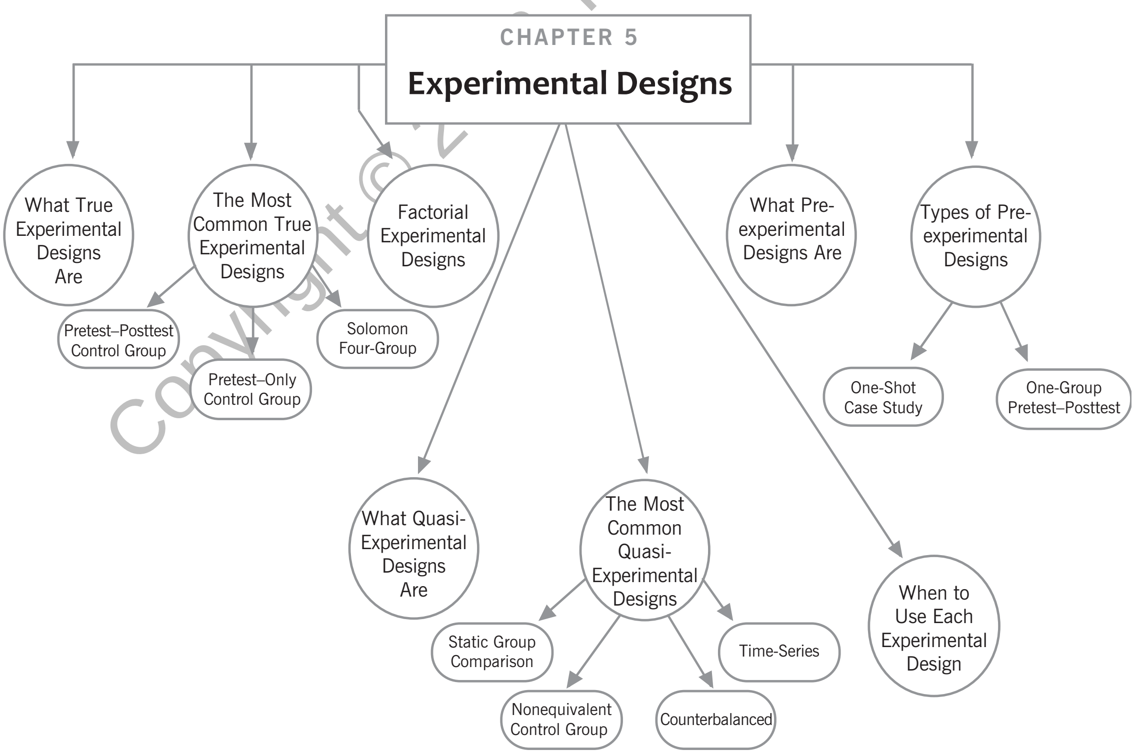
CHAPTER 5

Experimental Designs

OBJECTIVES

After studying this chapter, you should be able to . . .

1. State the goal of the experimental researcher and outline the four key features on which experimental designs depend to control for threats to internal and external validity.
2. Explain what is meant by true experimental designs.
3. Illustrate the most common true experimental designs.
4. Describe what is meant by factorial experimental designs.
5. Depict what is meant by quasi-experimental designs.
6. Outline the common quasi-experimental designs.
7. Express what is meant by preexperimental designs.
8. Illustrate the types of preexperimental designs.
9. Describe when researchers should use each experimental research design.



OVERVIEW

Researchers tend to be most interested in why things happen. Establishing cause-and-effect relationships among variables is typically the goal of most research. The experimental research designs described in this chapter test for the presence of a distinct cause-and-effect relationship between variables. In order to achieve this goal, experimental research must control extraneous variables. The results of a study using an experimental design indicate whether an independent variable produces or fails to produce changes in the dependent variable. For example, a simple experimental design would comprise two groups of participants randomly selected from a population in which one group (experimental group) receives the independent variable (intervention), and the other (control group) receives no intervention. Both groups are tested at the conclusion of the study to assess whether there is a difference in their scores. Assuming that the groups are equivalent from the start of the investigation, any observed differences at the conclusion of the investigation may reasonably be attributable to the independent variable, along with measurement and sampling error.

Findings from controlled experiments have led to much of our substantive knowledge in many areas, such as education, learning, memory, perception, and clinical psychology. The control achieved with experimental research designs enables researchers to ensure that the effects of the independent variable on the dependent variable are a direct (causal) consequence of the independent variable and not of other factors (i.e., extraneous variables or threats to internal validity). Of course, the experimental research designs used by researchers to establish cause-and-effect relationships differ widely in their ability to control for threats to internal and external validity (described in Chapter 2).

There are three categories of experimental research designs used by researchers in education and psychology: (1) true experimental, (2) quasi-experimental, and (3) preexperimental designs. These designs differ with regard to the level of experimental control they provide. Researchers must be concerned with the number of extraneous variables (threats to internal validity) that may effect changes in the performance of the experimental or control groups. Researchers must also be concerned with generalizing the results of studies to a broader population and set of conditions (external validity).

The ability of an experimental design to control for threats to internal and external validity is primarily dependent on four key features (Gall et al., 2007). The first key feature centers on the procedures used to select participants from a broader population. The usefulness of any given study is dependent on how well the results of that study can be generalized to a broader population and set of conditions. The second key feature involves the use of a control group or condition. Comparing the performance of an experimental group with that of a control group provides the basis for controlling many of the threats to internal and external validity. The comparison of conditions is critical for establishing whether there is a causal relationship between variables. The third key feature focuses on the initial equivalence of the experimental and control groups. Researchers must ensure this initial equivalence if they are to make relatively definitive conclusions regarding a causal relationship between the independent and dependent variables. The final key feature for experimental research centers on how effectively the investigation is conducted. Although researchers might employ a rigorous design, the results are useless if the investigation is not conducted well.

The three group experimental design categories differ on these design features that account for differences in the levels of experimental control. The final key feature (i.e., quality of implementation) cuts across all three categories of experimental designs. True experimental designs include clear procedures for addressing each of the first three key design features, including (1) the random selection of participants from a population (see Chapter 4), (2) the inclusion of a control group, and (3) the equivalence of the experimental and control groups. Quasi-experimental designs (with the exception of some time-series designs) include procedures for addressing the second key design feature (i.e., the inclusion of a control group). Preexperimental designs include none of these key design features. Table 5.1 summarizes the key design

TABLE 5.1. Key Design Features of True, Quasi-, and Preexperimental Research Designs

	Random selection of participants from a population	Random assignment of participants to conditions	Control group	Equivalence of groups
True	Yes	Yes	Yes	Yes
Quasi-	No	No	Yes	No
Preexperimental	No	No	No	No

features associated with true, quasi-experimental, and preexperimental research designs.

The remainder of this chapter includes a description of each of the true, quasi-experimental, and preexperimental research designs commonly used by researchers in education and psychology. Additionally, methods of analyzing data are described. (Refer to Figures 4.10, 4.11, and 4.12 for information regarding when to use each statistical analysis procedure to analyze data from investigations.) Finally, research examples are provided throughout the chapter to show true experimental and one quasi-experimental design, and an illustrative investigation is included at the end of the chapter for review.

Research Examples

Beginning with this chapter, we provide examples of research associated with the designs described in the readings. Think about how you would design research to address the questions/hypotheses from these research examples:

- Research Example 1: Do poor early reading skills decrease a child's motivation to read?
- Research Example 2: What are the effects of four first-grade math curricula?
- Research Example 3: What are the effects of pretesting and pedometer use on walking intentions and behavior?
- Research Example 4: What are the effects of sentence length on reading rate and the number of difficult words per paragraph on reading rate? Is there an interaction between the effects of sentence length and number of difficult words per paragraph on reading rates?
- Research Example 5: Do differences exist in dimensions related to quality of life for adults with developmental disabilities compared to adults in the general population?

Research examples appear throughout the chapter.

WHAT ARE TRUE EXPERIMENTAL DESIGNS?

True experimental designs are the only experimental designs that can result in relatively definitive statements about causal relationships between variables (Mertler & Charles, 2011). Researchers can argue rather decisively that there is a causal relationship between variables if they have effectively used a true experimental design. The beauty of true experimental designs rests in their simplicity in achieving the three requirements identified by Cook and Campbell (1979) in saying that one variable (independent variable) causes another (dependent variable). That is, true experimental designs ensure that (1) a change in the value of the independent variable is accompanied by a change in the value of the dependent variable, (2) how the independent variable affects the dependent variable is established a priori, and (3) the independent variable precedes the dependent variable.

There are three basic requirements for a research design to be considered a true experimental design. The first requirement is the random selection of participants from a population to form a sample. One of the sampling techniques discussed in Chapter 4 is to ensure that the participants selected are representative of the population. The random selection of participants from a population is a critical issue to the external validity of a study. It is important to note that researchers tend to be much more concerned with random assignment of participants to either experimental or control groups than with random selection of participants from a population to form a sample. This concern represents the tendency of researchers to be more mindful of a study's internal validity than its external validity.

The second requirement is that research participants must be randomly assigned to the experimental and control conditions. Random assignment helps to ensure that members of the experimental and control groups are equivalent to one another before the implementation of the independent variable. Ensuring that the experimental and control groups are equivalent is critical to the internal validity of the study. However, randomly assigning participants to groups does not necessarily guarantee initial equivalence between groups. Rather, random assignment only ensures the absence of systematic bias in the makeup of the groups. Assigning participants to experimental and control groups is accomplished by using one of the probabilistic sampling techniques discussed previously.

The equal treatment of members of experimental and control groups is the third requirement for a research design to be considered a true experimental design. Research participants in experimental and control groups must be treated equally in every way except in relation to the independent variable. In other words, participants in experimental and control groups are treated differently only with respect to the independent variable. However, note that the comparison of an independent variable (experimental group) with no independent variable (control group) is overly simplistic. The actual comparison in most true experimental studies is what occurs between the independent variable and the activities of the control group during the experimental time frame. Thus, the comparison might be better thought of as being between two *different* independent variables. It is important for critical research consumers to understand the strengths and weaknesses associated with true experimental designs.

WHAT ARE THE MOST COMMON TRUE EXPERIMENTAL DESIGNS?

We now look at three of the most common true experimental designs—pretest–posttest control-group, posttest-only control-group, and Solomon four-group designs. All three of these designs are presented in terms of comparing a single independent variable and dependent variable with a control condition. Designs with more than one independent variable also represent true experimental designs; we discuss these factorial designs separately.

Pretest–Posttest Control-Group Design

The **pretest–posttest control-group design** is one of the most common designs used in education and psychology to demonstrate a causal relationship between an independent variable and a dependent variable. This design begins with random selection of participants from a population to form a sample. Participants from the sample are then randomly assigned to experimental or control groups. Measurement of the dependent variable is taken prior to the introduction of the independent variable. The independent variable is then introduced, followed by postintervention measurement of the independent variable. Figure 5.1 depicts the form of the pretest–posttest control-group design.

The basic assumption of the pretest–posttest control-group design is that participants in the experimental and control groups are equivalent prior to introduction of the independent variable. Any differences observed at the end of the study are assumed to be due to the independent variable. The assumption that the experimental and control groups are equivalent is based on the notion that randomly assigning participants to either group ensures that they are equivalent at the start of the study. The extent to which this assumption is met is based on the technique used to randomly assign participants to groups and the number of participants in each group. Additionally, recall that randomly assigning participants to groups only ensures the absence of systematic bias in the makeup of the groups, not the initial equivalence of the groups.

The goal in a pretest–posttest control-group design is to keep the experiences of the experimental and control groups as identical as possible in all respects except for introduction of the independent variable to the experimental group. Changes in the pretest and posttest scores due to any other extraneous variables (e.g., maturation)

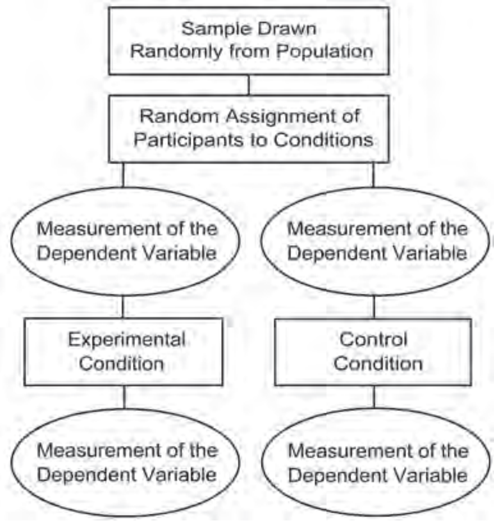


FIGURE 5.1. Pretest–posttest control-group design.

will be reflected in the scores of the control group. In other words, any changes in the posttest scores of the experimental group beyond the changes in the control group can be reasonably attributed to the independent variable.

In some studies using a pretest–posttest control-group design, the control group is administered only in the pretest and the posttest, and receives no specific intervention. In other studies using such a design, the control group is administered an equally desirable but alternative intervention or independent variable. Sometimes researchers refer to the control as a "comparison" group, if the group receives an intervention rather than being in a no-intervention condition. Furthermore, researchers may label the two groups in relation to the interventions the groups receive (e.g., direct instruction group, cooperative learning group).

Analysis of Data

Although the tests of statistical significance discussed in Chapter 4 can be used with the pretest–posttest control-group design, the ANCOVA is preferred. With ANCOVA, the posttest mean score of the experimental group is compared with the posttest mean score of the experimental group, with the pretest scores used as a covariate. Recall from Chapter 4 that ANCOVA statistically adjusts the posttest scores for initial differences between the experimental and control groups on the pretest. The ANCOVA is the preferred test of statistical significance because it is the most powerful. That is, ANCOVA increases the probability that researchers will detect the effects of the independent variable. Additionally, a nonparametric test, such as the Mann–Whitney *U* test, should be used if the data violate the assumptions underlying these parametric tests (i.e., homogeneity of variance, normal distribution of data, and interval or ratio scale data).

