only mild symptoms chose psychotherapy. In this case, the groups were different at the outset of the study; it would be no surprise to find that they are still different at the end of the study.

As we have said, the true experiment is the foundation of scientific research. Although not all research problems can be studied experimentally, when they can, such an approach is preferable because causal statements about the relationship between variables can be made.

Steps in Conducting an Experiment

Step 1. Formulate a Hypothesis

You will recall, from our discussion in Chapter 4, that a **hypothesis** is a statement about the expected relationships between variables. For example, perhaps we are interested in whether practice in mirror drawing with one hand might transfer to the other hand. In mirror drawing, the participant, while looking in a mirror, attempts to duplicate a figure, number, letter, and so on. Our hypothesis might be as follows:

Positive transfer to the nonpreferred hand will occur with training in mirror drawing with the preferred hand.

As you can see, this hypothesis is a statement about theoretical concepts (i.e., positive transfer and training). The next step is to decide how to measure these concepts. In other words, we have to operationalize, or make measurable, the variables.

Step 2. Select Appropriate Independent and Dependent Variables

In the example given above, the IV is amount of practice (practice vs. no practice) with the preferred hand. There are various ways we could measure (or operationalize) positive

136 METHODS IN PSYCHOLOGICAL RESEARCH

transfer. One way would be to count the number of errors on three trials before and after practice. Another way would be to measure the time to complete a trial. Of course, there are questions to answer. What is an error? How much practice?

Deciding which dependent measure is the most valid and reliable way to measure the behavior of interest is a matter of experience and familiarity with the available research. For more detail on measurement, see Chapter 5.

Once we have a testable hypothesis about the expected relationship between our IV and DV, we need to consider what other variables might be involved and find ways to control them.

Step 3. Limit Alternative Explanations for Variation

Remember, the goal of experimentation is to determine cause and effect. Does manipulation of the IV have a causal effect on the DV? Clearly, there is more than one variable influencing behavior at any one time. Let's look at an illustrative, if somewhat obvious, example. An angler wants to know which type of bait is best for trout. *Type of bait* then will be the IV. The *number of trout* caught in 4 hours will be the DV. Our angler goes to a lake and fishes from 8:00 a.m. until 12 noon on Sunday using flies. The following Sunday, he goes to another lake and fishes with worms from 2:00 p.m. until 6:00 p.m. He catches one trout at the first lake and six at the second lake. If we can demonstrate statistically that more fish were caught at the second lake, then we need to ask ourselves, have we demonstrated a *causal relationship* between the IV (type of bait) and the DV? Are there alternative explanations for the difference in number of fish caught that have nothing to do with the IV, type of bait, and, therefore, are confounded with it? We think so. The second lake might have more trout. The weather on the first Sunday might have been less conducive to fishing. Perhaps fish are more likely to bite in the afternoon than they are in the morning. We are sure you can think of other possible confounds. In an experiment (or any research), we want to control as many of these other variables as possible. Once we have done our best to think about, and limit, alternative explanations for our hoped-for effect, we can go to Step 4.

Step 4. Manipulate the IVs and Measure the DVs

In other words, carry out the experiment. We now have our data. What next?

Step 5. Analyze the Variation in the DVs

In the ideal experiment, all the variation in the dependent variable between groups receiving the independent variable (i.e., treatment groups) and groups not receiving the independent variable (i.e., control groups) should be caused by the independent variable. The objective of the exercise is to decrease variation among groups that is not a result of the manipulation of the IV (i.e., error variation). Techniques for reducing error variability are discussed more in Chapter 4.

We now must choose the appropriate statistical technique to analyze the variance in the DV. Which procedure we select depends on the kind of data we have and the questions we wish to answer. In Chapter 2, we discuss common analyses used by researchers in psychology today, and in Chapter 13, we discuss common statistical procedures students might use for their projects. Once we have selected and conducted the appropriate statistical analysis for our data, we can complete the final step.

Step 6. Draw Inferences About Relationship Between IVs and DVs

We use inferential statistical procedures to make statements about populations based on our sample findings. Conducting a true experiment allows us to make causal statements about the relationship between the IV and the DV. We can be confident in saying that the manipulated IV caused the changes in the behavior that we measured if we have carried out our experiment carefully by controlling other variables that could provide an alternative explanation of the changes in the DV, leaving our IV as the only likely causal variable.

Where We Do Experiments

Experiments can be conducted in the laboratory (**controlled experiment**) or in a natural setting (**field experiment**). In both controlled and field experiments, the IV is directly manipulated by the researcher. However, in a natural setting, which is where field experiments are conducted, it is more difficult to control all the secondary variables that might affect the results. As a consequence, many researchers prefer to do their work under laboratory conditions if they can. As you can see, there is a downside and an upside to conducting your experiment in both settings.

Controlled Experiments in the Laboratory

There are three broad advantages to controlled experimentation. First, our ability to control the independent variable is superior under laboratory conditions, improving internal validity. Second, we have superior control over secondary or extraneous sources of variation in the laboratory. For example, we can control noise and temperature. Last, we can more precisely measure our dependent variable under laboratory conditions. This kind of control over the IV, the DV, and secondary variables improves the internal validity of the study—that is, the probability that any changes in the DV are indeed a result of the manipulation of the IV.

Although the experiment is considered by researchers to be the best way to determine cause and effect, there are disadvantages to controlled experimentation. Some phenomena cannot be studied in a laboratory at all. The effect of September 11 on frequency of travel by Americans, for example, is a research topic that does not easily lend itself to laboratory study. Other research topics present ethical problems. For example, it would be unethical to conduct a laboratory study of the effects of sensory deprivation on human infants. There are also practical disadvantages to laboratory investigation. It can be costly and time-consuming.

But perhaps the most serious disadvantage to controlled experiments is that the outcomes may not be applicable to the real world. Behavior that occurs in a laboratory may be idiosyncratic to that environment and may not occur in a natural setting. If we decide to do our experiment in a natural setting, we are conducting a field experiment.

Experiments in the Field

When controlled experimentation is not possible or ethical, a field experiment may be the best choice. A field experiment is conducted in a natural setting (the field) where the experimenter directly manipulates the independent variable. Imagine that we are hired to determine if domestic violence intervention training for police officers reduces domestic assault. We might randomly assign a group of police officers to receive special training and a control group of

138 METHODS IN PSYCHOLOGICAL RESEARCH

officers who receive the standard training offered by the police department. We could then take various measures such as number of domestic assaults in areas served by the two groups or satisfaction of families served by those officers. Suppose we find that there was a difference. Fewer assaults and more family satisfaction occurred in the areas served by the specially trained officers. Although we would like to think this outcome was caused by our independent variable (training), we would have to be quite cautious in our inference, in part because in this field experiment, as in all field experiments, we have less control over extraneous, or secondary, sources of variation. The police officers would quickly learn that some of them were in a special training group and some were not. Perhaps the specially trained officers tried harder because they knew they were in the special training group, and it was their extra effort rather than the special training that caused the difference. These “clues” that lead participants to guess about the nature of the study and that may change their behavior are called *demand characteristics*. Researchers must try to anticipate what the demand characteristics associated with their study might be and try to limit their influence on the outcome.

