

# Ex Post Facto Research

Not all important questions in education can be answered with experimental research.

## INSTRUCTIONAL OBJECTIVES

After studying this chapter, the student will be able to:

- 1 Describe ex post facto research and compare it to experimental research.
- 2 State conditions needed to infer a causal relationship.
- 3 Describe alternative explanations in ex post facto research and identify cases in which these are or are not plausible.
- 4 Describe methods of partial control and identify cases in which they would be useful.
- 5 Distinguish between proactive and retroactive ex post facto designs.
- 6 Identify questions for which ex post facto research would be the method of choice.
- 7 State the major weakness of ex post factor research.

As researchers probe such educational questions as “Why are some children better readers than others?” “What is the effect of single-parent homes on achievement?” and “Why do some youths become delinquent while others do not?” they find that only some questions can be investigated through experimental research. If you want to investigate the influence of such variables as home environment, motivation, intelligence, parental reading habits, age, ethnicity, gender, disabilities, self-concept, and so forth, you cannot randomly assign students to different categories of these variables. Independent variables such as these are called **attribute independent variables**. An attribute variable is a characteristic that a subject has before a study begins.

In contrast, an independent variable that an investigator can directly manipulate is an **active independent variable**. An investigator can determine which students will have access to a computer laboratory and which will not or which will use program A to study a unit in algebra and which will use program B. When active independent variables are involved, an investigator can employ experimental or quasi-experimental research. When an investigation involves attribute independent variables that the researcher cannot manipulate, he or she must turn to *ex post facto* research. Ex post facto research is also appropriate when the variable actually could be manipulated but is not because it would be unethical or irresponsible to do so. For example, it would not be ethical to manipulate illegal drug use or use of alcohol or cigarettes to study their effects on human subjects. To study the effect of retention on subsequent achievement, you would not want to randomly assign some children to be retained and others to be promoted. Also, you would not manipulate the use of a drug such as Ritalin to study its effects on children's problem-solving behavior. In such cases, a researcher could use ex post facto research to compare the subsequent achievement of students who have been retained with the academic achievement of otherwise equivalent students who have been promoted or to compare the problem-solving behavior of a group of children already taking Ritalin with that of a matched group not taking the drug.

The designation *ex post facto*, from Latin for "after the fact," indicates that **ex post facto research** is conducted after variation in the variable of interest has already been determined in the natural course of events. This method is sometimes called *causal comparative* because its purpose is to investigate cause-and-effect relationships between independent and dependent variables. Researchers use it in situations that do not permit the randomization and manipulation of variables characteristic of experimental research. Thus, much of the basic rationale for experimental and ex post facto is the same. They both investigate relationships among variables and test hypotheses.

However, with an experiment it is possible to obtain much more convincing evidence for a **causal** (functional) **relationship** among variables than can be obtained with ex post facto studies. The effects of extraneous variables in an experiment are controlled by the experimental conditions, and the antecedent independent variable is directly manipulated to assess its effect on the dependent variable. If you observe  $Y$  to vary concomitantly with the variation in  $X$  in this controlled situation, then you have obtained evidence for the validity of the hypothesized antecedent-consequent relationship between  $X$  and  $Y$ . In an ex post facto investigation, in contrast, the researcher cannot control the independent variables by manipulation or by randomization: Changes in the variables have already taken place. Because of this lack of control, in an ex post facto study it is more hazardous to infer a genuine relationship between  $X$  and  $Y$ .

Let us illustrate the difference between ex post facto and an experimental approach by examining these two approaches to the same research question. Consider the question of the effect of students' anxiety in an achievement-testing situation on their examination performance. The ex post facto approach would involve measuring the already existing anxiety level at the time of the examination and then comparing the performance of "high anxious" and "low anxious" students. The weakness of such an approach is that you could not necessarily conclude that the students' anxiety produced the observed difference

in achievement examination performance. Both sets of scores may have been influenced by a third factor, such as knowledge of the subject matter being examined or general self-confidence. Knowledge or aptitude may be the major cause of both the level of anxiety and the achievement test results.