Internal Validity. The pretest–posttest control-group design in which the control group receives no intervention effectively controls for the eight threats to internal validity that result in changes in the performance of the experimental group. These threats to internal validity include (1) history, (2) maturation, (3) testing, (4) instrumentation, (5) statistical regression, (6) selection, (7) mortality, and (8) selection by maturation interaction. Testing may be a threat to the internal validity of the study if the pretest has a powerful effect on the intervention. The four additional threats to internal validity that cause changes in the performance of the control group are controlled if researchers provide the control group with an equally desirable but alternative intervention. These threats to internal validity include (1) experimental treatment diffusion, (2) compensatory rivalry by the control group, (3) compensatory equalization of treatments, and (4) resentful demoralization of the control group. Table 5.2 summarizes the potential threats to internal validity associated with each of the true experimental research designs.

External Validity. The pretest–posttest control-group design effectively controls for many of the threats to external validity if the study is conducted effectively. Table 5.3 summarizes the potential threats to external validity associated with each of the true experimental research

> *The* **pretest–posttest control-group design** *includes measurement of the dependent variable before and after implementation of the independent variable.*

designs. The threats to external validity that are controlled for include (1) multiple treatment interference, (2) novelty and disruption effects, (3) the Hawthorne effect, (4) pretest sensitization, and (5) posttest sensitization. It is important to note that novelty and disruption effects, as well as the Hawthorne effect, may be threats to the external validity of a study if the researchers did not provide an equally desirable but alternative intervention. Pretest sensitization also may be a threat to the external validity of the study if the pretest has a powerful effect on the intervention. The remaining threats to the external validity of a study are dependent on the particular characteristics of the study and how well the study was conducted (see Table 5.3). For example, as discussed earlier, if the participants have not been selected randomly from a population to form a sample, then the threats to external validity associated with the population represent potential problems.

Recall the research question regarding early reading failure and whether it decreases children's motivation to read? Read on to see how your design compares to how the actual study was designed.

Research Example 1: Early Reading Failure and Children's Motivation to Read

The pretest–posttest control group design was used to evaluate whether low performance in early reading decreased children's motivation to practice reading, and whether improving word identification skills with a peer tutor would bolster motivation (Morgan, Fuchs, Compton, Cordray, & Fuchs, 2008). Researchers sought to investigate the causal relationship between low performance in early reading and decreased motivation. Thirty classroom teachers were recruited in 15 schools in a large metropolitan area. Within teachers' classrooms, 75

TABLE 5.2. Threats to Internal Validity Associated with True Experimental Designs

	Research design		
Threat	Pretest–postttest control group	Posttest-only control group	Solomon four-group
1. Maturation	Controlled	Controlled	Controlled
2. Selection	Controlled	Controlled[c]	Controlled
3. Selection by maturation interaction	Controlled	Controlled	Controlled
4. Statistical regression	Controlled	Controlled	Controlled
5. Mortality	Controlled	Controlled	Controlled
6. Instrumentation	Controlled	Controlled	Controlled
7. Testing	Controlled[a]	Controlled	Controlled
8. History	Controlled	Controlled	Controlled
9. Resentful demoralization of the control group	Possible concern[b]	Possible concern[b]	Possible concern[b]
10. Diffusion of treatment	Possible concern[b]	Possible concern[b]	Possible concern[b]
11. Compensatory rivalry by the control group	Possible concern[b]	Possible concern[b]	Possible concern[b]
12. Compensatory equalization	Possible concern[b]	Possible concern[b]	Possible concern[b]

Note. This table is meant only as a general guideline. Decisions with regard to threats to internal validity must be made after the specifics of an investigation are known and understood. Thus, interpretations of internal validity threats must be made on a study-by-study basis.

[a]Although testing is generally controlled for, it may be a potential threat to the internal validity of a study if the pretest has a powerful effect on the intervention.

[b]These threats to internal validity are controlled if the control group received an equally desirable but alternative intervention.

[c]Although selection is controlled through the random assignment of participants to groups, the lack of a pretest precludes a statistical test of the equivalence of the groups.

TABLE 5.3. Threats to External Validity Associated with True Experimental Designs

	Research design		
Threat	Pretest–postttest control group	Posttest-only control group	Solomon four-group
1. Generalization across participants	Possible concern[a]	Possible concern[a]	Possible concern[a]
2. Interaction of personological variables and treatment effects	Possible concern[a]	Possible concern[a]	Possible concern[a]
3. Verification of independent variable	Possible concern	Possible concern	Possible concern
4. Multiple treatment interference	Controlled	Controlled	Controlled
5. Novelty and disruption effects	Controlled[b]	Controlled[b]	Controlled[b]
6. Hawthorne effect	Controlled[b]	Controlled[b]	Controlled[b]
7. Experimenter effects	Possible concern	Possible concern	Possible concern
8. Pretest sensitization	Possible concern	Controlled	Controlled
9. Posttest sensitization	Controlled	Possible concern	Controlled
10. Interaction of time of measurement and treatment effects	Possible concern	Possible concern	Possible concern
11. Measurement of the dependent variable	Possible concern	Possible concern	Possible concern
12. Interaction of history and treatment effects	Possible concern	Possible concern	Possible concern

Note. This table is meant only as a general guideline. Decisions with regard to threats to external validity must be made after the specifics of an investigation are known and understood. Thus, interpretations of external validity threats must be made on a study-by-study basis.

[a]Threats to external validity associated with the population are controlled if the sample has been drawn randomly from a population.

[b]These threats to external validity are controlled if the control group received an equally desirable but alternative intervention.

first graders were recruited, including 30 who were high-skilled readers, 15 who were low-skill readers, and 15 additional low-skill readers who received small-group tutoring. Of 30 low-skill participants, 12 were African American, 25 were European American, and 15 participated in Title I schools. Low-skill children were then randomly assigned to a tutoring or nontutoring condition. Tutoring consisted of a mean of 27 hours of small-group tutoring by trained graduate students in phonemic awareness, decoding, and fluency building. Researchers conducted observations measuring the fidelity of tutoring to ensure that tutors adhered to specific teaching steps. Tutoring fidelity was 98%. Tutoring was additive to standard reading instruction. Informal classroom observation confirmed that standard reading instruction was carried out during the study. Eight different measures were used in the pretest and posttest format. Emerging reading skills were measured by administering the Rapid Letter Naming Test from the Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Rashotte, 1999) and three additional tests. Reading skills were measured by the Word Attack and Word Identification subtests of the Woodcock Reading Mastery Test—Revised (Woodcock, 1987) and a first-grade Dolch Sight Word List. Reading motivation was measured by the Reading Self-Concept Scale (Chapman & Tunmer, 1995) and two additional tests. Reading practice was measured by the Reading Frequency Questionnaire (Bast & Reitsma, 1998) and one additional test. Psychometric data were reported on all tests. Results were analyzed using repeated measures ANOVA. Researchers found consistent evidence that reading skill and reading motivation were related. Low-skill readers reported lower reading self-concept than high-skill readers. Teachers rated low-performers as less motivated and more task avoidant during reading. Teachers also rated low-skill readers as less likely to independently practice reading. However, low-skill readers in the tutoring group did not have concomitant changes in reading self-concept, motivation, or task orientation. Although their reading improved, they did not increase their practice of reading. Researchers recommended combining reading tutoring with strategies directly targeting motivation.

Posttest-Only Control-Group Design

The concept of applying a pretest in experimental research is very common in education and psychology. Although it is difficult for researchers to give up the comfort of knowing for sure whether the experimental and control groups are equivalent, applying a pretest is not essential to conducting a true experimental research study. Randomization typically leads to equivalent experimental and control groups (Campbell & Stanley, 1963). The **posttest-only control-group design** is very similar to the pretest–posttest-only control-group design except that pretests of the dependent variable are not administered to the experimental and control groups. The posttest-only control-group design begins with the random selection of participants from a population to form a sample. Participants from the sample are then randomly assigned to the experimental or control groups. The independent variable is then introduced followed by postintervention measurement of the dependent variable. Figure 5.2 depicts the form of the posttest-only control-group design.

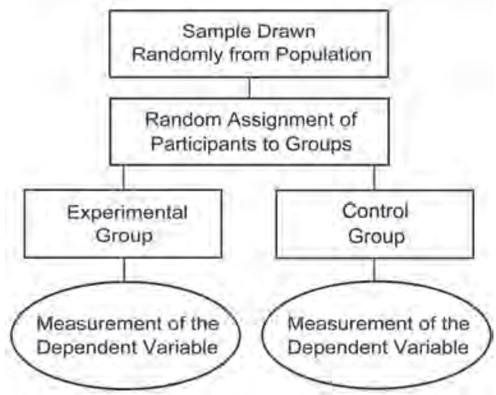


FIGURE 5.2. Posttest-only control-group design.

The basic assumption of the posttest-only control-group design is the same as the pretest–posttest control-group design. That is, the participants of the experimental and control groups are equivalent prior to the introduction of the independent variable. Any differences observed at the end of the study are assumed to be due to the independent variable. Of course, one of the weaknesses of the posttest-only control-group design is that researchers cannot be certain that random assignment of participants resulted in the initial equivalence of the groups.

The goal of the posttest-only control group design is the same as the pretest–posttest control-group design. This goal is to keep the experiences of the experimental and control groups as identical as possible in all respects except for the introduction of the independent variable to the experimental group. Changes in the posttest scores due to any other extraneous variables will be reflected in the scores of the control group. Additionally, as with pretest–posttest control-group designs, the control group in a posttest-only control-group design receives no specific intervention in some cases. In other cases, the control group is administered an equally desirable but alternative intervention.

The posttest-only control-group design is used when there is a possibility that the pretest will have an effect on the independent variable, or when researchers are unable to identify a suitable pretest measure (Gall et al., 2007). The posttest-only control-group design is most often used in research exploring the effects of different interventions on the beliefs or attitudes of individuals. For example, say we are interested in studying the effects of a new substance abuse prevention program on high school students' beliefs about alcohol. A pretest in this case might influence participants' responses on the posttest. Thus, participants would be randomly assigned to either the experimental or the control condition. The experimental group would receive the new substance abuse prevention program, and the control group would receive no intervention. At the end of the intervention, a posttest would be administered to both the experimental and the control group. Any differences (statistically significant) in the responses of the experimental and control groups could then be attributed to the effects of the substance abuse prevention program.

Analysis of Data

The data from a posttest-only control-group design are typically analyzed by doing a *t*-test comparing the means of the posttest scores of the experimental and control groups. ANOVA can be used if more than two groups have been studied. If the researcher has collected data on one or more variables unrelated to the purpose of the study, such as gender or IQ, then ANCOVA can be used. Additionally, a nonparametric test such as the Mann–Whitney *U* test should be used if the data violate the assumptions underlying these parametric tests (i.e., homogeneity of variance, normal distribution of data, and interval or ratio scale data).

Internal Validity. The posttest-only control-group design in which the control group receives no intervention effectively controls for the same eight threats to internal validity as the pretest–posttest control-group design (i.e., history, maturation, testing, instrumentation, statistical regression, selection, mortality, and selection by maturation interaction) that result in changes in the performance of the experimental group (see Table 5.2). A disadvantage of the posttest-only control-group design is that it does not enable researchers to check the initial equivalence of the experimental and control groups. Additionally, the posttest-only control-group design effectively controls for the four threats to internal validity (i.e., experimental treatment diffusion, compensatory rivalry by the control group, compensatory equalization of treatments, and resentful demoralization of the control group) that result in changes in the performance of the control group, if that group received an equally desirable but alternative intervention (see Table 5.2).

External Validity. As with the pretest–posttest control-group design, the posttest-only control-group design controls for many of the threats to external validity if the study is conducted effectively (see Table 5.3). A particular advantage is that the posttest-only control group design is able to control for pretest sensitization. That is, because there is no pretest, it is unlikely that participants can gain information allowing them to perform better on the posttest. Novelty and disruption effects and the Hawthorne effect may be threats to the external validity of a study if the researchers did not provide an equally desirable but alternative intervention. Posttest sensitization also may be a threat to the external validity of the study if the posttest has a powerful effect on the intervention. The remaining threats to the external validity of a study are dependent on the particular characteristics of the study and on how well it was conducted (see Table 5.3).

Recall the research question regarding comparison of four different math curricula? Read on to see how your design compares to how the actual study was designed.

Research Example 2: Comparison of Fourth-Grade Math Curricula

A posttest-only control-group design was used to compare four first-grade math curricula (Agodini & Harris, 2010). Over 8,000 first and second graders in 110 schools in 12 districts of 10 states participated. All participants were randomly assigned to one of four curricula. Random assignment was conducted at the school level. A random sample of 10 participants per classroom was included in the analysis. Because of the large number of participants and random assignment, researchers were relatively certain that performance across curriculum groups was equivalent at the outset. Independent variables were the math curricula including (1) Investigations in Number, Data, and Space; (2) Math Expressions; (3) Saxon Math, and (4) Scott Foresman–Addison Wesley Mathematics. Investigations in Number, Data, and Space is a student-centered approach that focuses on student understanding rather than specific problem-solving procedures. Math Expressions is a blend of student-centered and teacher-directed approaches. Saxon Math is a teacher-directed approach, with scripted lessons and daily student practice providing lessons on solving

The **posttest-only control-group design** *includes measurement of the dependent variable after implementation of the independent variable.*

problems. Scott-Foresman–Addison Wesley Mathematics is an approach combining teacher-directed instruction with differentiated student activities, allowing teachers to select relevant and appropriate materials, including manipulatives. Dependent measures were end-of-year test scores on a nationally normed math assessment developed for the Early Childhood Longitudinal Study for first and second graders. For first graders, researchers found that mean math achievement on the posttest was higher for Math Expressions and Saxon Math, although neither was statistically significant. For second graders, researchers found a statistically significant difference with highest performance by the Saxon Math group.

Solomon Four-Group Design

The **Solomon four-group design** offers researchers the greatest amount of experimental control. Like the pretest–posttest control-group and posttest-only control-group designs, the Solomon four-group design assesses the effects of the independent variable relative to a control condition. Unlike these designs, the Solomon four-group design enables researchers to assess the presence of both pretest sensitization and an interaction between the pretest measures and the independent variable. Although the posttest-only control-group design controls for pretest sensitization and for an interaction between the pretest measures and the independent variable, it does not enable researchers to determine their presence. Thus, the Solomon four-group design not only controls for testing (threat to internal validity) and pretest sensitization (threat to external validity), but it also enables researchers to assess their effects on the intervention outcomes.