Back to our example. Perhaps there was a factory shutdown in the neighborhoods served by the officers who did not receive special training. When people are out of work, domestic problems probably escalate. It is possible that the greater number of domestic assaults in those neighborhoods might be a result of this factor rather than the training factor. In a laboratory situation, it is often easier to control these kinds of variables by using single- and double-blind procedures.

The outcome we may find in a controlled laboratory setting may not generalize to the external (i.e., natural) world. When you bring a phenomenon into the laboratory, you may be interfering with how it operates naturally. Findings that do not hold true in a natural setting are not **externally valid** and may be of little interest. On the other hand, as you saw in our domestic violence example, field experiments, which may be more externally valid, have a major disadvantage in that they may lack **internal validity** because it is much more difficult to control the IV, the DV, and secondary sources of variation.

Choosing the best setting in which to conduct an experiment requires considerable thought. Some things to think about include pragmatic considerations, including cost, control over variables, and validity considerations. We discuss validity in much more detail in Chapter 5.

As we often say, it is a matter of balance.

Conceptual Exercise 7B

A researcher is interested in the uninhibited behavior she has often observed at rock concerts. She wonders what kinds of variables influence this behavior (loudness of band, proximity of audience to band, etc.). Should she consider a controlled or a field experiment? Why?

We have discussed the steps that researchers follow and the things they must think about as they consider how best to answer the questions they have. The rest of this chapter deals

with basic experimental designs where different participants serve in each level(s) of an independent variable(s)—independent groups or between-participant designs.

HOW WE DO EXPERIMENTS: INDEPENDENT GROUPS DESIGNS

We cannot stress enough the importance of the assumption of initial equivalence of groups in experimental design. If we cannot assume that our groups were equivalent prior to treatment (i.e., before manipulation of our IV), then we have no basis at all for any causal inferences about differences after treatment. So, how can we assume that our groups are equivalent at the start?

One common way is to assign participants randomly and independently to conditions—an independent groups or between-participants design. In an **independent groups** or **between-participants design**, participants are randomly and independently assigned to each level of the independent variable. Because the participants were independently assigned to groups or levels of the IV, we can feel confident that the groups were initially equivalent. If we then treat each group differently (i.e., the treatment variable) and find there is a significant difference in the outcome measure (i.e., the dependent variable), then we can confidently infer that the outcome, differences in DV, was causally related to our manipulation (i.e., levels of the IV).

With independent groups or between-participants designs, each score is independent of every other score, hence the name—independent groups designs. Because different participants are assigned to the different levels of the independent variable, their scores are assumed to be independent of each other. The simplest independent groups design is one where the researcher is interested in one independent variable with two or more levels—a completely randomized groups design.

Completely Randomized Groups Designs: One IV

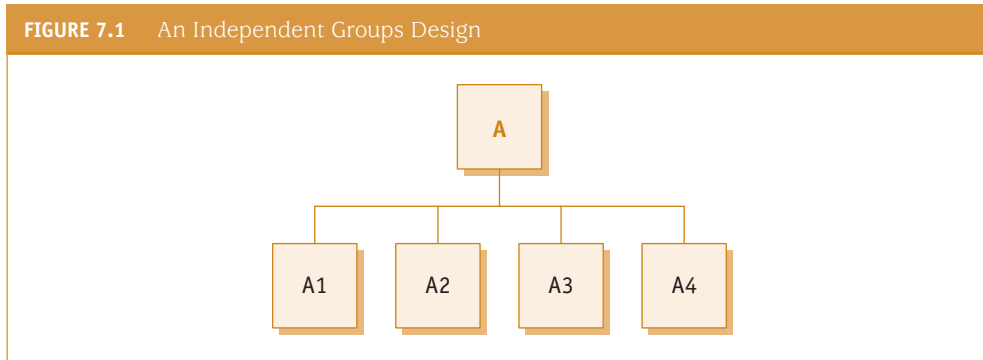
In a **completely randomized design**, research participants are randomly assigned to different levels of *one* independent variable. Figure 7.1 illustrates how we might diagram this type of design.

As you can see, there are four levels of the A independent variable, and n participants would be randomly assigned to each group.

The simplest completely randomized groups design would be a two-group design where participants are randomly selected and independently assigned to either an experimental group or a control group (i.e., two levels of one independent variable). Such designs allow us to answer one question: Did the manipulation of the independent variable affect the

NOTE: The American Psychological Association recommends that the word *participants* be used when referring to human participants in research. Although the word *subjects* has been used, and typically still is used in many books and published research articles, to refer to various types of research designs such as between-subjects and within-subjects designs, we have followed the recommendation of the APA in this regard. We use the word *subjects* only when we are referring to animals, not humans.

FIGURE 7.1 An Independent Groups Design



dependent variable? Perhaps in more typical language, we might ask, did the treatment of the various groups cause a difference in some response measure, the DV? Let's examine a recently published research paper to see how a between-participants design was used in an effort to answer a specific question.

Randomized Groups Design: One IV With Two Levels

The Research Problem. We have all heard the phrase “conflict of interest.” We expect a person who is making a judgment about the value of something to be fair, impartial, and objective. A conflict arises when a person's judgment about a primary interest (such as which drug works better) is influenced by a secondary interest (such as ownership of shares in a drug company).

Chaudry, Schroter, Smith, and Morris (2002) were interested in how readers of a medical research periodical judge the scientific value of that research. Specifically, they wanted to know if readers would judge a research study as less interesting, important, and valid if they believed that the researchers who conducted the study had a conflict of interest.

Hypotheses. Chaudry et al. (2002) hypothesized that readers who believed that a research study had been conducted by researchers with competing interests would judge the study as less interesting, important, relevant, valid, and believable than readers who had no such belief.

Selection of Participants and Assignment to Conditions. Chaudry et al. (2002) randomly selected 300 readers of the *British Medical Journal*. All readers were sent a short report about the substantial impact of pain on the lives of sufferers of herpes zoster (cold sores). For half the readers, the authors of this paper were declared to be employees of, and held stock options in, a fictitious pharmaceutical company. For the rest of the readers, the authors of the paper were claimed to be employees of a care center with no competing interest.

The IV and DVs. The independent variable was the information about the authors, whether they had competing interest (held company stock) versus no competing interest. Readers rated the study described in the paper on several variables, including interest, importance,

relevance, validity, and believability on a 5-point scale. These ratings were the dependent variables.

The Design. Because the readers were randomly assigned to receive each version of the research paper and could only be assigned to one condition, this is an independent groups design. Table 7.1 illustrates the design.

The Statistical Analyses. The researchers chose to compare the means of their groups with independent *t* tests.

The Results. Some of the descriptive statistics reported by Chaudry et al. (2002) are presented in Table 7.2.

Because Chaudry et al. (2002) had two independent groups, they chose independent *t* tests as their inferential analyses. The outcome of these tests indicated that readers in the competing interest group rated the research paper as significantly less interesting, relevant, valid, and believable (all *ps* < .001) and less important (*p* < .02) than did the readers in the no competing interest groups.

TABLE 7.1 Independent Groups Design

Group	
Competing interest	No competing interest

Source: Adapted from Chaudry et al. (2002).