In an experimental approach to the same problem, the investigator could randomly assign subjects to two exam conditions that are identical in every respect except that one is anxiety arousing and the other is neutral. The experimenter can induce anxiety by telling the subjects that their final grade may depend on their performance, that they may find the test to be extremely difficult, or that the test can be used to identify the incompetent. The neutral group would merely be told that their cooperation is needed for the experiment. The investigator could randomly assign subjects to the two conditions. Then, if the anxious group performed better than the neutral group, it could be concluded that the induced anxiety had a facilitating effect on test performance. If the neutral group performed better than the anxious group, it could be concluded that the induced anxiety had a debilitating effect. A conclusion could be legitimately drawn because of the control provided by the random assignment of groups to treatments and by the experimenter's direct manipulation of the independent variable. Anxiety is one of the few variables that can be either an active or an attribute independent variable. You can manipulate it actively, as described (experimental approach), or you can classify subjects on the basis of their scores on an anxiety measure (ex post facto approach).

Ex post facto research, unlike experimental research, does not provide the safeguards that are necessary for making strong inferences about causal relationships. Mistakenly attributing causation based on a relationship between two variables is called the **post hoc fallacy**. An investigator who finds a relationship between the variables in an ex post facto study has secured evidence only of some concomitant variation. Because the investigator has not controlled  $X$  or other possible variables that may have determined  $Y$ , there is less basis for inferring a causal relationship between  $X$  and  $Y$ .

If you wish to reach a conclusion that one variable ( $X$ ) is the cause of another variable ( $Y$ ), three kinds of evidence are necessary:

1. A statistical relationship between  $X$  and  $Y$  has been established.
2.  $X$  preceded  $Y$  in time.
3. Other factors did not determine  $Y$ .

A statistical relationship is one in which a change in one variable can be predicted from a change in the other. However, we must note that such a relationship between two variables, by itself, is not sufficient evidence of cause and effect. You must proceed to look for evidence on the other two criteria.

Therefore, the investigator must also establish the time sequence; that is, you must consider whether  $Y$  might have occurred before  $X$  and hence could not be an effect of  $X$ . If  $X$  is a cause of  $Y$ , then a change in  $X$  must precede a change in  $Y$ . Decisions about the time relationship between  $X$  and  $Y$  can be made either on a logical basis or as a result of measurements that show the groups did not differ on  $Y$  before exposure to  $X$ .

The third kind of evidence shows that no other plausible explanation can account for the observed effect. It is extremely important that the investigator

consider whether factors other than  $X$  might have determined  $Y$ . This is the most difficult evidence to obtain. You proceed to check this possibility by introducing other relevant variables into the analysis and observing how the relationship between  $X$  and  $Y$  is affected by these additional variables. You may find that the relationship between  $X$  and  $Y$  holds up even when the other variables are introduced. In this case, you have evidence to support a causal inference. However, you may find that other variables may account for the apparent relationship between  $X$  and  $Y$ . In this case, you would conclude that the relationship between  $X$  and  $Y$  is spurious. A **spurious relationship** is one in which the two variables really have no effect on each other but are related because some other variable influences both. For example, a positive relationship between the number of churches and the number of armed robberies in the cities within a state does not mean that building more churches will increase armed robberies or that increasing armed robberies will cause more churches to be built. Here, the extraneous variable is city size. Large cities have more churches and more armed robberies, whereas small cities have fewer armed robberies and fewer churches.

### THINK ABOUT IT 12.1

1. Why does the current administration of the U.S. Department of Education prefer randomized experimental research to ex post facto research?
2. Why do researchers conduct ex post facto research?

#### Answers

1. Because active independent variables are factors that one can change; therefore, true experimental research yields information on which proposed programs have sufficient evidence to obtain funding.
2. Because many important independent variables in education cannot be deliberately manipulated; or, in some cases they could be manipulated, but it would be unethical to do so.

## PLANNING AN EX POST FACTO RESEARCH STUDY

1. *The first step in an ex post facto study is to state the research problem, usually in the form of a question.* What is the relationship between variable A and variable B? Or, what is the effect of variable A on variable B? Again, in this type of research the variables involved are those that the researcher does not directly manipulate. The researcher then states a hypothesis about the expected relationship and defines the variables in operational terms.
2. *Next, select two or more groups to be compared.* Recall that investigators doing ex post facto research achieve the variation they want not by directly manipulating the variable but by selecting individuals in whom the variable is present or absent, strong or weak. Thus, these two groups should differ on the variable of interest in that one group should possess the characteristic and the other group should not, but they should be similar on any relevant extraneous variables. Differential (subject) selections pose a major

threat to the internal validity of ex post facto investigations because you have no control over the selection of subjects into the two groups. They are selected because they already possess the variable of interest, for example, smoker/nonsmoker and retained/not retained. Whenever assignment is not random, there is always an opening for other variables to enter to explain the observed difference between the groups. The way to deal with this threat is to collect data to show that the groups are similar on other extraneous variables that might affect the variable of interest. For example, if you were studying the effect of preschool attendance on the social maturity of kindergarteners, you would have to control any other factors that might have been shown to influence social maturity. Some of these might be age, gender, socioeconomic status, and aptitude. You use logic and previous research to determine what factors need to be controlled in an ex post facto study.