The Solomon four-group design essentially combines the pretest–posttest control-group design and the posttest-only control-group design (compare Figure 5.3 with Figures 5.1 and 5.2). The Solomon four-group design begins with the random selection of participants from a population to form a sample. Participants from the sample are then randomly assigned to one of four groups. Two groups serve as experimental groups, and two serve as control groups. Measurement of the dependent variable is taken prior to the introduction of the independent variable with one of the experimental groups and one of the control groups. The independent variable is then introduced, followed by postintervention measurement of the dependent variable for all four of the groups. Figure 5.3 depicts the form of the Solomon four-group design.

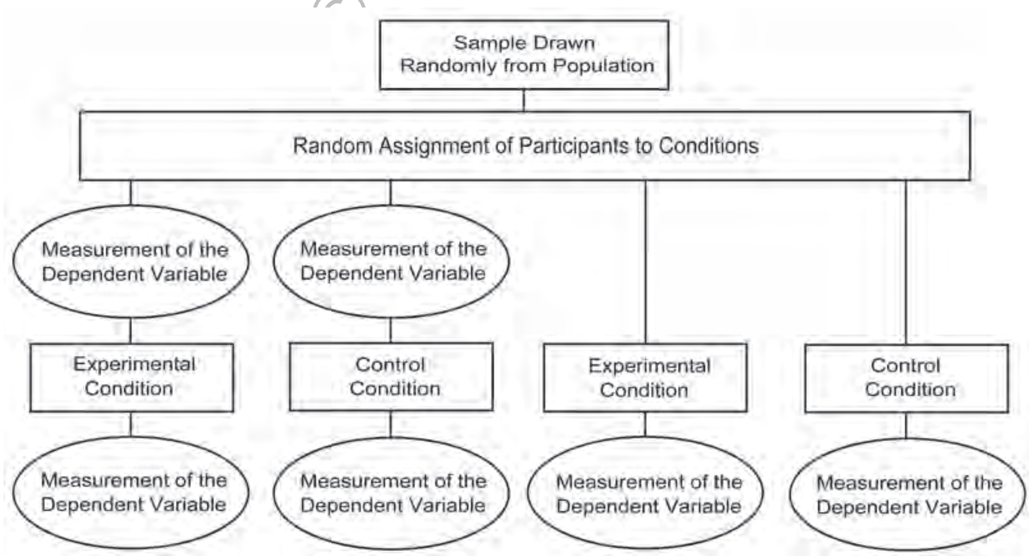


FIGURE 5.3. Solomon four-group design.

The basic assumption of the Solomon four-group design is the same as the pretest–posttest and posttest-only control-group designs, in that the participants of the experimental and control groups are equivalent prior to the introduction of the independent variable. Any differences observed at the end of the study are assumed to be due to the independent variable and, in some cases, the preintervention measures of the dependent variable. As we might expect, the Solomon four-group design requires a great deal more effort and resources to implement than the pretest–posttest and the posttest-only control-group designs. The extra effort and resources needed to implement a Solomon four-group design may be worth it in cases in which it is critical to determine the effects of pretest sensitization and the interaction of the pretest and the independent variable.

The goal of the Solomon four-group design is the same as that of the pretest–posttest and posttest-only control-group designs; that is, the goal is to keep the experiences of the experimental and control groups as identical as possible except for the introduction of the independent variable. Changes in the posttest scores due to any other extraneous variables will be reflected in the scores of the control group.

Analysis of Data

The Solomon four-group design, in its simplest form, is essentially a 2 × 2 factorial design (i.e., two or more independent variables [called factors] affect the dependent variable either independently [main effect] or in combination with each other [interaction effect]), in which the presence or absence of a pretest is one factor (signified by the first "2"), and the presence or absence of the independent variable is the second factor (signified by the second "2"). (*Note*. Factorial designs are discussed below.) Thus, data from a Solomon four-group design is typically analyzed with an ANOVA. Any significant main effects (e.g., differences between the presence and absence of the independent variable) and interaction effects (e.g., the presence or absence of the pretest differentially affects the independent variable such that the independent variable is more or less effective when the pretest is provided) are then explored using post hoc analysis procedures such as the Scheffé test. Additionally, a nonparametric test, such as the Kruskal–Wallis test, should be used if the data violate the assumptions underlying parametric tests (i.e., homogeneity of variance, normal distribution of data, and interval or ratio scale data).

Internal Validity. The Solomon four-group design controls for all eight of the threats to internal validity (i.e., history, maturation, testing, instrumentation, statistical regression, selection, mortality, and selection by maturation interaction) that result in changes in the performance of the experimental groups (see Table 5.2). As discussed before, an important advantage of the Solomon four-group design over the pretest–posttest and the posttest-only control-group designs is that it enables researchers to determine the effects of the preintervention measurement of the dependent variable on the postintervention measurement of the dependent variable. Additionally, the Solomon four-group design effectively controls for the four threats to internal validity (i.e., experimental treatment diffusion: compensatory rivalry by the control group, compensatory equalization of treatments, and resentful demoralization of the control group). These may result in changes in the performance of the control group if the control group received an equally desirable but alternative intervention (see Table 5.2).

External Validity. The Solomon four-group design controls for many of the threats to external validity if the investigation is conducted effectively (see Table 5.3). Novelty and disruption effects and the Hawthorne effect may be threats to the external validity of a study if researchers do not provide an equally desirable but alternative intervention. Although pretest sensitization is controlled for, posttest sensitization may be a threat to the external validity of the study if the posttest has a powerful effect on the intervention. The remaining threats to the external validity of a study are dependent on the

> *The* **Solomon four-group design** *is a combination of the pretest–posttest control-group design and the posttest-only control-group design.*

particular characteristics of the study and how well the study was conducted.

Recall the research question regarding the effects of pretesting and pedometers? Read on to see how your design compares to how the actual study was designed.

Research Example 3: Effects of Pretests and Pedometer Use

A Solomon four-group design was used to assess effects of pretesting and pedometers on walking intentions and behavior (Spence, Burgess, Rodgers, & Murray, 2009). The purpose of the study was to determine whether the administration of a pretest questionnaire differentially influenced the posttest self-report of walking behavior of participants as a result of using a pedometer. Participants were 63 female university students randomly assigned to one of four conditions: (1) pedometer and pretest, (2) pedometer and no pretest, (3) no pedometer and pretest, and (4) no pedometer and no pretest. There were 16 participants in all conditions except no pedometer and pretest (n = 15). Pretest conditions included questions on walking, intentions to walk, and self-efficacy for walking. The Solomon four-group configuration was used to control for pretest sensitization. Pedometer and pretest participants completed the walking behavior and self-efficacy questionnaires. Pedometer and pretest participants, as well as pedometer-only participants, were then given a pedometer and provided with information about the device. No pedometer and pretest, as well as pedometer and no-pretest participants, were told they would have the opportunity to wear the pedometer later. Posttest conditions were identical to pretest conditions and were administered 1 week later. Data were analyzed using a 2 × 2 (Pedometers × Pretest) ANOVA. The researchers note that significant pretest × pedometer interactions would have indicated the presence of pretest sensitization; however, no such results were obtained. After controlling for pretest reports of walking, wearing pedometers resulted in statistically significant increases in self-reports of walking. About 75% of pedometer users returned log sheets indicating that average number of steps per day was 10,293. Positive correlations were found between self-reported walking and average number of steps taken per day, and total steps taken per week.

WHAT ARE FACTORIAL EXPERIMENTAL DESIGNS?

Up to now, we have focused on experimental research designs that incorporate one independent variable (i.e., single factor). The goal in such designs is to establish a causal relationship between two variables. However, researchers are often interested in assessing the effects of two or more independent variables (factors) and in identifying which participants benefit most from the independent variable. Researchers use **factorial experimental designs** to assess the effects of two or more independent variables (described in this chapter) or the interaction of participant characteristics with the independent variable (described in Chapter 6, this volume; Graziano & Raulin, 2010). As such, researchers use factorial designs to determine whether the effect of a particular variable studied concurrently with other variables will have the same effect as it would when studied in isolation.

Analysis of Data

The factorial designs and associated statistical analysis procedures used by researchers can be quite complex. For example, a factorial design in a study that compares three methods of teaching reading (i.e., direct instruction, whole language, and language experience) could be extended to include a comparison of short (i.e., periods of intensive instruction) versus massed (i.e., extended period of intensive instruction) learning conditions. The effect on achievement of three methods of teaching and two learning conditions could be investigated with a 3 (direct instruction vs. whole language vs. language experience) × 2 (short vs. massed learning) ANOVA.

It is informative to look more closely at the preceding example to identify the two factors included in factorial designs. The first and second factors are the three methods of reading instruction (i.e., direct instruction, whole language, and language experience) and the type of instruction (short vs. massed learning), respectively. These factors are considered independent variables and are manipulated (termed *stimulus variables*). Additionally, it is important to note that factorial designs may involve repeated measurement of the same participants. The repeated measurement of participants on one or more factors can greatly complicate a factorial design.

A key advantage of factorial designs is that information is obtained about the interaction between factors. Going back to the previous example, one method of teaching reading may interact with a condition of learning and render that combination either better or worse than any other combination. Of course, one of the disadvantages of the factorial design is that the number of combinations may become quite unwieldy to conduct and difficult to interpret. In education and psychology, it is advisable to avoid overly complex factorial experiments. We direct the reader to Ferguson (1989) and McBurney and White (2009) for more information on factorial designs. The particular factorial design used by researchers primarily depends on the following factors:

1. The number of independent variables.
2. The attributes of the independent variable (referred to as a *between-subjects factor* when comparing different groups).
3. Repeated measurement of participants (also referred to as a *within-subjects factor*).
4. Mixing within (repeated measures) and between (comparison of group means) subjects factors.
5. Relative number of participants in each intervention group.

Internal Validity

The threats to internal validity that a factorial design controls for are dependent on the basic underlying true experimental design (see Table 5.2). For example, if the basic, underlying true experimental design is a pretest–posttest control group design, the eight threats to internal validity (i.e., history, maturation, testing, instrumentation, statistical regression, selection, mortality, and selection by maturation interaction) that result in changes in the performance of the experimental group are controlled. Additionally, the four threats to internal validity that cause changes in the performance of the control group are controlled for in factorial designs, because the different experimental groups are receiving some form of the intervention.

External Validity

One of the greatest strengths of a factorial design is that it enhances the external validity of the study. Analyzing the effects of the intervention on different subsets of the sample increases the extent to which the results can be generalized "across participants" and provides an understanding of the interaction of personological variables and treatment effects. As with all of the true experimental designs, factorial designs control for many of the threats to external validity if the investigation is conducted effectively (see Table 5.3).

Recall the research question regarding the effects of sentence length and number of difficult words per paragraph on reading rates? Read on to see how your design compares to how the actual study was designed.

Research Example 4: Effects of Sentence Length and Number of Difficult Words per Paragraph on Reading Rates

This example was not drawn from experimental literature; it is hypothetical. However, we include it to illustrate important characteristics and processes related to factorial designs. We consider an example comparing a single-factor experimental design and a two-factor design that address the same research question to show how researchers use factorial designs. Suppose we are interested in assessing the effects of sentence length and number of difficult words per paragraph on the reading rates

> **Factorial experimental designs** *assess the effects of two or more independent variables or the interaction of participant characteristics with the independent variable.*

of elementary-school-age students. The two independent variables of interest are sentence length and number of difficult words per paragraph. Suppose we choose two sentence lengths (≤ 20 words and > 20 words) and two levels of the number of difficult words per paragraph (≤ 2 and > 2). We would need to use two separate studies to assess the effects of these two independent variables using single-factor experimental designs. Table 5.4 presents what these two experimental designs might look like. The study on the top (Design 1) is designed to assess the effects of sentence length on reading rate, and the study in the middle (Design 2) is designed to assess the effects of the number of difficult words per paragraph. In both studies, 20 participants would be assigned to each of the experimental conditions, for a total of 40 participants per study (total of 80 participants). Except for students in the sentence length and the number of difficult words per paragraph conditions, all of the participants would be treated the same. At the completion of the study, we would analyze the data and be able to make a statement regarding the influence of sentence length and number of difficult words per paragraph on rate of reading.

Compare these two single-factor studies with the two-factor factorial design (Design 3) at the bottom of Table 5.4. The two independent variables (i.e., sentence length and number of difficult words per paragraph) are manipulated simultaneously. Because both independent variables have two levels, there are four (2 × 2) unique groups. This design would be called a 2 × 2 factorial design. Inspection of Table 5.4 reveals that the sample size in each group is 10. This number was chosen to provide a direct comparison with the single-factor studies (i.e., we start with 40 participants, then randomly assign 10 participants to serve in each of the four groups). At the completion of the study, we would analyze the data and make a statement regarding the influence of sentence length and number of difficult words per paragraph on rate of reading (main effects). We would make a statement regarding the interaction of sentence length and number of difficult words per paragraph (i.e., interaction effects).

Our comparison of two single-factor designs with a two-factor factorial design is only correct up to a point. The factorial design provides the same information as the two single-factor designs only when there is no interaction

TABLE 5.4. Comparison of Single-Factor and Two-Factor Experimental Designs

Design 1

Sentence length	
≤ 20 words 20 participants	> 20 words 20 participants

Design 2

Average number of difficult words per paragraph	
≤ 2 words 20 participants	> 2 words 20 participants

Design 3

	Average number of difficult words per paragraph	
Sentence length	≤ 2 words	> 2 words
≤ 20 words	10 participants	10 participants
> 20 words	10 participants	10 participants

between the two independent variables. There are no interaction effects when the effect of one of the independent variables is the same at each level of the other independent variable. Returning to our example, the effects of sentence length are the same regardless of the number of difficult words per paragraph and vice versa. On the other hand, if the effects of sentence length are different across the different levels of the number of difficult words per paragraph, there is an interaction, and the information provided by the main effect is not the same as that in the single-factor designs. This finding is not problematic, because the researchers would have discovered information that is not available from the single-factor designs—how the two independent variables combine to influence reading rate. Additionally, researchers will not be as interested in the main effects when there is an interaction effect, because anything they say with regard to the effects of one independent variable would have to be qualified by its differential effect across the levels of the other independent variable.

Finally, it is important to address the concepts of main and interaction effects, because they are key concepts associated with factorial designs. We examine an illustrative example in which a 2 × 2 factorial design yields only main effects, and in which the same 2 × 2 factorial design yields interaction effects. It is important to note that we are presenting a simplified explanation of main and interaction effects here. We direct the reader to Keppel (1973) and Ferguson (1989) for more information on main and interaction effects.