TABLE 7.2 Descriptive Statistics: Reader Perceptions of Research Study

Variable	Group	
	Mean Ratings (Standard Deviations)	
	Competing Interest	No Competing Interest
Interest	2.59 (0.87)	2.99 (0.91)
Importance	2.59 (0.90)	2.94 (0.96)
Relevance	2.70 (1.10)	3.15 (1.10)
Validity	2.53 (0.99)	3.04 (1.00)
Bievability	2.73 (1.10)	3.33 (0.92)

Source: Chaudry et al. (2002).

Note: Low scores indicate low interest, importance, and so on.

142 METHODS IN PSYCHOLOGICAL RESEARCH

The Conclusions. The authors concluded that readers judge scientific research findings as less credible when they perceive that the people conducting the research have a conflict of interest.

An independent groups design with two groups is a basic design where the researcher is interested in comparing an experimental group, which received some sort of treatment, with a control group, which did not. Often we are interested in more levels of the independent variable. Rather than comparing a group of participants who received drug treatment with participants who received a placebo, we may want to compare groups of participants who received various dosages of a drug to determine which dosage is more effective in treating some disorder. In this case, we would have an independent groups design with *several levels of one independent variable*. Here is an example from the literature.

Randomized Groups Design: One IV With More Than Two Levels

The Research Problem. Teachers are not likely to use classroom interventions that they think are unacceptable. Jones and Lungaro (2000) noted that, although many studies have revealed that teachers find reinforcement-based intervention to be more acceptable than punishment-based interventions, little research had been done to examine the nature of the reinforcement strategy. Specifically, they wondered if reinforcement strategies that were clearly connected to the assessment of the problem would be seen as more acceptable than strategies where the reinforcer was not clearly connected to the problem (i.e., more arbitrary).

Consider a child who is acting out in class and receives attention from his or her peers for this behavior. A differential reinforcement procedure could be used where acting-out behavior is ignored by the teacher (rather than the peer group), and desirable behavior is rewarded by points that could be traded in for rewards. In this case, the reinforcers (i.e., teacher attention and points) are not linked to the original problem situation where the behavior seemed to be reinforced by *peer attention*, not *teacher attention*. On the other hand, a differential reinforcement procedure could be used where acting-out behavior is ignored by peers and desirable behavior is rewarded with attention from peers. Given that the acting-out behavior had probably been reinforced by attention from classmates originally, this procedure then links the reinforcer (i.e., peer attention) to the problem situation. This was the basic procedure that Jones and Lungaro (2000) used. Their interest was in the reinforcer itself. If peer attention was reinforcing acting-out behavior, then withdrawing peer attention rather than some other event such as teacher attention might be a more acceptable strategy. And likewise, teachers might feel that rewarding desirable behavior with peer attention rather than teacher attention is a better strategy. This was the problem that Jones and Lungaro decided to investigate.

Hypotheses. Jones and Lungaro (2000) expected that teachers would rate treatments linked to an assessment of the behavior problem to be more acceptable than treatments not linked to such an assessment. They were also interested in the likelihood that teachers would use such strategies. Although they did not specifically say so, we suspect they expected teachers to say they were more likely to use strategies that they perceived as being more acceptable.

Selection of Participants and Assignment to Conditions. One hundred and eleven public school teachers volunteered to participate in the experiment. The teachers were randomly assigned to one of three conditions with 34, 38, and 39 teachers in the groups. All of the teachers read vignettes about a child, Jesse, who was disruptive during class by getting out of his seat. They read that the school psychologist observed the behavior and noted that Jesse was out of his seat about once every 10 minutes and that 90% of the time when he was out of his seat, classmates talked to, laughed at, or teased Jesse. The vignette went on to describe a differential reinforcement strategy recommended by the psychologist in one of three ways.

Teachers in the peer attention group (PA) read that the strategy recommended was that classmates were instructed to ignore Jesse when he was out of his seat. In addition, they were told that Jesse would receive a point for every 10-minute interval that he remained in his seat. At the end of the class, Jesse could have 1 minute of free time with a classmate for every point he had earned during class. This strategy then was linked to the problem situation; peer attention, which seemed to be reinforcing the out-of-seat behavior, was used to reinforce in-seat behavior.

Teachers in the teacher attention group (TA) read that the strategy recommended was that the teacher would ignore Jesse when he was out of his seat. Points were given for in-seat behavior as in the PA group, but those points could be traded in for free time with the teacher, not a classmate, at the end of the class period. This strategy then was linked to the attention part of the problem but used teacher rather than peer attention.

Teachers in the tangible reinforcer group (TR) read that Jesse could earn points for remaining in his seat and that those points could be traded in for desirable items in a grab bag. The reinforcer in this condition, then, was not linked to the problem. Rather, the reinforcer was arbitrary.

The IV and DVs. The independent variable was type of reinforcement strategy. Jones and Lungaro (2000) compared three strategies: peer attention, teacher attention, and tangible reinforcer.

Jones and Lungaro (2000) measured two dependent variables: acceptability of strategy and likelihood of use.

The Design. The design, an independent groups design with three levels of the IV, is illustrated in Table 7.3.

Group		
PA	TA	TR

Source: Jones and Lungaro (2000).

144 METHODS IN PSYCHOLOGICAL RESEARCH

The Statistical Analysis. Because the design was an independent groups design with three levels of one independent variable, the researchers analyzed each DV with a separate one-way analysis of variance (ANOVA).

The Results. The descriptive statistics from Jones and Lungaro's (2000) experiment are presented in Table 7.4. Jones and Lungaro did not provide descriptive statistics for the likelihood of use dependent measure.

Because Jones and Lungaro (2000) had three levels of one independent variable, they used a one-way ANOVA to analyze each dependent measure and a Tukey HSD test to compare pairs of means. You will recall, we hope, from your statistics course that a significant F from an ANOVA tells us only that at least two means are significantly different. The Tukey test tells us which pairs of means are different. The outcome of the ANOVA that Jones and Lungaro conducted was significant on the acceptability of the strategy dependent measure, $F(2, 108) = 6.66, p < .01$. Jones and Lungaro reported that the acceptability scores for the peer attention group were significantly higher than those in the teacher attention and tangible reinforcer conditions (inferential statistics were not reported for the comparisons).

The outcome of the second ANOVA, conducted on the likelihood of use measure, was also significant, $F(2, 108) = 4.20, p < .01$, and again, the Tukey comparisons indicated that teachers in the peer attention group said they were more likely to use the procedure than teachers in the other two groups (they reported the finding to be significant but did not report the Tukey statistics).

The Conclusions. The authors concluded that teachers report that a behavioral intervention strategy is more acceptable and that they are more likely to use that strategy if it is linked to an assessment of the problem behavior situation than if it is not linked. Although the strategy used in the teacher attention condition seems more linked to the behavior problem than the strategy used in the tangible reinforcement condition, there were no significant differences in acceptability or likelihood of use scores between these two groups.

To this point, we have talked about independent groups designs where participants have been assigned to two or more levels of one independent variable. But what if we are interested in a second or a third or more independent variables? As you will see in the next sections, there are important benefits to including more than one IV in a single experiment.