Using the same example noted earlier, Table 5.5 presents Study 1, in which a main effect for sentence length was obtained, and Study 2, in which an interaction effect between sentence length and number of difficult words per paragraph was obtained. Inspection of the results of Study 1 reveals that changes in participants' reading rates for sentences with ≤ 20 words and > 20 words were relatively (statistically similar) consistent across the number of difficult words per paragraph. Participants generally read faster when the sentences contained < 20 words. These results are consistent with only obtaining a main effect. On the other hand, inspection of the results of Study 2 reveals a different pattern. Participants' reading rates differed across the number of difficult words per paragraph. The reading rates of participants were similar regardless of sentence length if there were less than two difficult words per paragraph. However, the reading rates of participants were significantly lower when there were more than two difficult words per paragraph.

TABLE 5.5. Comparison of Main Effects and Interaction Effects

Study 1

	Average number of difficult words per paragraph	
Sentence length	≤ 2 words	> 2 words
	Mean	*Mean*
≤ 20 words	112	114
> 20 words	90	88

Study 2

	Average number of difficult words per paragraph	
Sentence length	≤ 2 words	> 2 words
	Mean	*Mean*
≤ 20 words	110	115
> 20 words	112	85

WHAT ARE QUASI-EXPERIMENTAL DESIGNS?

Quasi-experimental designs differ from true experimental designs in two ways. First, participants are not randomly selected from a specified population. Second, participants are not randomly assigned to experimental and control groups. Participants are intact groups, such as all sophomores in a high school classroom or members of an afterschool study group. Nevertheless, quasi-experimental designs provide a relatively high degree of experimental control in natural settings, and they clearly represent a step up from preexperimental designs (described later), because they enable researchers to compare the performance of the experimental group with that of a control group. In other words, quasi-experimental designs enable researchers to move their experimentation out of the laboratory and into a natural context. It is important for critical research consumers to understand the strengths and weaknesses associated with quasi-experimental designs.

WHAT ARE THE COMMON QUASI-EXPERIMENTAL DESIGNS?

We examine four common quasi-experimental designs—static-group comparison, nonequivalent control-group, counterbalanced, and time-series designs. Note that these four designs are presented in terms of a single independent variable and dependent variable.

Static-Group Comparison Design

The **static-group comparison design** begins with the identification of two naturally assembled experimental and control groups (e.g., students in two classrooms). The naturally assembled experimental and control groups should be as similar as possible, and the assignment to one group or the other is assumed to be random. The independent variable is then introduced to the experimental group, followed by the postintervention measurement of the dependent variable. Figure 5.4 depicts the form of the static-group comparison design.

Analysis of Data

The data from a static-group comparison design can be analyzed with a *t*-test of the difference between the posttest mean scores of the experimental and control groups. Additionally, a nonparametric test such as the Mann–Whitney *U* test should be used if the data violate the assumptions underlying these parametric tests (i.e., homogeneity of variance, normal distribution of data, and interval or ratio scale data).

Internal Validity. The use of a comparison group in the static-group comparison design enhances its experimental control in comparison to preexperimental designs (described later). However, in the absence of random assignment of participants to groups, the lack of a pretest greatly weakens its ability to control for a number of threats to internal validity. Inspection of Table 5.6 reveals that the static-group comparison design controls for four of the threats to internal validity (i.e., statistical regression, instrumentation, testing, and history) that may result in changes in performance of the experimental group. The remaining threats to internal validity (i.e., maturation, selection, selection by maturation interaction, mortality, resentful demoralization of the control group, diffusion of treatment, compensatory rivalry by the control group, and compensatory equalization) represent possible concerns to the internal validity of the static-group comparison design.

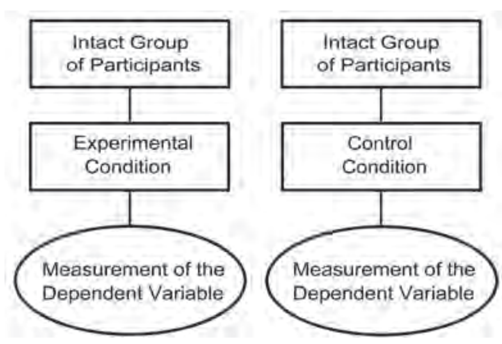


FIGURE 5.4. Static-group comparison design.

TABLE 5.6. Threats to Internal Validity Associated with Quasi-Experimental Designs

	Research design			
Threat	Static group comparison	Nonequivalent control group	Counterbalanced	Time-series
1. Maturation	Possible concern	Controlled	Controlled	Controlled
2. Selection	Possible concern	Controlled	Controlled	Controlled
3. Selection by maturation interaction	Possible concern	Controlled	Possible concern	Controlled
4. Statistical regression	Controlled	Possible concern	Controlled	Controlled
5. Mortality	Possible concern	Controlled	Controlled	Controlled
6. Instrumentation	Controlled	Controlled	Controlled	Controlled
7. Testing	Controlled	Controlled	Controlled	Controlled
8. History	Controlled	Controlled	Controlled	Controlled
9. Resentful demoralization of the control group	Possible concern[a]	Possible concern[a]	Controlled	Not applicable
10. Diffusion of treatment	Possible concern[a]	Possible concern[a]	Controlled	Not applicable
11. Compensatory rivalry by the control group	Possible concern[a]	Possible concern[a]	Controlled	Not applicable
12. Compensatory equalization	Possible concern[a]	Possible concern[a]	Controlled	Not applicable

Note. This table is meant only as a general guideline. Decisions with regard to threats to internal validity must be made after the specifics of an investigation are known and understood. Thus, interpretations of internal validity threats must be made on a study-by-study basis.

[a]These threats to internal validity are controlled if the control group received an equally desirable but alternative intervention.

External Validity. Inspection of Table 5.7 reveals that the threats to external validity associated with the population (i.e., generalization across participants and interaction of personological variables and treatment effects) are a concern, because participants were not randomly selected from a specified population. Novelty and disruption effects, as well as the Hawthorne effect, may be threats to the external validity of the study if the researchers did not provide an equally desirable but alternative intervention. Posttest sensitization also may be a threat to external validity if the posttest has a powerful effect on the independent variable. The remaining threats to the external validity of a study are dependent on the particular characteristics of the study, and on how well the study was conducted.

Recall the research question regarding quality of life for adults with and without disabilities? Read on to see how your design compares to how the actual study was designed.

Research Example 5: Differences in Quality of Life

A static-group comparison design was used to investigate whether differences exist in dimensions related to *quality of life* for adults with developmental disabilities compared to adults in the general population (Sheppard-Jones, Prout, & Kleinart, 2005). *Quality of life* has been defined as "general feelings of well-being, feelings of positive social involvement, and opportunities to

Quasi-experimental designs *are similar to true experimental designs except that participants are neither selected from the specified population nor randomly assigned to groups.*

The **static-group comparison design** *is the same as the posttest-only control-group design described earlier, except for the absence of the random selection of participants from a population and random assignment of participants to groups.*

TABLE 5.7. Threats to External Validity Associated with Quasi-Experimental Designs

	Research design			
Threat	Static group comparison	Nonequivalent control group	Counterbalanced	Time-series
1. Generalization across participants	Concern	Concern	Concern[b]	Concern
2. Interaction of personological variables and treatment effects	Concern	Concern	Concern[b]	Concern
3. Verification of independent variable	Possible concern	Possible concern	Possible concern	Possible concern
4. Multiple treatment interference	Controlled	Controlled	Controlled	Controlled
5. Novelty and disruption effects	Controlled[a]	Controlled[a]	Possible concern	Controlled
6. Hawthorne effect	Controlled[a]	Controlled[a]	Possible concern	Controlled
7. Experimenter effects	Possible concern	Possible concern	Possible concern	Controlled
8. Pretest sensitization	Controlled	Possible concern	Controlled	Controlled
9. Posttest sensitization	Possible concern	Controlled	Controlled	Controlled
10. Interaction of time of measurement and treatment effects	Possible concern	Possible concern	Possible concern	Possible concern
11. Measurement of the dependent variable	Possible concern	Possible concern	Possible concern	Possible concern
12. Interaction of history and treatment effects	Possible concern	Possible concern	Possible concern	Possible concern

Note. This table is meant only as a general guideline. Decisions with regard to threats to external validity must be made after the specifics of an investigation are known and understood. Thus, interpretations of external validity threats must be made on a study-by-study basis.

[a]These threats to external validity are controlled if the control group received an equally desirable but alternative intervention.

[b]These threats to external validity are not a concern if the researcher has randomly selected the participants from a specified population.

achieve personal potential" (Shalock et al., 2002, p. 458). Two groups were compared: (1) consumers, that is, individuals with intellectual or other developmental disabilities (e.g., autism spectrum disorder) and (2) the general population. The consumer group comprised 502 randomly selected adults with disabilities receiving state-funded services. Consumers had a mean age of 40 years, were largely single (95%), and lived in varied environments (40% in group homes, 40% with family, 15% in institutions or nursing homes, 5% other). The general population group comprised 576 adults with a mean age of 45, who were largely married (60%) and living on their own (99%). Groups were similar in terms of gender and race. The measurement instrument, called the Core Indicators Consumer Survey (Human Services Research Institute, 2001), was completed in a face-to-face or telephone interview. Items on the survey related to relationships, safety, health, choice-making opportunities, community participation, well-being and satisfaction, and rights. Individuals responded to survey items using a 3-point Likert-type scale or by responding "yes" or "no." Individuals in the consumer group were interviewed directly, although in 194 cases (38.6%), proxies such as legal guardians responded on behalf of the consumers because of limitations in communication. Individuals in the general population group, interviewed via telephone, were selected randomly in a random-digit dialing procedure. The response rate for the telephone survey was 36.4%. Researchers used MANOVA to analyze survey data. Numerous variables were statistically significant. Results indicated that consumer group members were more lonely, desirous of more work hours, and afraid at home in their neighborhood. Consumers were less likely to have a choice as to where and with whom they lived or to exercise choice in free-time activities. They had fewer options from which to choose in terms of activities, transportation, and place of residence.

Nonequivalent Control-Group Design

The **nonequivalent control-group design** begins with the identification of naturally assembled experimental and control groups. Again, the naturally occurring experimental and control groups should be as similar as possible, and the assignment to one group or the other is assumed to be random. Measurement of the dependent variable is taken prior to the introduction of the independent variable. The independent variable is then introduced, followed by the postintervention measurement of the dependent variable. Figure 5.5 depicts the form of the nonequivalent control-group design.

Analysis of Data

The data from a nonequivalent control-group design are analyzed using ANCOVA, because the primary threat to the internal validity of the nonequivalent control-group design is the possibility that differences on the posttest scores of the experimental and control groups are the result of initial differences rather than the effects of the independent variable. ANCOVA statistically equates initial differences between the experimental and control groups by adjusting the posttest means of the groups.

Internal Validity. Although the nonequivalent control-group design does not provide the same level of experimental control as the pretest–posttest control-group design, it enables researchers to address many of the threats to internal validity adequately. The effectiveness of the nonequivalent control-group design in addressing the threats to internal validity increases with the similarity of the pretest scores of the experimental and control groups. Inspection of Table 5.6 reveals that the nonequivalent control-group design controls for seven of the threats to internal validity (i.e., maturation, selection, selection by maturation interaction, mortality, instrumentation, testing, and history) that result in changes in the performance of the experimental groups. The nonequivalent control-group design does not control for statistical regression that can result in changes in the performance of the experimental group. The four threats to internal validity (i.e., experimental treatment diffusion, compensatory rivalry by the control group, compensatory equalization of treatments, and resentful demoralization of the control group) that result in changes in the performance of the control group are controlled if researchers provide the control group an equally desirable but alternative intervention.

External Validity. As with the static-group comparison design, the threats to external validity associated with the population (i.e., generalization across participants and interaction of personological variables and treatment effects) are a concern, because participants were not randomly selected from a specified population (see Table 5.7). Novelty and disruption effects and the Hawthorne effect may be threats to the external validity of the study if the researchers did not provide an equally desirable but alternative intervention. Pretest and posttest sensitization also may be threats to the external validity if the pretest and posttest have a powerful effect on the intervention. The remaining threats to the external validity of a study are dependent on the particular characteristics of the study.

The **nonequivalent control-group design** *is similar to the pretest–posttest control-group design described previously, except for the absence of the random selection of participants from a population and the random assignment of participants to groups.*

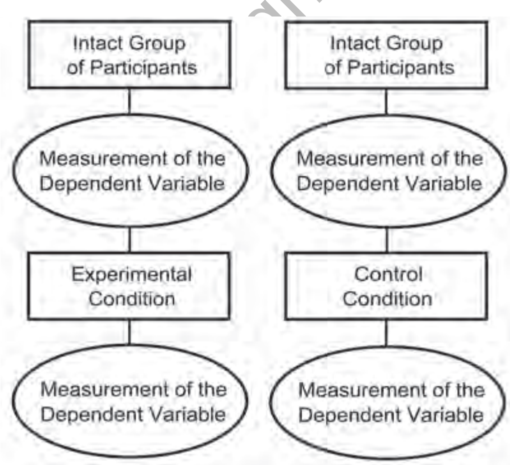


FIGURE 5.5. Nonequivalent control-group design.

Counterbalanced Designs

Counterbalanced designs encompass a wide range of designs in which independent variables are introduced to all participants. Counterbalanced designs are useful when researchers are interested in studying multiple variables. Counterbalanced designs are also useful when it is not possible to assign participants randomly to the experimental and control groups, and when participant attrition may be a problem. The level of experimental control achieved with a counterbalanced design is greater than that of the nonequivalent control-group design, because each participant serves as his/her own control. Counterbalanced designs are also referred to as "rotation experiments," crossover designs, and "switchover designs."

In its simplest form, the counterbalanced design begins with the identification of two naturally assembled groups. Measurement of the dependent variable is taken prior to the introduction of the independent variable. The independent variable is introduced to one of the groups, followed by postintervention measurement of the dependent variable. The independent variable is then introduced to the other group followed by the postintervention measurement of the dependent variable. Figure 5.6 depicts the form of the two-group counterbalanced design using naturally assembled groups. (Researchers may randomly assign participants to groups.)

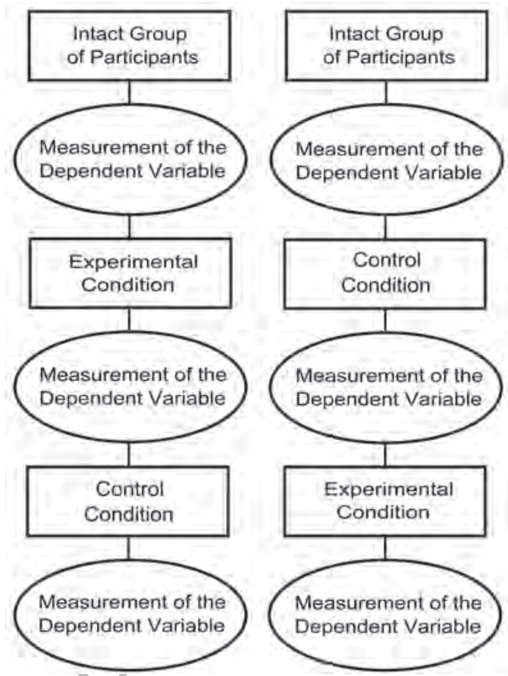


FIGURE 5.6. Two-group counterbalanced design.

Analysis of Data

Although data from counterbalanced designs can be analyzed with a number of statistical tests, a repeated measures ANOVA is most commonly used. The advantage of any repeated measurement analysis procedure is that it controls for differences between participants. That is, repeated measurement analyses eliminate differences between participants from the experimental error. Additionally, a nonparametric measure such as the Kruskal–Wallis test should be used if the data violate the assumptions underlying parametric tests (i.e., homogeneity of variance, normal distribution of data, and interval or ratio scale data).

Internal Validity. The counterbalanced design provides a higher level of experimental control than does the nonequivalent control-group design. This greater degree of experimental control is achieved because the counterbalancing of the independent variable enables researchers to compare participants' performance in groups rather than in experimental and control groups; thus, counterbalanced designs eliminate the need for groups to be equivalent.

Inspection of Table 5.6 reveals that counterbalanced designs effectively deal with seven of the threats to internal validity (i.e., maturation, selection, statistical regression, mortality, instrumentation, testing, and history) that result in changes in performance of the experimental group. Counterbalanced designs do not effectively control for selection by maturation interaction effects, because there is no nonintervention control group. The four threats to internal validity (i.e., experimental treatment diffusion, compensatory rivalry by the control group, compensatory equalization of treatments, and resentful demoralization by the control group) that result in changes in performance of the control group are controlled for if researchers ensure that there is no treatment diffusion.

External Validity. The threats to external validity associated with the population (i.e., generalization across participants and interaction of personological variables and treatment effects) are not a concern when participants have been randomly selected from a specified population (see Table 5.7). If this is not the case, then these threats to external validity are of concern. Novelty and disruption, the Hawthorne effect, and experimenter effects may be evident because of the ongoing implementation of different independent variables. Pretest and posttest sensitization are unlikely threats to the external validity of a study because of the repeated measurement of the dependent variable. The remaining threats to the external validity of a study are dependent on the particular characteristics of the study and how well it was conducted.

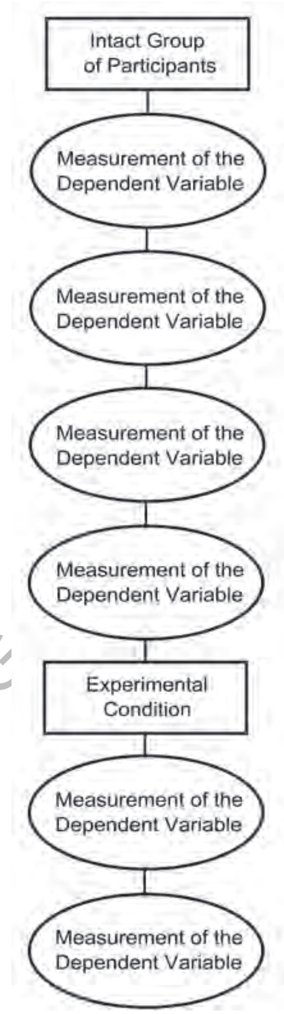


FIGURE 5.7. Time-series design.

Time-Series Designs

Any effects attributable to the independent variable are indicated by discontinuity in the preintervention and postintervention series of scores. **Time-series designs** differ from single-case designs (see Chapters 11 to 13) in that the unit of analysis is a group of individuals, not single participants. Additionally, time-series designs can be used in an ex post facto or experimental fashion. It is important to note that a variety of time-series designs involve different numbers of groups, independent variables, and so on.

A time-series design begins with the identification of a naturally assembled group. Measurement of the dependent variable occurs a number of times prior to the introduction of the independent variable, which is followed by measurement of the dependent variable a number of more times. Figure 5.7 depicts the form of time-series designs.

In **counterbalanced designs**, *two or more groups get the same independent variables, but the independent variables are introduced in different orders.*

Time-series designs *are quasi-experimental designs involving a series of repeated measurements of a group of research participants.*

Analysis of Data

Data from time-series designs can be analyzed in a variety of ways. Researchers may graph the scores and look for changes in the preintervention and postintervention pattern of scores. Each score is plotted separately on the graph, and scores are connected by a line. A vertical line is inserted into the series of scores at the point the independent variable is introduced. Researchers then use the graphed data to compare the level, slope, and variation in the preintervention and postintervention scores. Visual analysis methods for assessing intervention effects are described more completely in our presentation of single-case designs (see Chapters 11 to 13). Statistical methods used with data from time-series designs can range from multiple regression to log linear analysis procedures. These statistical analysis procedures are all aimed at determining whether the preintervention and postintervention patterns of scores differ statistically from one another. We direct the reader to Box, Hunter, and Hunter (2005) and to Glass, Wilson, and Gottman (2008) for complete descriptions of statistical procedures for analyzing data from time-series studies.

Internal Validity. Time-series designs provide a high degree of experimental control even though they do not employ a control group. Inspection of Table 5.6 reveals that time-series designs control for the eight threats to internal validity (i.e., maturation, selection, selection by maturation interaction, statistical regression, mortality, instrumentation, testing, history) that result in changes in performance of the experimental group. Instrumentation might be a concern if the calibration of the measurement device employed by the researcher changed over the course of the study. Of course, it would be unlikely for such a change to occur in direct connection with the introduction of the independent variable. The four threats to internal validity that result in changes in the performance of the control group are not applicable with time-series designs.

External Validity. As with all quasi-experimental designs, the threats to external validity associated with the population (i.e., generalization across participants and interaction of personological variables and treatment effects) are a concern, because participants have not been randomly selected from a specified population (see Table 5.7). Novelty and disruption, the Hawthorne effect, and experimenter effects tend not to be a problem because of the ongoing measurement of the dependent variable. Additionally, pretest and posttest sensitization are unlikely threats to the external validity of a study because of the repeated measurement of the dependent variable. The remaining threats to the external validity of a study are dependent on the particular characteristics of the study and on how well the study was conducted.

WHAT ARE PREEXPERIMENTAL DESIGNS?

Preexperimental designs primarily differ from true experimental designs and two of the quasi-experimental designs (i.e., counterbalanced and time-series designs) in that they do not include a control group. The lack of a control group essentially eliminates researchers' ability to control for any of the threats to internal validity. The lack of control over possible extraneous variables that may cause changes in the dependent variable renders preexperimental designs almost useless to furthering knowledge in education and psychology. In short, researchers cannot ensure that a change in the value of the independent variable is accompanied by a change in the value of the dependent variable, which is one of the three key requirements for asserting that a causal relationship exists between two variables.

Nevertheless, we detail preexperimental designs because they are used extensively in educational and psychological research despite these obvious problems. It is important for critical research consumers to understand the weaknesses associated with preexperimental designs.

WHAT ARE THE TYPES OF PREEXPERIMENTAL DESIGNS?

We consider two preexperimental designs—one-shot case study and one-group pretest–posttest design. Both designs are presented in terms of a single independent variable and dependent variable.

One-Shot Case Study

In the one-shot case study, an independent variable is introduced to a group of participants. The one-shot case study begins with the identification of a naturally assembled group. The independent variable is then administered, followed by the measurement of the dependent variable. Figure 5.8 depicts the form of the one-shot case study.

Analysis of Data

The data from a one-shot case study may be analyzed with the one-sample *t*-test if there is a specified population mean. In essence, the obtained mean of the study group is compared to the specified population mean. The results of the one-sample *t*-test indicates whether the obtained mean of the study group differs from that of the specified population mean.

Internal Validity. Researchers utilizing one-shot case studies often collect extensive information or use standardized measures in an effort to document the effects of the independent variable. Although such efforts may seem a reasonable replacement for the experimental control associated with the true experimental designs, they do not lead to any definitive assessment of the potential extraneous variables that may have resulted in changes in the scores of individuals.

Inspection of Table 5.8 reveals that the one-shot case study has low internal validity. Four of the eight threats to internal validity (i.e., history, maturation, selection, mortality) that may result in changes in performance of the experimental group are concerns. The remaining four threats to internal validity (i.e., selection by maturation interaction, instrumentation, testing, and statistical regression) that may result in changes in performance of the experimental group and the four threats to internal validity (i.e., experimental treatment diffusion, compensatory rivalry by the control group, compensatory equalization of treatments, and resentful demoralization of the control group) that result in changes in performance of the control group are not applicable.

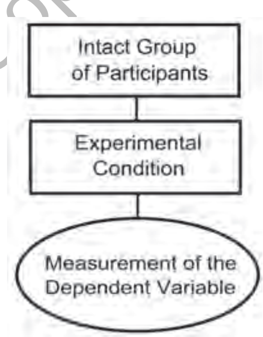


FIGURE 5.8. One-shot case study.

External Validity. The one-shot case study does not control for most of the threats to external validity (see Table 5.9). Pretest sensitization is the only threat to external validity that can be controlled. Researchers who are limited to studying one group of participants should include a pretest in their experimental design (this design is described in what follows).

One-Group Pretest–Posttest Design

The one-group pretest–posttest design differs from the one-shot case study in that a pretest measure is administered prior to introduction of the independent variable. Effects of the independent variable are determined by comparing the pretest and posttest scores of the group of participants. The one-group pretest–posttest design begins with the identification of a naturally assembled group. Measurement of the dependent variable occurs prior to the introduction of the independent variable. The independent variable is then introduced followed by measurement of the dependent variable. Figure 5.9 depicts the form of the one-group pretest–posttest design.

Analysis of Data

Data from a one-group pretest–posttest design may be analyzed with a correlated *t*-test. A paired set of variables is the minimum requirement for a correlated *t*-test. The pretest scores are compared to the posttest scores. The results of the correlated *t*-test indicate whether the

> **Preexperimental designs** *do not include a control group.*

TABLE 5.8. Threats to Internal Validity Associated with Preexperimental Designs

	Research design	
Threat	One-shot case study	One-group pretest–posttest
1. Maturation	Concern	Concern
2. Selection	Concern	Controlled
3. Selection by maturation interaction	Not applicable	Concern
4. Statistical regression	Not applicable	Concern
5. Mortality	Concern	Controlled
6. Instrumentation	Not applicable	Concern
7. Testing	Not applicable	Concern
8. History	Concern	Concern
9. Resentful demoralization of the control group	Not applicable	Not applicable
10. Diffusion of treatment	Not applicable	Not applicable
11. Compensatory rivalry by the control group	Not applicable	Not applicable
12. Compensatory equalization	Not applicable	Not applicable

Note. This table is meant only as a general guideline. Decisions with regard to threats to internal validity must be made after the specifics of an investigation are known and understood. Thus, interpretations of internal validity threats must be made on a study-by-study basis.

TABLE 5.9. Threats to External Validity Associated with Preexperimental Designs

	Research design	
Threat	One-shot case study	One-group pretest–posttest
1. Generalization across participants	Possible concern	Possible concern
2. Interaction of personological variables and treatment effects	Possible concern	Possible concern
3. Verification of independent variable	Possible concern	Possible concern
4. Multiple treatment interference	Not applicable	Not applicable
5. Novelty and disruption effects	Possible concern	Possible concern
6. Hawthorne effect	Possible concern	Possible concern
7. Experimenter effects	Possible concern	Possible concern
8. Pretest sensitization	Not applicable	Possible concern
9. Posttest sensitization	Possible concern	Possible concern
10. Interaction of time of measurement and treatment effects	Possible concern	Possible concern
11. Measurement of the dependent variable	Possible concern	Possible concern
12. Interaction of history and treatment effects	Possible concern	Possible concern

Note. This table is meant only as a general guideline. Decisions with regard to threats to internal validity must be made after the specifics of an investigation are known and understood. Thus, interpretations of internal validity threats must be made on a study-by-study basis.

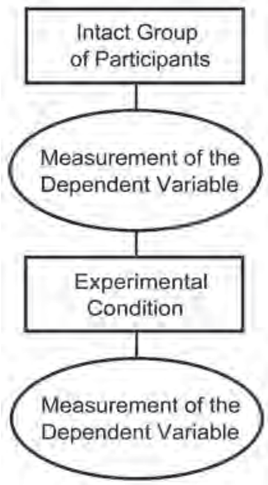


FIGURE 5.9. One-group pretest–posttest design.

pretest and posttest means of the study group differ from one another. Additionally, a nonparametric test such as the Wilcoxon signed rank test should be used if the data violate the assumptions underlying these parametric tests (i.e., homogeneity of variance, normal distribution of data, and interval or ratio scale data).

Internal Validity. Although the use of a pretest in the one-group pretest–posttest design renders it better than the one-shot case study, it still does not provide much experimental control. Inspection of Table 5.8 reveals that the one-group pretest–posttest design does not control for six of the eight threats to internal validity (i.e., history, maturation, testing, instrumentation, statistical regression, and selection by maturation interaction) that may result in changes in performance of the experimental group. The one-group pretest–posttest design does control for selection and mortality that may result in changes in the dependent variable. As with the one-shot case study, the remaining four threats to internal validity (i.e., experimental treatment diffusion, compensatory rivalry by the control group, compensatory equalization of treatments, and resentful demoralization of the control group) that result in changes in performance of the control group are not applicable, because there is no control group.