TABLE 7.4 Descriptive Statistics: Acceptability of Intervention Strategies

Variable	Group		
	Mean Score Out of 90 (Standard Deviation)		
	PA	TA	TR
Acceptability	61.1 (11.9)	48.3 (17.5)	49.3 (18.5)

Source: Jones and Lungaro (2000).

Note: Higher scores indicate greater acceptability.

Randomized Factorial Groups Designs: More Than One IV

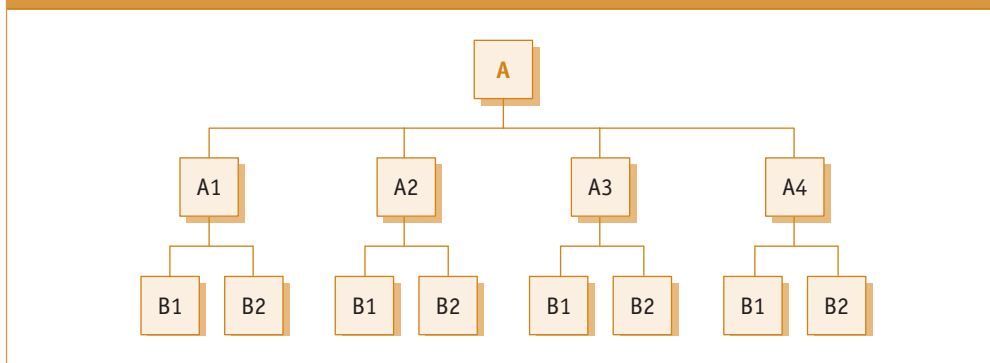
How were your grades in high school? Do you think they accurately reflected your level of achievement? Perhaps your classmates, teachers, or textbooks influenced your grades. Human behavior is complex and influenced by all sorts of variables. Experiments designed to assess the effects of more than one independent variable on performance are probably more like the real world and more likely to be externally valid (i.e., generalize to real-world settings and situations).

In a **randomized factorial design**, participants are randomly assigned to each level of *more than one* independent variable (or factor). These designs allow us to assess the effects of more than one independent variable and to assess the interaction between independent variables. A *simple randomized groups design* provides the answer to *one* question: Did the independent variable affect the dependent variable? *Factorial designs* allow us to find answers to *several* questions. What effect, if any, did each independent variable have on the dependent variable, and how did the combination of levels of the independent variables affect the dependent variable? The statistical analysis of a factorial design, then, allows us to assess the effects of each independent variable on the dependent variable (called the **main effects**), and it will indicate the interactions between those independent variables (called the **interaction effect**). Interaction effects are very important; indeed, if significant interactions between independent variables are present, the interpretation of the effects of those independent variables becomes more complicated. Figure 7.2 illustrates how we might diagram a randomized factorial design.

As you can see, this design has two IVs: A has four levels, and B has two levels. Participants would be independently assigned to one of the eight groups.

Before we look at some real randomized groups experiments, let's examine a hypothetical experiment. Many people, including us, find that reading under incandescent light is easier on the eyes than reading under fluorescent light. But is text easier to read (i.e., more legible) under incandescent light? Word processors offer many different fonts. We find certain fonts to be easier to read than others. Perhaps some fonts are easier to read under incandescent light and others under fluorescent light. We could design an experiment to determine if our anecdotal experience with light and fonts is supported empirically. We

FIGURE 7.2 Randomized Factorial Design



146 METHODS IN PSYCHOLOGICAL RESEARCH

might randomly assign readers to incandescent light conditions and others to fluorescent light conditions, and we might have them read text typed in three different fonts under each type of light condition. With two levels of our first independent variable (light) and three levels of our second independent variable (font), we would have six different conditions in our experiment. If we randomly assigned different readers to each condition, we would have a 2×3 randomized factorial design (i.e., two levels of one IV and three levels of a second IV) that might look like the following (see Table 7.5).

As you can see, readers in the IT condition would read text in the Times font under incandescent light, and readers in the FG condition would read the same text in the Geneva font under fluorescent light. Following the reading part of our study, perhaps we have asked our participants to rate the text for readability. In a randomized factorial design, we have three questions. In our hypothetical experiment, the first question to ask is, “Did the first independent variable (i.e., light) affect the dependent variable (readability)?” In other words, forgetting about the different fonts used for the moment, did light condition affect readability overall? We could examine the mean readability ratings calculated over all font conditions. We would simply determine the mean readability rating for *all* the participants who read the text under incandescent light regardless of the font used and compare those ratings with the mean readability ratings of *all* the participants who read the text under fluorescent light. Let’s plot these ratings in a graph. Figure 7.3 shows the main effect of light for our hypothetical experiment. Higher ratings indicate the text was more readable.

It seems that text is more readable under incandescent than under fluorescent light. In a real experiment, we could analyze these data with a two-way ANOVA to determine if the difference was statistically significant.

Our second question is, “Does our second independent variable (i.e., font) make a difference, overall?” Now, we will look at the readability scores for each level of font calculated over both light conditions. In other words, we will find the mean score for *all* the readers who read the Times font, *all* those who read the Courier font, and *all* those who read the Geneva font, regardless of light condition. Figure 7.4 is a graph of the main effect of font.

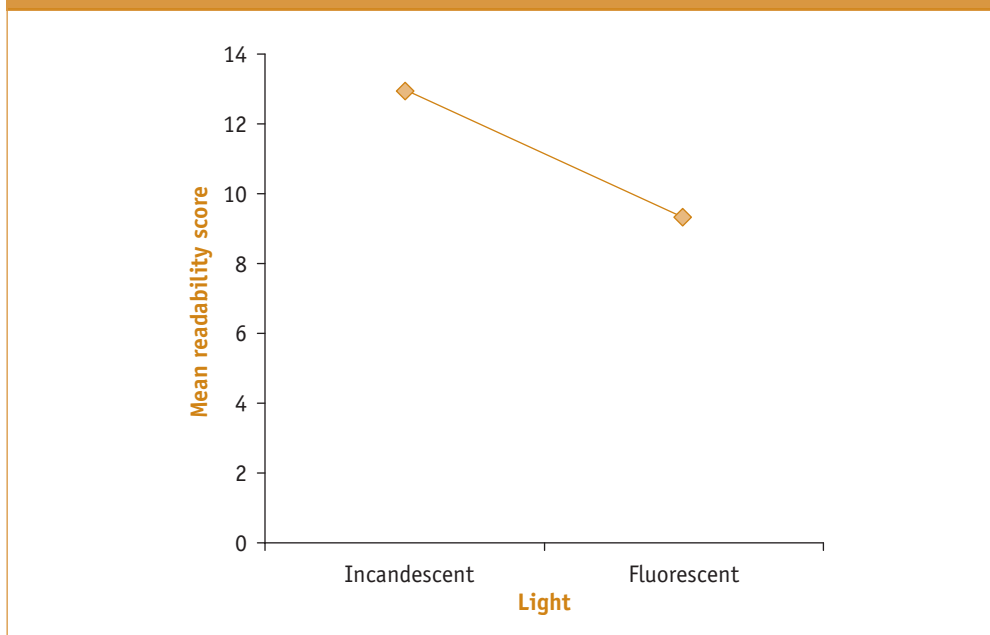
Our graph indicates that the Times font is easiest to read and the Courier font is the hardest to read. Our two-way ANOVA would tell us if the groups differ significantly.