External Validity. As with the one-shot case study, researchers often provide a great deal of detail in an effort to document the effects of the dependent variable. These efforts should not alleviate concerns regarding the external validity of the one-group pretest–posttest design. Inspection of Table 5.9 shows that the one-group pretest–posttest design does not control for any of the threats to external validity.

WHEN SHOULD RESEARCHERS USE EACH EXPERIMENTAL RESEARCH DESIGN?

Experimental research designs are used to establish causal relationships. It should be clear by now that there is a clear difference in the extent to which true experimental, quasi-experimental, and preexperimental research designs allow researchers to assert with confidence that there is a causal relationship between variables. True experimental research designs provide the highest level of confidence and are the designs of choice to establish a causal relationship between variables. The internal and external validity of true experimental designs are high, because these research designs rely on the random selection of participants from a population and the random assignment of participants to experimental and control groups. The random selection of participants from a population increases the external validity of the study, and the random assignment of participants to experimental and control groups increases the internal validity of the study. Establishing the initial equivalence of the experimental and control groups enables the researcher to conclude confidently that any statistically significant differences are due to the independent variables. Thus, researchers who are attempting to demonstrate a cause-and-effect relationship should use a true experimental design if they can randomly select and assign participants.

Although quasi-experimental research designs do not provide the same level of control as true experimental research designs, they control reasonably well for threats to the internal and external validity of studies. Quasi-experimental research designs are extremely useful, because they enable researchers to conduct representative research—that is, research that replicates "real-world" conditions. Quasi-experimental research designs should be employed when it is critical for researchers to conduct a representative study. Quasi-experimental research designs may also be employed when it is impossible for researchers to select participants randomly from a population or randomly assign participants to the experimental and control groups.

The preexperimental research designs have a lack of experimental control that essentially renders them useless for establishing causal relationships between variables. Indeed, we have detailed preexperimental research designs only because they continue to be used inappropriately by researchers to infer causal relationships between variables. In contrast to true experimental and quasi-experimental research designs, preexperimental research designs do not enable reasonable comparisons to be made. Thus, any causal claims made by researchers are clearly inappropriate. Researchers should use preexperimental designs as a last resort; that is, when they are not able to use true experimental or quasi-experimental designs due to limitations (e.g., limited financial or human resources, teacher, administrative, or parental concerns, scheduling difficulties).

SUMMARY

❖ Establishing cause-and-effect relationships among variables is typically the goal of most research. There are three categories of experimental research designs used by researchers in education and psychology: (1) true experimental, (2) quasi-experimental, and (3) preexperimental designs.

❖ The ability of an experimental design to control for threats to internal and external validity is primarily dependent on four key features: (1) the procedures used to select participants from a broader population; (2) the use of a control group or condition; (3) the initial equivalence of the experimental and control groups; and (4) how effectively the investigation was conducted.

❖ True experimental designs are the only experimental designs that can result in relatively definitive statements about causal relationships between variables. True experimental designs ensure that (1) a change in the value of the independent variable is accompanied by a change in the value of the dependent variable, (2) how the independent variable affects the dependent variable is established a priori, and (3) the independent variable precedes the dependent variable.

❖ There are three basic requirements for a research design to be considered a true experimental design: (1) random selection of participants from a population to form a sample; (2) random assignment of research participants to the experimental and control conditions; and (3) equal treatment of members of the experimental and control groups.

❖ Three of the most common true experimental designs include (1) pretest–posttest control-group, (2) posttest-only control-group, and (3) Solomon four-group designs.

❖ The pretest–posttest control-group design has the following five attributes: (1) random selection of participants from a population to form a sample; (2) random assignment of selected participants to the experimental or control groups; (3) measurement of the dependent variable prior to the introduction of the independent variable; (4) introduction of the independent variable; and (5) postintervention measurement of the independent variable. Any differences observed at the end of the study are assumed to be due to the independent variable.

❖ A posttest-only control-group design is a true experimental design similar to the pretest–posttest-only control-group design, with the exception that pretests of the dependent variable are not administered to the experimental and control groups. This design has the following four attributes: (1) random selection of participants from a population to form a sample; (2) random assignment of the selected participants to the experimental or control groups; (3) introduction of the independent variable; and (4) postintervention measurement of the dependent variable. Any differences observed at the end of the study are assumed to be due to the independent variable.

❖ The Solomon four-group design is a true experimental design that essentially combines the pretest–posttest control-group design and the posttest-only control-group design; two groups serve as experimental groups and two groups serve as control groups.

Many of the threats to internal validity are controlled in true experimental designs, while many of the threats to external validity remain. The Solomon four-group design not only controls for testing (threat to internal validity) and pretest sensitization (threat to external validity) but also enables researchers to assess their effects on the intervention outcomes.

❖ A factorial experimental design is used to assess the effects of two or more independent variables or the interaction of participant characteristics with the independent variable. It is used to determine whether a particular variable has the same effect when studied concurrently with other variables as it would when studied in isolation. The threats to internal validity controlled for by a factorial design are dependent on the basic underlying true experimental design. One of the greatest strengths of a factorial design is that it enhances the external validity of the study.

❖ In quasi-experimental designs, participants are neither randomly selected from the specified population nor randomly assigned to experimental and control groups. These designs differ from true experimental designs in two ways: (1) Participants are not randomly selected from a specified population, and (b) participants are not randomly assigned to experimental and control groups.

The static-group comparison design is the same as the posttest-only control group design except for the absence of random selection of participants from a population and random assignment of participants to groups.

The nonequivalent control-group design is similar to the pretest–posttest control-group design described previously except for the absence of random selection of participants from a population and random assignment of participants to groups.

Counterbalanced designs encompass a wide range of designs in which independent variables are introduced to all participants. In counterbalanced designs, two or more groups get the same independent variables, but the independent variables are introduced in different orders. Counterbalanced designs are also useful when it is not possible to assign participants randomly to the experimental and control groups, and when participant attrition may be a problem.

Time-series designs are quasi-experimental designs involving a series of repeated measurements of a single group of research participants in the following three steps: (1) Measure the dependent variable a number of times prior to the introduction of the independent variable; (2) introduce the independent variable; and (3) measure the dependent variable a number of more times.

In the absence of random assignment of participants to groups, the possibility of pretreatment differences is present. The lack of a pretest (as in the static group comparison) greatly weakens the ability to control for a number of threats to internal validity. However, the level of experimental control achieved with a counterbalanced design is greater than that achieved with other designs, because each participant serves as his/her own control, and time-series designs provide a high degree of experimental control even though they do not employ a control group. External validity threats are present with these designs, just as they were in the true experimental designs.

❖ Preexperimental designs primarily differ from true experimental designs and two of the quasi-experimental designs (i.e., counterbalanced and time-series designs) in that they do not include a control group. The lack of a control group essentially eliminates researchers' ability to control for any of the threats to internal validity.

The one-shot case study involves the following three steps: (1) identification of a naturally assembled group, (2) administration of the independent variable, and (c) measurement of the dependent variable. The one-group pretest–posttest design differs from the one-shot case study in that a pretest measure is administered prior to the introduction of the independent variable. The effects of the independent variable are determined by comparing the pretest and posttest scores of the group of participants. Preexperimental designs control for threats to neither internal validity nor external validity. Conclusions regarding the results of investigations using these designs must be made with caution.

❖ Researchers who are attempting to demonstrate a cause-and-effect relationship should use a true experimental design if they can randomly select and assign participants. Quasi-experimental research designs should be employed when it is critical for researchers to conduct a representative study, or when it is impossible for researchers to select participants randomly from a population or randomly assign participants to the experimental and control groups. Researchers should use preexperimental designs as a last resort; that is, when researchers are not able to use true experimental or quasi-experimental designs due to limitations.

DISCUSSION QUESTIONS

1. What is meant by a true experimental design? (In your response, discuss the basic design features of true experimental designs.)
2. What are quasi-experimental designs? How are true experimental and quasi-experimental designs similar, and how are they different?
3. Can quasi-experimental designs determine cause-and-effect relationships? Explain.
4. How are threats to the internal validity of the experiment controlled with a quasi-experimental design?
5. How do true experimental designs control for threats to external validity? What are some concerns you might have?
6. Are quasi-experimental designs more or less useful than true experimental designs in determining the external validity of the investigation? Why or why not?
7. What type of design is recommended if you are concerned that males and females have different attitudes toward math and may perform differently in math? Why?
8. What type of design have you used when you pretest a single class, provide the instruction, then posttest the class? What are the problems with this design? What conclusions could you make in terms of internal and external validity?
9. For the example in Question 8, what would be a better design for you to use? How would you do this? How would your conclusions differ?
10. Of the designs discussed in this chapter, which design(s) seem to have the strongest control over threats to internal validity? Which design(s) do you think would be the most difficult to implement in an applied setting? Why?

Note. Beginning in this chapter, we offer an illustrative example of research associated with the designs described in the readings. The example is followed by questions about the research. Finally, there are references to three additional research studies associated with the designs in the readings.

ILLUSTRATIVE EXAMPLE

The Effects of Learning Strategy Instruction on the Completion of Job Applications by Students with Learning Disabilities

J. Ron Nelson, Deborah J. Smith, and John M. Dodd

The purpose of this study was to assess the effects of learning strategy instruction on the completion of job applications by students identified as learning disabled. Thirty-three students (average age 15 years 6 months) were randomly assigned by grade and gender to one of two experimental conditions: learning strategy instruction or traditional instruction. The result was 16 students (10 boys and 6 girls) being placed under the learning strategy instruction condition and 17 students (10 boys and 7 girls) being placed under the traditional instruction condition. Results indicated that in addition to statistically significant lower numbers of information omissions and information location errors, holistic ratings of the overall neatness of the job applications were significantly higher for those students under the learning strategy instruction condition. In addition to these positive changes in the performance measures, social validity data suggest that students under the learning strategy condition would be more likely to receive an invitation for a job interview. The findings and future research needs are discussed.

Because employers often receive numerous applications for a single advertised position, the quality of the employment application materials has a direct effect on an individual's ability to secure employment. Employers most often use the employment application form and, when applicable, the personal resume to decide whom to interview for a position. Regardless of otherwise equal qualifications, the content, completeness, and neatness of these employment application materials have an effect on whether an individual is given the opportunity to interview for a specific job (Field & Holley, 1976). Indeed, the skills involved in completing employment application materials may be the foundation upon which other job-finding skills, such as interviewing, are built (Azrin & Philip, 1979; Mathews, Whang, & Fawcett, 1981).

Despite the importance of a complete and accurate employment application, researchers have mostly focused on the effects of training procedures on an individual's job interview performance (e.g., Furman, Geller, Simon, & Kelly, 1979; Hall, Sheldon-Wildgen, & Sherman, 1980; Hollansworth, Dressel, & Stevens, 1977); relatively few studies have been conducted on teaching disadvantaged individuals (Clark, Boyd, & MaCrae, 1975) or those identified as learning disabled (Mathews & Fawcett, 1984) to complete job application forms. Approaches employed in these studies were applied behavioral instruction techniques designed to teach the skills involved in completing employment applications. Mathews and Fawcett, for example, taught three high school seniors identified as having learning disabilities (LD) to complete a job application and write a resume. The training sequence involved the student, with assistance from the experimenter, reading a set of instructional materials containing detailed written specifications for the behaviors, including a rationale and examples for each task. The student then practiced each task with feedback from the experimenter. Following training, each of the three students showed significant changes in the percentage of application items completed accurately.

Although the results of this study demonstrate that students with LD can be taught to more accurately complete job application forms, the students were reported to need assistance with reading the procedural text accompanying the employment application materials. More important, procedures required 2.5 hours of individual instructional time. This would suggest that an effective group training procedure on employment application skills should be designed to facilitate both the acquisition and the expression of information. Students must be taught to understand the procedural text included on job application forms and to provide all of the requested information accurately.

Learning strategy instruction is one instructional approach that might be used to teach students these two related skills that are necessary to independently complete job applications. For the purposes of this article, a learning strategy will be defined as a collection of specific skills that one uses in a particular situation to facilitate the acquisition or expression of knowledge or skills. This type of instruction appears to be especially beneficial for students with LD because these students have been characterized as lacking active task engagement and persistence (Harris, 1986) and as lacking the skills necessary to execute and monitor the cognitive processes central to academic success (Baumann, 1986).

Although, to date, there appears to have been no empirical work conducted with learning strategies designed to facilitate both the understanding of procedural text that is included on a job application and the expression of the requested information, the present study can be placed in the context of recent work on learning strategy instruction in reading (e.g., Borkowski, Weyhing, & Carr, 1988) and writing (e.g., Englert, Raphael, Anderson, Anthony, & Stevens, 1991). Research on reading strategy instruction has focused on how to teach students to (a) determine the main idea (Baumann, 1986; Cunningham & Moore, 1986; Williams, 1986); (b) summarize the information contained in text (Day, 1980; Hare & Borchardt, 1984; Nelson, Smith, & Dodd, 1992; Palincsar & Brown, 1984; Taylor, 1982; Taylor & Beach, 1984; Taylor & Berkowitz, 1980); (c) draw inferences about what they have read (Hansen, 1981; Pearson, 1985; Raphael & McKinney, 1983; Raphael & Pearson, 1985; Raphael & Wonnacott, 1985); (d) generate questions about what they have read (Andre & Anderson, 1978–79; Brown & Palincsar, 1985); and (e) monitor their comprehension of the text (Baker & Anderson, 1982; Vosniadou, Pearson, & Rogers, 1988). In sum, this work has shown that students with LD, as well as other students, can be taught strategies to facilitate their reading and understanding of literary and expository text.

However, following written directions such as those included on job applications differs from other kinds of reading, in that the goal of the reader is to *do* something rather than to *learn about* something. In the case of procedural text or written directions, a partial understanding is insufficient—mastery of the content is required. Reading written directions is further complicated by the fact that writers of directions often overestimate the reader's experience with directions, omit intermediate steps, use technical vocabulary, and employ complex syntax (Henk & Helfeldt, 1987). Furthermore, written directions often contain unclear directional and location cues for entering information in specific places on the application form. In addition, according to Henk and Helfeldt, there is no immediate transfer of academic reading skills to following written directions. Good readers follow written directions well only 80% of the time, and poor readers achieve less than a 50% success rate (Fox & Siedow, 1980).