In a randomized factorial design, the main effects must be interpreted in light of any interaction effects. When we examine a main effect, we are looking at means for each level of that IV calculated over all the levels of the other IVs, and so we cannot see any differences

TABLE 7.5 Hypothetical Randomized Factorial Design Comparing Different Kinds of Light and Font

Font (Second IV)	Light (First IV)	
	Incandescent	Fluorescent
Times	IT	FT
Courier	IC	FC
Geneva	IG	FG

FIGURE 7.3 Main Effect of Light



that might exist at the different levels of the other IVs. The following is our third question: “Is the dependent variable affected by the combination of levels of each IV?”

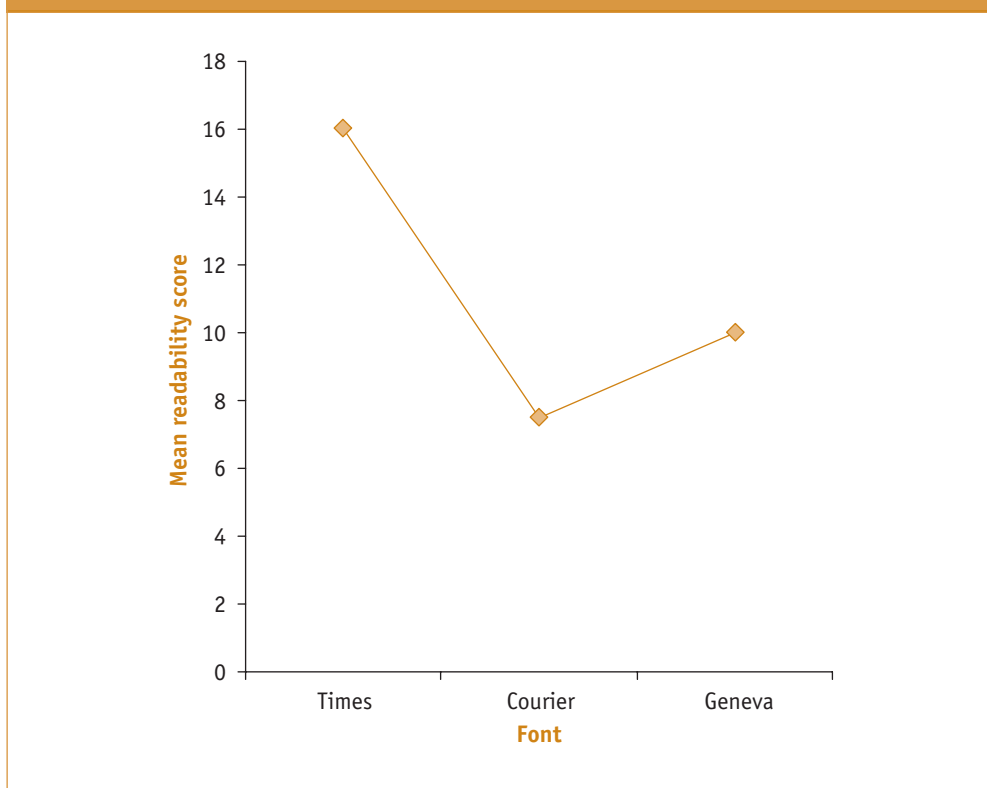
Let’s now graph the mean readability scores for *each* level of each IV to see if there is an interaction effect (see Figure 7.5).

You can see that things are not quite as simple as they seemed when we looked at the graphs of the two main effects. If the two independent variables did not interact, we would see two more or less parallel lines. These lines are not parallel, so it seems that we have an interaction going on here. Of course, the statistical analysis would tell us if the interaction is statistically significant, but let’s examine our graph. Readability of the Times and Geneva fonts seems to be little affected by light conditions. Readability of text in these two fonts is not much affected by the light conditions, but this is not the case for the Courier font. It seems to be much harder to read the Courier font under fluorescent light than under incandescent light. The interaction effect tells us that the dependent variable (readability) is affected differently under the different combinations of the independent variables. This is the great advantage of a factorial design. We can determine how different combinations of levels of two (or more) independent variables affect the dependent variable. We could not determine this effect with a simple independent groups design.

Are you ready for a real experiment?

The Research Problem. Benzodiazepines are central nervous system depressants that tend to reduce aggressive behavior in animals, but some researchers have found the opposite

FIGURE 7.4 Main Effect of Font

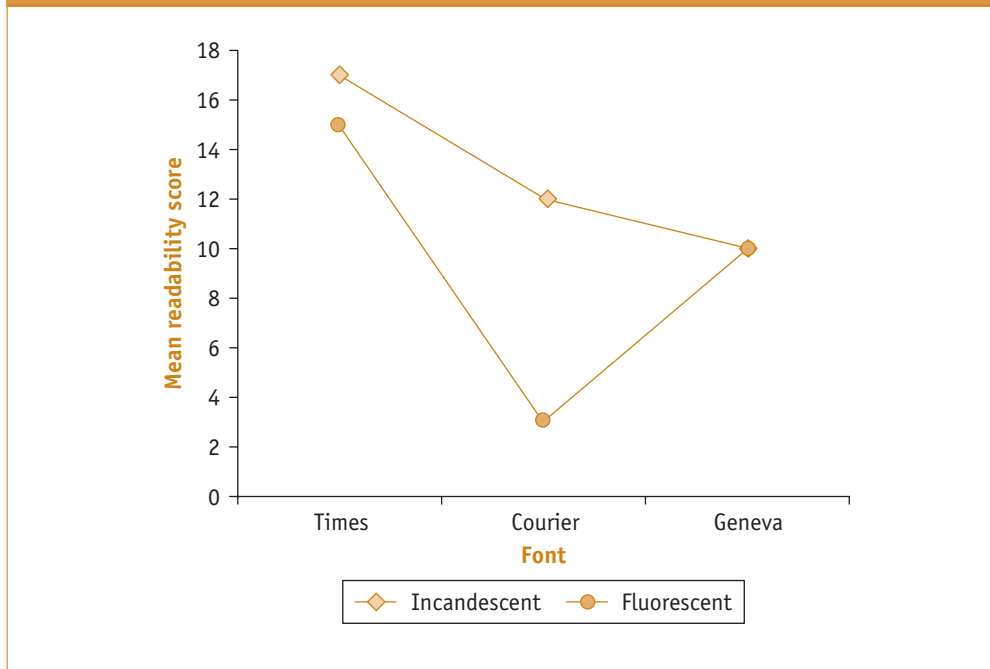


effect, increased aggressiveness, when sustained low doses are used (e.g., Fox, Webster, & Guerriero, 1972, as cited in Renzi, 1982). Renzi (1982) decided to examine the effects of single versus repeated doses of a depressant drug on aggressive behavior of mice that had been induced, by a low-voltage shock, to fight with each other. Renzi was also interested in whether size of dose would make a difference in aggressive behavior and whether size of dose would interact with the number of doses variable.

The Hypotheses. Renzi (1982) did not specifically state his hypotheses, so we will simply specify the three questions his research was intended to answer.

1. Is aggressive behavior affected by number of doses of drug? (Main effect of number)
2. Is aggressive behavior affected by dosage (amount of drug given)? (Main effect of dose)
3. Is aggressive behavior affected by different combinations of dosage and number of doses? (Interaction effect)

FIGURE 7.5 Interaction of Font With Light



Selection of Subjects (Mice, Not People) and Assignment to Conditions. Renzi (1982) reported that mice were assigned to six conditions, with 16 mice in each condition. He did not specify how they were assigned, but it seems likely that they were randomly assigned.

The IVs and DVs. One independent variable in this experiment was dosage or amount of drug given. The IV had three levels: 0, no drug (saline); 2.5 mg/kg of body weight of depressant; and 5 mg/kg of depressant.

A second independent variable was number of doses, and this variable had two levels: single dose (S) and repeated dose (R) (one dose a day for 10 days).

The dependent variable was aggressive behavior, and Renzi (1982) measured number of bites during a specified period of time.

The Design. Renzi (1982) assessed all combinations of all levels of each variable. This particular design, like our hypothetical design, was a 3×2 randomized factorial design. There were three levels of the first independent variable (dosage) and two levels of the second independent variable (number of doses). Three groups of mice received a single dose, with one of the three groups receiving saline, one receiving 2.5 mg/kg of drug, and one receiving 5 mg/kg of drug. Three other groups of mice received repeated doses, with one of the three receiving saline, another group receiving the lower dose, and the third group receiving the higher dose. Table 7.6 illustrates the six conditions in this experiment.

TABLE 7.6 Randomized 3×2 Factorial Design

Number of doses	Dosage Group (mg/kg)		
	0	2.5	5
S (single)	S0	S2.5	S5
R (repeated)	R0	R2.5	R5

Source: Renzi (1982).

Mice in the S0 group received a single dose of saline with no depressant, and mice in the R5 group received repeated doses (one per day for 10 days) of 5 mg/kg of the drug.

The Statistical Analysis. This design has two independent variables, and different mice were assigned to each of the six conditions. An appropriate statistical analysis is a two-way ANOVA, which is the analysis Renzi (1982) used.

The Results. Some of the descriptive statistics from this experiment are presented in Table 7.7. We estimated these statistics from a graph included in the report.

The two-way ANOVA provided answers to the three questions of interest to the researcher: Does number of doses affect the behavior? Does dosage level influence aggressive behavior? And, do dosage and number of doses interact?

To address the first inferential question, we examine the outcome of the ANOVA for the first main effect. The ANOVA revealed that there was a significant main effect of number of doses, $F(1, 42) = 29.78, p < .001$. Figure 7.6 illustrates this main effect.

As you can see, there was certainly more aggressive behavior when the mice were given repeated doses of the drug.

The second question was, "Does aggressive behavior differ depending on the dosage level given?" The statistical analysis revealed that dosage level did make a significant difference in the aggressiveness of the mice, $F(2, 42) = 13.52, p < .001$. Let's examine Figure 7.7, the graph of the main effect of dosage.

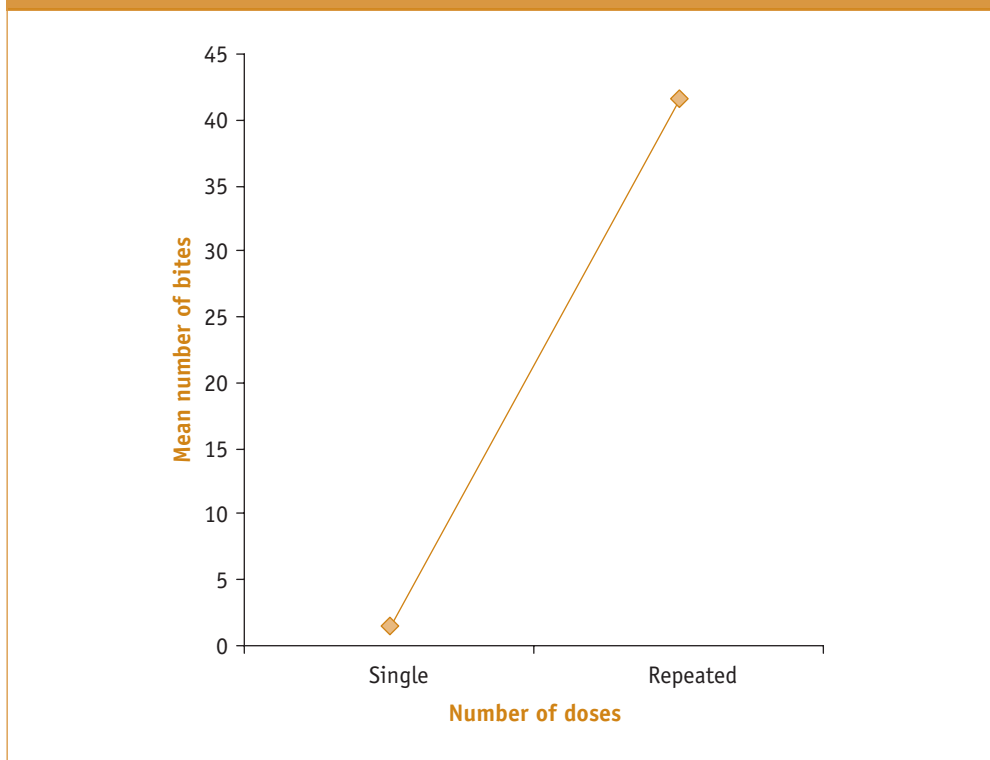
We can see that the mice that received the highest dose of the drug were the most aggressive. The mice that received the saline did not bite at all, it appears. The significant F from the ANOVA tells us that at least two means are different, so we know that the 5-mg group

TABLE 7.7 Descriptive Statistics:
Aggressive Behavior (Mean Number of Bites)

Number of doses	Dosage Group (mg/kg)		
	0	2.5	5
Single	~ 0	~ 3.5	~ 90
Repeated	~ 0	~ 1	~ 3

Source: Renzi (1982).

FIGURE 7.6 Main Effect of Number of Doses



was more aggressive than the 0 group, but Renzi (1982) did not include a post hoc comparison between the 2.5-mg and the 5-mg groups, so we do not know if those two groups are significantly different.

To really understand how the two variables operated, we must examine the interaction illustrated in Figure 7.8.

Renzi (1982) reported that the interaction was significant, $F(2, 42) = 11.43, p < .01$, and as you can see, the lines are not parallel. A single dose, either saline or the depressant drug, had little effect on the behavior of the mice regardless of the amount of drug given, but when repeated doses of the depressant were administered, aggressive behavior *increased* with the dosage level.

The Conclusions. Renzi (1982) speculated that effects of the depressant drug on aggressive behavior might be a result of an accumulation of the drug or its active metabolites that would only occur with its repeated administration.

You can see that the interpretation of the main effects must be made in light of any interaction. The factorial design allows us to investigate these more complicated and often more interesting relationships between variables.