Paralleling work on reading comprehension, researchers have developed a number of learning strategies designed to facilitate students' abilities to generate expository text (e.g., Englert et al., 1991; Graham & Harris, 1989; Harris & Graham, 1985; Schmidt, Deshler, Schumaker, & Alley, 1988). Schmidt et al., for example, taught high school students with LD four written expression learning strategies: sentence writing, paragraph writing, error monitoring, and theme writing. The results showed improvements both in the quality of themes and in the mechanics of the written text.

Because research and theory suggest that students should be taught to apply different learning strategies to different types of situations (Brandt, 1989), students must possess specific strategies for both understanding and following procedural text or written directions. In other words, the reading and writing strategies demanded by the task requirements of a job application depart from those that students apply elsewhere. The purpose of the present study was to develop a learning strategy and study its effects on completion of job applications by students with LD. This is important because there are significant societal and personal costs associated with the unemployment and underemployment of individuals with disabilities, and, as noted, regardless of otherwise equal qualifications, the content, completeness, and neatness of a job application can determine whether an individual has an opportunity to even interview for a specific job.

METHOD

Subjects

Thirty-three students (20 boys and 13 girls) with LD served as participants in the study. All were receiving special education services in a public high school in a city in the Northwest (population 180,000) and were classified as learning disabled by a school district multidisciplinary evaluation team. Criteria for special education classification include deficits in oral expression (as measured by the Northwestern Syntax Screening Test), listening comprehension (as measured by the Carrow Test for Auditory Comprehension of Language), and/or written expression (as measured by the Comprehensive Tests of Basic Skills). Criteria also included a significant discrepancy (at least 2 years below grade placement) between the student's estimated ability and academic performance.

Students were generally from low-SES families (qualified for free and reduced lunch). Table 1 provides additional descriptions of the participants' sex, age, race, grade level, years in special education, percentage of each school day spent in special education, IQ, and achievement.

TABLE 1. Subject Description

	Learning strategy instruction ($n = 16$)	Traditional instruction ($n = 17$)
Gender		
Male	10	10
Female	6	7
Age		
Mean	15.9	16.3
Range	14.5–17.3	14.3–17.5
Race		
White	15	17
African American	1	0
Grade level		
12th	2	3
11th	6	7
10th	6	5
9th	2	2
Years in special education		
Mean	5.9	5.3
Range	4–8	4–9
Percentage of day in special education		
Mode	.50	.50
Range	33–83	33–83
Intelligence[a]		
Mean	98.5	96.2
Range	88–106	84–105
Reading comprehension[b]		
Mean *T*	35.1	33.4
Range	22–43	25–41

[a]Stanford–Binet Intelligence Scale.
[b]Iowa Test of Basic Skills.

Setting

All participants were enrolled in a pre-vocational education class for students with learning disabilities. The class was taught by a certificated special education teacher with 6 years of teaching experience at the high school level. The classroom aide was a high school graduate with 8 years of classroom assistance experience. The teacher conducted the experimental sessions during two 60-minute instructional periods. The classroom was approximately 10 m by 15 m and had 25 individual desks at which the participating students sat during the experimental sessions.

Dependent Measures

Student Performance Measures

Three mutually exclusive measures were employed to assess the effects of the learning strategy instruction on the completion of job applications by students: information omissions, information location errors, and a holistic rating of overall neatness of the job application. An omission was scored when a required item was not completed. A location error was scored when the correct information was entered in the wrong location (e.g., writing the information on the line directly below where the information was to be placed). A 5-point Likert-type scale (1 = *very messy* to 5 = *very neat*) was used to obtain a holistic rating of the overall neatness of the job application.

Interscorer agreement for omissions and location errors was determined by having two scorers independently score all of the job applications. The scorers' records were compared item by item. For omissions, agreement was noted when both scorers had marked a response as not present. Similarly, an agreement was noted when both the scorers marked the location of the information as correct or if both scorers had marked the location of the information as incorrect. Percentage of agreement for each measure was computed by dividing the number of agreements by the number of agreements plus disagreements. The percentage of agreement was 100% in both cases.

Interscorer agreement was computed for the holistic rating by having two raters independently rate all of the job applications. A Pearson product moment correlation was then calculated to estimate the reliability of the ratings. The correlation was .78, $p < .05$.

Social Validity Measure

To assess the social validity of the effects of the training, the supervisor of classified personnel at a local university employing approximately 1,200 classified staff was asked the following: "Based on this job application, if you had a position open, would you invite this person in for an interview?" The rating was completed on a 5-point Likert-type scale (1 = *very unlikely*, 3 = *undecided*, 5 = *very likely*). The supervisor rated each application and was unaware of whether it was completed under the learning strategy or traditional instruction condition.

Design

A pretest–posttest control group design was employed. Students were randomly assigned by age and gender to one of two experimental conditions: learning strategy instruction or traditional instruction. This resulted in 16 students (10 boys and 6 girls) being assigned under the learning strategy instruction condition and 17 students (10 boys and 7 girls) under the traditional instruction condition. The results of a preliminary analysis revealed that there were statistically nonsignificant differences in characteristics (i.e., intelligence, achievement, age, years in special education, and percentage of each school day spent in special education) between the two groups.

Procedure

Job Applications

Job applications for entry-level jobs were obtained from eight local businesses. Two of these job applications were selected for the pretest and posttest; two additional applications were used to conduct the training sessions (demonstration and independent practice). Although these job applications were designed to elicit the same general information, the format (e.g., sequence of information and location cues) differed. The same pretest, posttest, and training job applications were used under the learning strategy instruction and traditional instruction conditions.

Preskill Instructional Module

Students under both conditions (described below) received a prepared instructional module designed to provide the relevant prerequisite vocabulary knowledge necessary to complete a job application. This instruction was conducted, and job application information collected (discussed below), prior to pretesting. The teacher presented the prerequisite vocabulary knowledge module, using a written script, to students under both conditions. The prerequisite information included definitions for the following job application vocabulary words: (a) *birth place*, (b) *nationality*, (c) *previous work experience*, (d) *references*, (e) *maiden name*, (f) *marital status*, (g) *citizenship*, (h) *salary*, and (i) *wage*. Instruction continued until all of the students earned 100% correct on a paper-and-pencil test in which the words were matched with their respective definitions.

Students under both experimental conditions also compiled the information necessary for them to complete a job application, including (a) birth date, (b) social security number, (c) complete address, (d) telephone number, (e) educational experience, (f) previous work experience, (g) references, and (h) felony convictions (if applicable). Students then constructed a job application information card containing this information.

Students under both experimental conditions then completed the pretest job application. The teacher asked them to complete the job application as if they were applying for an actual job. She also explained that typically no one is available to help people complete job applications, and they were to use their job information card for the task. Students were provided as much time as they needed to complete the application. The teacher did not provide the students any assistance during this time. The pretest session was conducted 1 day prior to the training and posttest sessions.

Learning Strategy Instruction Condition

The job application learning strategy taught in this investigation was designed after analyzing the nature of items included on standard job applications for entry-level jobs obtained from a number of local businesses, and after completing a task analysis of the steps involved in completing a job application. The strategy was also designed in accordance with the needs and skill levels of the students. The principle steps were then sequenced and a first-letter mnemonic device was developed to facilitate students' recall of the strategy steps. This resulted in a six-step strategy called "SELECT."

Students first *S*urvey the entire job application and look for the *E*mphasized words that indicate the type of information requested (e.g., previous experience) and think to themselves, "What information do I have to have to complete the job application?" and "Do I have all of the necessary information to complete the application (check job application information card)?" If not, "What additional information do I need to get?" The students then look closely at the items on the job application for *L*ocation cues that indicate where the requested information is to be entered (e.g., line immediately below the request for information) and think to themselves, "Where does the information go?" Next, they think to themselves, "How much space do I need for the information—How big should I print the information?" and then carefully *E*nter the information requested in the appropriate location. After completing the application, the students then *C*heck to see if the information is accurate (compare with job information card) and that the job application is completed, and think to themselves, "Did I put the right information in the right locations?" If not, "I need to complete another job application." Then, "Did I complete the job application?" If not, "Complete the job application." Finally, the students *T*urn the completed job application in to the appropriate individual.

The special education teacher used a five-step procedure to teach the students the job application strategy during an approximately 1-hour instructional session. First, the teacher discussed the goal of the job application strategy instruction procedure (i.e., to help students accurately complete a job application) and why it is important to know how to accurately complete a job application. She also explained how they would be able to use the strategy whenever they applied for a job.

Second, an overhead transparency was used to introduce and discuss the six-step job application strategy. The teacher and students discussed the use of the strategy until it was clear that the students fully understood the steps. This was accomplished through choral responding by the students and informal checks by the teacher.

Third, using an overhead transparency, the teacher modeled the job application strategy by completing a standard job application while "thinking out loud." To actively engage the students, the teacher used prompts to encourage an interactive dialogue with the students throughout the demonstration, for example, "What is it I have to do? I need to . . . " and "How am I doing?" The students were encouraged to help the teacher. After modeling, the teacher and students discussed the importance of using self-questioning statements while completing a job application.

Fourth, students were required to verbally practice the job application strategy steps, including the self-questioning statements, until they were memorized. All of the students were able to do this correctly within a 15- to 20-minute rehearsal period. They were then required to write down the steps and associated self-questioning statements as they worked through a job application. Students were provided only one practice attempt. They were allowed to ask any questions at this time and the teacher provided corrective feedback only upon demand by the students throughout the training session.

Finally, students independently completed the posttest job application. As under the pretest condition, the teacher asked the students to complete the job application as if they were applying for an actual job. She also explained to the students that because there typically is no one there to help them complete job applications, they were to use only their job information card to complete the job application, and that they had as much time as they needed to complete the application. The teacher did not provide the students any assistance during this time. After they completed the posttest job application, the students were asked to independently describe the steps they had used, in an attempt to check whether they had employed the learning strategy. All of the students verbally stated, in sequence, the steps and associated self-questioning statements included in the learning strategy.

Traditional Instruction Condition

The same job application forms used under the learning strategy condition were used for the traditional instruction condition. During an approximately 1-hour instructional session, the special education teacher (same teacher) first discussed the goal of the job application instruction (i.e., to help students accurately complete a job application) and why it is important to know how to accurately complete a job application. She also explained how they would be able to use the things they learned whenever they applied for a job.

Next, the teacher used an overhead transparency to model how to complete a standard job application. Throughout the demonstration, the teacher explained why it was important to accurately complete job applications and instructed the students to be careful to complete all of the information and to be sure that they put the information in the correct place. To actively engage the students, the teacher used prompts throughout the demonstration, such as "What is it I have to do? I need to . . ." and "How am I doing?" The students were encouraged to help the teacher complete the job application. Students were then required to practice completing a job application. They were allowed only one practice attempt, and they were allowed to ask any questions during this time. The teacher provided corrective feedback only upon request throughout the session.

Finally, the students independently completed the posttest job application. The teacher did not provide the students any assistance during this time. Once again, these conditions (job application, instructions, and amount of time) were the same as those employed under the pretest and learning strategy instruction conditions.

Fidelity of Implementation

Fidelity of implementation was assessed under both experimental conditions by observing the teacher on the day of instruction to ensure that she followed the teaching steps associated with each of the experimental conditions. The primary researcher used a checklist to track whether the teacher fully completed the teaching functions described above under each condition.

RESULTS

Preliminary analyses indicated that there were nonsignificant differences between the groups on the pretest measures. Posttest measures were analyzed in condition (traditional, strategy) by gender (male, female) analyses of variance (ANOVAs). For every dependent measure, only a significant main effect for condition was obtained. The *F* values for these effects, along with the means and standard deviations for each of the dependent measures, are presented in Table 2.

The findings indicate that students who received instruction in the learning strategy condition made statistically significant lower numbers of information omission errors and location errors than students under the job application instruction condition. Additionally, these students received statistically significant higher holistic ratings on their job applications than their counterparts.

TABLE 2. Mean Number of Information Omissions and Location Errors, and Mean Holistic Rating of Overall Application Neatness

Dependent measure	Group A	Group B	*F*(1, 31) (Condition)
Omissions	5.35 (2.55)	0.63 (0.63)	15.29*
Location errors	1.35 (0.99)	0.25 (0.25)	5.29**
Neatness rating	3.37 (1.05)	4.46 (0.51)	7.25***

Note. Group A refers to the traditional instruction condition and Group B refers to the strategy instruction condition. Numbers in parentheses are standard deviations.
*$p < .001$. **$p < .05$. ***$p < .01$.

There were statistically nonsignificant main effects for gender and nonsignificant condition by gender interactions for all of the dependent measures.

Confidence in these results is strengthened by the results of the checks for fidelity of implementation conducted under both experimental conditions. These findings showed that the teacher fully completed the teaching functions described above under each condition.

The social validity measure was analyzed in a condition (traditional, strategy) by gender (male, female) ANOVA. A significant main effect for condition was obtained, $F(1,31) = 6.12$, $p < .05$. There were statistically nonsignificant main effects for gender and condition by gender interactions for the social validity measure. The effects of the job application training on the ratings (1 = *very unlikely* to 5 = *very likely*) by the supervisor of classified personnel suggest that students under the learning strategy condition (mean = 4.21; *SD* = 0.46) would be more likely to receive invitations for job interviews after training than those under the traditional condition (mean = 2.88; *SD* = 1.02).

DISCUSSION

Past research on learning strategies has focused on skills that were general in nature and that apply across subject matters (Brandt, 1989). Recent work on learning strategies, however, has focused on studying how people learn particular things in particular environments. The present study was designed to develop and assess the effects of a learning strategy designed specifically to help students with LD understand the procedural text or written directions included on a job application and provide the requested information.

The results of this study suggest that a sample of students identified as learning disabled according to the state of Washington and federal guidelines were capable of mastering a six-step job application learning strategy in a relatively short time. Because the accurate completion of job applications constitutes an important component in the job search process, these procedures may be very beneficial in facilitating successful job acquisition by students with learning disabilities.

The findings of this study support those of other researchers (e.g., Clark et al., 1975; Mathews & Fawcett, 1984), demonstrating the beneficial effects of teaching students employment application skills. In addition to statistically significant lower numbers of information omissions and information location errors, holistic ratings of the overall neatness of the job applications were much higher for those students under the learning strategy instruction condition. Confidence in the findings of this study are strengthened by the fidelity of implementation data that indicate that the experimental conditions were fully implemented under both conditions.