152 METHODS IN PSYCHOLOGICAL RESEARCH

FIGURE 7.7 Main Effect of Dosage

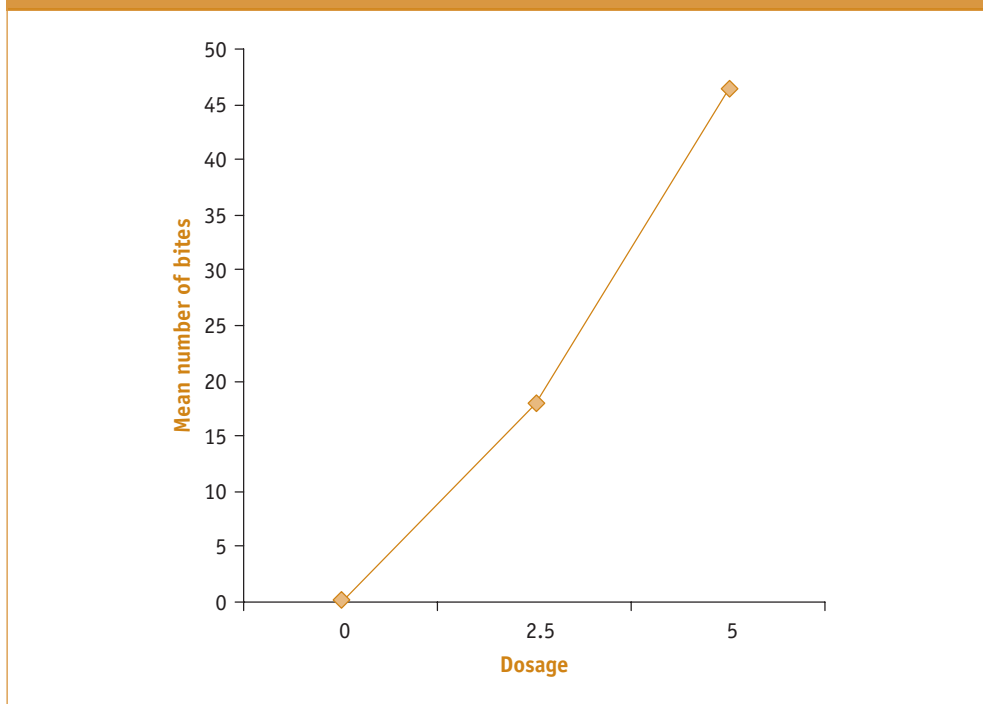
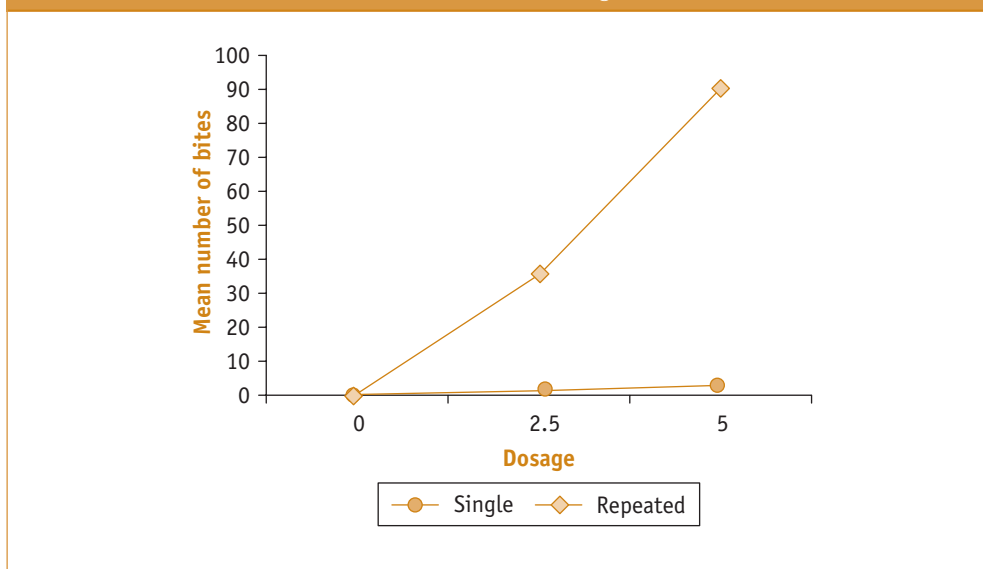


FIGURE 7.8 Interaction of Number of Doses With Dosage Level



Renzi's (1982) 3×2 factorial design involved two variables with three and two levels, respectively. The complexity of a factorial design increases with the number of levels of each variable and with the number of independent variables. With only two variables, the design yields two main effects and one interaction effect. With three independent variables, the design yields three main effects, three two-way interaction effects, and one three-way interaction effect. As you can imagine, the more variables, the more complicated the interactions, and the more difficult it is to interpret the findings. But given the complexity of human behavior, designs involving multiple variables are the rule in social science research today.

In addition to increasing complexity, the number of participants also dramatically increases with the addition of each new variable. Renzi (1982) had 16 single-dose mice and 16 repeated-dose mice in each of the three dosage-level conditions: 96 mice in total. What if he were to add a third independent variable; perhaps he also wanted to see if level of shock used to induce aggressive behavior made a difference. If he tested the mice under two shock conditions (high vs. low, perhaps), he would be adding six more groups to the experiment. If he used 16 mice per group, he would have needed 192 mice in his design!

Between-participants designs where participants are randomly and independently assigned to conditions are excellent experimental designs, particularly if there is any concern that one treatment condition might contaminate another. However, just because participants have been randomly assigned to conditions does not guarantee initial equivalence of groups. Nevertheless, random assignment of participants to conditions is the most common technique researchers use to deal with initial differences between participants. Another approach to ensuring equivalence of groups is addressed in Chapter 8.

CHAPTER SUMMARY

Experiments are the first choice of most behavioral researchers because they allow us to infer a **causal relationship** between a manipulated **independent variable** and some measure of behavior, the **dependent variable**. We conduct experiments to evaluate theories, satisfy our curiosity, and demonstrate behavioral phenomena and the factors influencing them.

When we conduct experiments, we begin with a **hypothesis**. We select appropriate independent and dependent variables to test our hypothesis. We make every attempt to control alternative sources of variation, and then we carry out our experiment. After analyzing the data, we are then in a position to draw **inferences** about the relationship between our manipulated independent variable and the observed behavioral change.

Experiments can be conducted in a laboratory and are called **controlled experiments**. When we conduct experiments in a natural setting, these are called **field experiments**. Controlled experiments tend to have greater **internal validity** than field experiments, but field experiments may have greater **external validity**. Some research problems are better examined in a natural setting, whereas others are better examined in the controlled conditions of the laboratory.

Independent groups or **between-participants designs** are used to compare different groups of participants, all of whom have been independently assigned to treatment

154 METHODS IN PSYCHOLOGICAL RESEARCH

groups. The simplest of these is the **completely randomized design** with one independent variable with two levels. When participants have been independently assigned to all combinations of more than one independent variable, we have a **randomized factorial design**. Factorial designs allow the simultaneous assessment of more than one IV and the interactions between IVs. The effects of each IV on the DV are called **main effects**, and the effect of combinations of levels of IVs on the DV is the **interaction effect**.