Most important, in addition to statistically significant positive changes in the three performance measures, the social validity data suggest that the learning strategy instruction resulted in job application forms that would be more likely to elicit invitations for job interviews. The supervisor of classified personnel at the local university indicated that he would be likely to give the students under the learning strategy condition an invitation for a job interview. In contrast, he was significantly less likely to grant a job interview to students under the traditional instruction condition.

It is important to note several limitations of the study. First, the present study provides only one comparison of many potential instructional approaches. Thus, conclusions regarding the efficacy of the learning strategy instruction over any other instructional practices must be made cautiously. Second, because maintenance was not assessed, the long-term impact of the training is uncertain. Third, although students under the learning strategy condition verbally stated that they had employed and articulated the six-steps included in the learning strategy to complete the job applications, the subjective nature of this self-report data does not fully substantiate their claim. Fourth, the relatively small number of subjects limits conclusions regarding the effectiveness of this strategy with students from other areas of the country or, more important, students with other types of disabilities and abilities. Finally, the limited nature of the task also limits conclusions regarding the effectiveness of this type of instruction for other procedural types of texts. The skills required to understand procedural text and perform the required functions accompanying technological devices, such as videocassette recorders, personal computers, programmable microwave ovens, and so forth, differ from those required to complete a standard job application. Procedural text for these devices, with accompanying illustrations, require an individual to fully understand sequence and direction and location concepts. Readers, for example, must be sensitive to sequence cues such as "then," "next," and "finally," and complex direction and location concepts, such as "down," "outside," "against," "inside," and "up." These complex directions, when combined with manual operations and the need to monitor progress, pose an instructional dilemma.

In summary, the skills addressed in the present study, although important, are relatively simple compared to demands that students may encounter regarding the understanding of procedural text. Further research is needed not only to clarify the results of the present study, but also to address the instructional requirements for preparing students with disabilities and others to effectively manage procedural text. Given our rapidly expanding technological society, the complexity of procedural text is only going to increase. Thus, researchers and teachers must continue to develop instructional procedures to facilitate students' understanding of procedural text.

REFERENCES

Andre, M. E., & Anderson, T. H. (1978–79). The development and evaluation of a self-questioning study technique. *Reading Research Quarterly, 14*, 605–623.

Azrin, N. H., & Philip, R. A. (1979). The Job Club methods vs. a lecture–discussion-role play method of obtaining employment for clients with job-finding handicaps. *Rehabilitation Counseling Bulletin, 23*, 144–155.

Baker, L., & Anderson, R. L. (1982). Effects of inconsistent information on text processing: Evidence for comprehension monitoring. *Reading Research Quarterly, 17*, 281–294.

Baumann, J. F. (1986). The direct instruction of main idea comprehension ability. In J. F. Bauman (Ed.), *Teaching* *main idea comprehension* (pp. 133–178). Newark, DE: International Reading Association.

Borkowski, J. G., Weyhing, R. S., & Carr, M. (1988). Effects of attributional retraining on strategy-based reading comprehension in learning-disabled students. *Journal of Educational Psychology, 80*, 46–53.

Brandt, R. (1989). On learning research: A conversation with Lauren Resnick. *Educational Leadership, 46*(4), 12–16.

Brown, A. L., & Palincsar, A. S. (1985). *Reciprocal teaching of comprehension strategies: A natural history of one program to enhance learning* (Tech. Rep. No. 334). Urbana: University of Illinois, Center for the Study of Reading.

Clark, H. B., Boyd, S. B., & MaCrae, J. W. (1975). A classroom program teaching disadvantaged youths to write biographical information. *Journal of Applied Behavior Analysis, 8*, 67–75.

Cunningham, J. W., & Moore, D. W. (1986). The confused world of main idea. In J. B. Baumann (Ed.), *Teaching main idea comprehension* (pp. 1–17). Newark, DE: International Reading Association.

Day, J. D. (1980). *Teaching summarization skills: A comparison of training methods*. Unpublished doctoral dissertation, University of Illinois, Urbana–Champaign.

Englert, C. S., Raphael, T. E., Anderson, L. M., Anthony, H. M., & Stevens, D. D. (1991). Making strategies and self-talk visible: Writing instruction in regular and special education classrooms. *American Educational Research Journal, 28*, 337–372.

Field, H. S., & Holley, W. H. (1976). Resume preparation: An empirical study of personnel managers' perceptions. *Vocational Guidance Quarterly, 25*, 229–237.

Fox, B. J., & Siedow, M. D. (1980). Written directions for content area classrooms—Do students understand the teacher? *The Clearing House, 54*, 101–104.

Furman, W., Geller, M., Simon, S. J., & Kelly, J. A. (1979). The use of a behavior rehearsal procedure for teaching job-interview skills to psychiatric patients. *Behavior Therapy, 10*, 157–167.

Graham, S., & Harris, K. R. (1989). Component analysis of cognitive strategy instruction: Effects on learning disabled students' skills at composing essays: Self-instructional strategy training. *Exceptional Children, 56*, 201–214.

Hall, C., Sheldon-Wildgen, J., & Sherman, J. A. (1980). Teaching job interview skills to retarded clients. *Journal of Applied Behavior Analysis, 13*, 433–442.

Hansen, J. (1981). The effects of inference training and practice on young children's reading comprehension. *Reading Research Quarterly, 16*, 391–417.

Hare, V. C., & Borchardt, K. M. (1984). Direct instruction of summarization skills. *Reading Research Quarterly, 20*, 62–78.

Harris, K. R., & Graham, S. (1985). Improving learning disabled students' composition skills: Self-control strategy training. *Learning Disability Quarterly, 8*, 27–36.

Henk, W. A., & Helfeldt, J. P. (1987). How to develop independence in following written directions. *Journal of Reading, 30*, 602–607.

Hollansworth, J. G., Dressel, M. E., & Stevens, J. (1977). Use of behavioral versus traditional procedures for increasing job interview skills. *Journal of Counseling Psychology, 24*, 503–510.

Mathews, R. M., & Fawcett, S. B. (1984). Building the capacities of job candidates through behavioral instruction. *Journal of Community Psychology, 12*, 123–129.

Mathews, R. M., Whang, P. L., & Fawcett, S. B. (1981). Behavioral assessment of job-related skills. *Journal of Employment Counseling, 18*, 3–11.

Nelson, J. R., Smith, D. J., & Dodd, J. M. (1992). *The effects of a summary skills intervention on students' comprehension of science text*. Manuscript submitted for publication.

Palincsar, A. M., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*, 117–175.

Pearson, P. D. (1985). Changing the face of reading comprehension instruction. *The Reading Teacher, 38*, 724–738.

Raphael, T. E., & McKinney, J. (1983). An examination of 5th and 8th grade children's question answering behavior: An instructional study in metacognition. *Journal of Reading Behavior, 15*(5), 67–86.

Raphael, T. E., & Pearson, P. D. (1985). Increasing students' awareness of sources of information for answering questions. *American Educational Research Journal, 22*, 217–235.

Raphael, T. E., & Wonnacott, C. A. (1985). Heightening fourth-grade students' sensitivity to sources of information for answering questions. *Reading Research Quarterly, 20*, 282–296.

Schmidt, J., Deshler, D., Schumaker, J., & Alley, G. (1988). The effects of generalization instruction on the written performance of adolescents with learning disabilities in the mainstream classroom. *Reading, Writing, and Learning Disabilities, 4*, 291–309.

Taylor, B. M. (1982). Text structure and children's comprehension and memory for expository materials. *Journal of Educational Psychology, 15*, 401–405.

Taylor, B. M., & Beach, R. W. (1984). The effects of text structure instruction on middle-grade students' comprehension and production of expository text. *Reading Research Quarterly, 19*, 134–146.

Taylor, B. M., & Berkowitz, B. S. (1980). Facilitating childrens' comprehension of content material. In M. L. Kamil & A. J. Moe (Eds.), *Perspectives in reading research and instruction* (pp. 64–68). Clemson, SC: National Reading Conference.

Vosniadou, S., Pearson, P. D., & Rogers, T. (1988). What causes children's failures to detect inconsistencies in text? Representation vs. comparison difficulties. *Journal of Educational Psychology, 80*, 27–39.

Williams, J. P. (1986). Extracting important information from text. In J. A. Niles & R. V. Lalik (Eds.), *Solving problems in literacy: Learners, teachers, and researchers* (pp. 11–29). Rochester, NY: National Reading Conference.

ABOUT THE AUTHORS

J. Ron Nelson, PhD, is an assistant professor in the Department of Applied Psychology at Eastern Washington University. He received his doctoral degree in special education from Utah State University. His major research interests include teaching methods for exceptional students and students at risk for school failure, as well as students' social and intellectual reasoning. **Deborah J. Smith, PhD,** is project director in Disabled Student Services and adjunct professor in the Department of Applied Psychology at Eastern Washington University. She received her doctoral degree in special education from Utah State University. Her major research interests include self-management strategies for classroom deportment and academic performance, as well as cognitive instructional strategies. **John M. Dodd, EdD,** is a professor in the Department of Reading and Special Education at Eastern Montana College. He received his doctoral degree in human development at the University of Kansas. His major research interests include culturally sensitive identification and programming for linguistically and culturally diverse students. Address: J. Ron Nelson, Department of Applied Psychology, MS/92, Eastern Washington University, Cheney, WA 99004.

Source: From Nelson, J. R., Smith, D. J., & Dodd, J. M. (1994). The effects of learning strategy instruction on the completion of job applications by students with learning disabilities. *Journal of Learning Disabilities*, *27*, 104–110.

ILLUSTRATIVE EXAMPLE QUESTIONS

1. Are there any problems with the way participants were selected and assigned to the two conditions?
2. What type of experimental design was used in the study?
3. Why was fidelity of implementation assessed? Does it increase your confidence in the findings? Why?
4. Why do you think it was important for the authors to use a variety of dependent measures?
5. Are there any problems with treatment diffusion, compensatory rivalry, or resentful demoralization of the traditional instruction group? Why?
6. Was it necessary for the authors to assess the initial equivalence of the two groups?
7. What other statistical analysis procedure could the authors have used to ensure the initial equivalence of the experimental and control groups?
8. What was the purpose of providing the preskill lesson to both of the groups?
9. What could the authors have done to improve the internal validity of the study?
10. What could the authors have done to improve the external validity of the study?

ADDITIONAL RESEARCH EXAMPLES

1. Cheng, C., Wang, F., & Golden, D. L. (2011). Unpacking cultural differences in interpersonal flexibility: Role of culture-related personality and situational factors. *Journal of Cross-Cultural Psychology, 42*, 425–444.
2. Horrey, W. J., Lesch, M. F., Kramer, A. F., & Melton, D. F. (2009). Effects of a computer-based training module on drivers' willingness to engage in distracting activities. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 51*, 571–581.
3. Kellow, J. T., & Jones, B. D. (2008). The effects of stereotypes on the achievement gap: Reexamining the academic performance of African American high school students. *Journal of Black Psychology, 34*, 94–120.

THREATS TO INTERNAL VALIDITY

Circle the number corresponding to the likelihood of each threat to internal validity being present in the investigation and provide a justification.

1 = definitely not a threat	2 = not a likely threat	3 = somewhat likely threat
4 = likely threat	5 = definite threat	NA = not applicable for this design

Results in Differences within or between Individuals

1. Maturation 1 2 3 4 5 NA

   Justification ______________________________

   ____

2. Selection 1 2 3 4 5 NA

   Justification ______________________________

   ____

3. Selection by Maturation Interaction 1 2 3 4 5 NA

   Justification ______________________________

   ____

4. Statistical Regression 1 2 3 4 5 NA

   Justification ______________________________

   ____

5. Morality 1 2 3 4 5 NA

   Justification ______________________________

   ____

6. Instrumentation 1 2 3 4 5 NA

   Justification ______________________________

   ____

7. Testing 1 2 3 4 5 NA

   Justification ______________________________

   ____

(continued)

8. History 1 2 3 4 5 NA

Justification ____

____

9. Resentful Demoralization of the Control Group 1 2 3 4 5 NA

Justification ____

____

Results in Similarities within or between Individuals

10. Diffusion of Treatment 1 2 3 4 5 NA

Justification ____

____

11. Compensatory Rivalry by the Control Group 1 2 3 4 5 NA

Justification ____

____

12. Compensatory Equalization of Treatments 1 2 3 4 5 NA

Justification ____

____

Abstract: Write a one-page abstract summarizing the overall conclusions of the authors and whether or not you feel the authors' conclusions are valid based on the internal validity of the investigation.

THREATS TO EXTERNAL VALIDITY

Circle the number corresponding to the likelihood of each threat to external validity being present in the investigation according to the following scale:

1 = definitely not a threat	2 = not a likely threat	3 = somewhat likely threat
4 = likely threat	5 = definite threat	NA = not applicable for this design

Also, provide a justification for each rating.

Population

1. Generalization across Subjects 1 2 3 4 5 NA

   Justification ______________________________

   ____

2. Interaction of Personological Variables and Treatment 1 2 3 4 5 NA

   Justification ______________________________

   ____

Ecological

3. Verification of the Independent Variable 1 2 3 4 5 NA

   Justification ______________________________

   ____

4. Multiple Treatment Interference 1 2 3 4 5 NA

   Justification ______________________________

   ____

5. Hawthorne Effect 1 2 3 4 5 NA

   Justification ______________________________

   ____

6. Novelty and Disruption Effects 1 2 3 4 5 NA

   Justification ______________________________

   ____

7. Experimental Effects 1 2 3 4 5 NA

   Justification ______________________________

   ____

(continued)

8. Pretest Sensitization 1 2 3 4 5 NA

   Justification ______________________________

   ____

9. Posttest Sensitization 1 2 3 4 5 NA

   Justification ______________________________

   ____

10. Interaction of Time of Measurement and Treatment Effects 1 2 3 4 5 NA

    Justification ______________________________

    ____

11. Measurement of the Dependent Variable 1 2 3 4 5 NA

    Justification ______________________________

    ____

12. Interaction of History and Treatment Effects 1 2 3 4 5 NA

    Justification ______________________________

    ____

Abstract: Write a one-page abstract summarizing the overall conclusions of the authors and whether or not you feel the authors' conclusions are valid based on the external validity of the investigation.