ANSWERS TO CONCEPTUAL EXERCISES

Conceptual Exercise 7A

1. Comparisons between offending and nonoffending youth are participant comparisons; this is a quasi-experiment.
2. We might assume that the children were independently assigned to each type of film, and so this would be a true experiment.
3. It seems reasonable to assume that the pigeons were independently assigned to groups—a true experiment.
4. Gender is a participant variable; this is a quasi-experiment.

Conceptual Exercise 7B

This research problem is probably best studied in the field. It would be difficult to simulate rock concert conditions in a laboratory setting. And it is doubtful that the same kinds of behavior would occur in a laboratory as those in the natural setting.

FAQ

Q1: Most psychology research is conducted on psychology students. Is that a problem?

A1: Using psychology students in research is a problem that affects the external validity of your study. To the degree that these participants are different from the population you want to generalize to, your study lacks external validity. Practically speaking, most researchers do not consider this a problem because their focus is on testing a null hypothesis. Whether experimental results conducted on university students will generalize to other groups is perhaps the topic of another experiment.

Q2: What is the difference between random assignment and random selection?

A2: Random assignment is a procedure that is used to assign participants to treatment groups (or levels). The goal is to form groups of participants who are initially equivalent on measures of the dependent variable. That is, if you were to measure the dependent variable at the start of the experiment, the groups should not differ. Random selection is a procedure for selecting a sample from a population. The goal is to make your sample representative of the population. Random assignment affects the internal validity of your study, and random selection affects the external validity.

Q3: Why is random assignment so important?

A3: Random assignment of participants to treatment groups (or levels of treatments) is crucial to the internal validity of the experiment. As mentioned in Question 2, random assignment is used to ensure that the groups are equivalent on any variable that may influence the dependent variable. For example, if your treatment and control groups differ on measures of the dependent variable, then it cannot be said that the difference was caused by the independent variable.

Q4: What is the difference between an experimental and a nonexperimental design?

A4: The main difference between an experimental and a nonexperimental design is control of the independent variable. In an experiment, the researcher has control over who receives the treatment. In a nonexperimental design, the participants have, in a sense, already been assigned to their group. For example, if you were conducting an experiment on the effects of alcohol on cognitive ability, you would assign sober participants to consume two, four, or six drinks and also assign some individuals to a nonalcoholic drink control group. Once everyone had consumed his or her assigned amount, you would administer your dependent variable measure. This is an experiment because *you* have formed the groups. If you were to do this study using a nonexperimental design, you could simply enter a bar and ask people how many drinks they have had and then administer your dependent variable measure. In the experiment, you could make causal statements about how the alcohol had influenced cognitive ability. In the nonexperimental design, you could make statements of relationship but not statements of causation. This is because you did not assign the participants to the conditions; they assigned themselves. And it could be that the individuals differ on cognitive ability even before they began drinking (i.e., perhaps their differences in cognitive ability are driving them to drink excessively).

Q5: If the experiment is the cornerstone of scientific research, why use any other approach?

A5: Some research problems simply do not lend themselves to an experimental approach. Psychologists are interested in many variables that are inherent in the participants and, therefore, cannot be manipulated. Developmental psychologists, for example, often study variables such as gender and age. Such participant variables can only be studied with a quasi-experimental design. In addition, there may be ethical problems or practical reasons that cause the researcher to choose a nonexperimental approach.

Q6: Are treatment levels used in both experimental and nonexperimental designs?

A6: There are treatment groups in quasi-experimental designs; however, the researcher does not form them. When the researcher does not have control over the assignment of participants to groups, such a design is not considered a true experimental design, and strong statements of causal relationship cannot be made.

Q7: When doing field experiments, do you have to tell people they are being observed?

A7: Although you must have your research examined by an Institutional Review Board, generally you do not have to obtain consent when observing people in a public setting. However, you must be able to guarantee anonymity of those observed. In other words, if you use video, it must be done in such a way that no one can be identified.

156 METHODS IN PSYCHOLOGICAL RESEARCH

CHAPTER EXERCISES

1. What are the main reasons for doing experiments?
2. What are the steps involved in conducting an experiment?
3. What is the difference between a controlled laboratory experiment and a field experiment?
4. Describe the advantages and disadvantages between controlled and field experiments.
5. How are participants assigned to groups in independent groups designs, and what is the purpose of that method?
6. What does it mean when we say there is a significant main effect?
7. What does it mean when we say there is a significant interaction effect?

CHAPTER PROJECTS

1. A social psychologist is interested in the effects of video games on children's hand-eye dexterity. Design a controlled experiment to investigate this research problem.
 - a. Specify IV (operationalize).
 - b. Specify DV (operationalize).
 - c. What is your research hypothesis?
 - d. Specify how you will select participants and assign them to conditions.
 - e. What is your statistical hypothesis and how will you test it?
2. Design a field experiment to investigate the research problem described in Question 1.
 - a. Specify IV (operationalize).
 - b. Specify DV (operationalize).
 - c. What is your research hypothesis?
 - d. Specify how you will select participants and assign them to conditions.
 - e. What is your statistical hypothesis and how will you test it?
3. Design an independent groups experiment to evaluate the following conceptual hypothesis: Children diagnosed with attention-deficit disorder (ADD) are more distractible on group tasks than on individualized tasks.
 - a. Specify IV (operationalize).
 - b. Specify DV (operationalize).
 - c. What is your research hypothesis?
 - d. Specify how you will select participants and assign them to conditions.
 - e. What is your statistical hypothesis and how will you test it?
4. You have conducted an experiment to determine how children diagnosed with ADD perform on group and individualized tasks. In addition, the tasks are classified as difficult or easy. The children were independently assigned to each of the four conditions (difficult group, easy group, difficult individualized, easy individualized). You have measured mean performance of the groups on talk solution.
 - a. What kind of design have you used?
 - b. What are the IVs and what is the DV?
 - c. Using the data on the next page, graph each main effect and the interaction using group means. Describe what seems to have occurred. Higher scores are better.

Type of Task	Task Difficulty	
	Easy	Difficult
Individual	10.0	7.0
Group	9.0	2.0

REFERENCES

- Chaudry, S., Schroter, S., Smith, R., & Morris, J. (2002). Does declaration of competing interests affect readers' perceptions? *British Medical Journal*, *325*(7377), 1391–1392.
- Fox, K. A., Webster, J. C., & Guerriero, F. J. (1972). Increased aggression among grouped male mice fed nitrazepam and flurazepam. *Pharmacological Research Communications*, *4*, 157–162.
- Jones, K. M., & Lungaro, C. J. (2000). Teacher acceptability of functional assessment-derived treatments. *Journal of Educational & Psychological Consultation*, *11*(3/4), 323–332.
- Renzi, P. (1982). Increased shock-induced attack after repeated chlordiazepoxide administration in mice. *Aggressive Behavior*, *8*, 172–174.



Visit the study site at www.sagepub.com/evansmprstudy for practice quizzes and other study resources.