3. *Determine whether your question requires a proactive or a retroactive design.* There are two types of ex post facto research designs—proactive and retroactive.

	Independent Variables	Dependent Variables
Proactive design	Begin with subjects grouped on the basis of preexisting independent variables.	Test hypothesis concerning possible dependent variables.
Retroactive design	Begin with subjects grouped on the basis of preexisting dependent variables.	Test hypothesis concerning possible independent variables.

- a. The **proactive ex post facto research** design begins with subjects grouped on the basis of an independent variable such as father present/father not present or retained/promoted. The researcher then compares these preexisting groups on measures of dependent variables such as self-confidence, mental health, and academic performance.
- b. **Retroactive ex post facto research** seeks possible antecedent causes (independent variables) for a preexisting dependent variable. Whether students graduate from high school or drop out is a variable that cannot be manipulated. Therefore, a researcher would use retroactive ex post facto research to investigate hypotheses about possible causes, such as truancy, attitude toward school, ambition, and parents' education.

## ALTERNATIVE EXPLANATIONS IN EX POST FACTO RESEARCH

When investigators can control the treatment ( $X$ ) and then observe the dependent variable ( $Y$ ) as in experimental research, they have reasonable evidence that  $X$  influences  $Y$ . Ex post facto research, on the other hand, lacks control of the independent variable and thus has lower internal validity. If researchers cannot control ( $X$ ), they may be led to inappropriate conclusions. When interpreting ex post facto research, one should consider alternative explanations, such as common cause, reverse causality, and the presence of other independent variables.

### Common Cause

In an *ex post facto* investigation, you must consider the possibility that both the independent variable and the dependent variable of the study are merely two separate results of a third variable—that they have a **common cause**. For example, if you use a school's total budget as an independent variable and cases of diagnosed learning disability as a dependent variable, you might find a positive correlation between the two variables. Does this mean that an increase in total school budget leads to an increase in cases of learning disability? A more plausible explanation is that the relationship is spurious. An increase in school size/number of children attending could account for both the budget and the cases of diagnosed learning disability because funding is tied to the number of students. It is well established that the average income of private high school graduates is much higher than the average income of public and parochial high school graduates. Does this mean that private schools better prepare students for financial success? Or is the difference due to the fact that those families with enough money to send their children to private schools are also able to finance their children's professional training and set them up in business?

When doing *ex post facto* research, you must always consider the possibilities of common cause or causes accounting for an observed relationship. In our examples, fairly obvious common causes could be identified. However, in *ex post facto* research there is always a nagging doubt that maybe common causes no one has thought of explain a relationship. Research has shown that the injury rate of drivers who use safety belts is lower than the injury rate of drivers who do not. Is this because using safety belts reduces injury or because cautious drivers have fewer injury-causing accidents and also are more likely to use safety belts?

### Reverse Causality

In interpreting an observed relationship in an *ex post facto* study, the researcher must consider the possibility of **reverse causality**—that the reverse of the suggested hypothesis could also account for the finding. Instead of saying that *X* causes *Y*, perhaps it is the case that *Y* causes *X*. For instance, the proportion of Episcopalians listed in *Who's Who in America* is much greater than their proportion in the general population. Does this mean Episcopalians leads to the kind of success that results in being listed in *Who's Who*? It is just as plausible to hypothesize that successful people tend to gravitate to the Episcopal Church. Similarly, if you find that college students who drink have a lower grade point average (GPA) than nondrinkers, you cannot automatically conclude that alcohol consumption depresses academic performance. Perhaps bad grades drive students to drink. (Or, of course, any number of common causes could lead to both drinking and poor grades.)

Investigations on the effects of child-rearing practices have revealed that children who are frequently punished show more aggressive behavior. Can you conclude that parental punishment leads to aggressive children? Or are aggressive children more likely to be punished? A study conducted at the University of Southern California (Hotz, 2002) reported that children who are outgoing and adventurous as toddlers have substantially higher IQs and reading ability by the time they are preteens (age 11 years). The difference in mental ability was true for boys and girls and across variations in ethnic background and income levels.

The researchers correctly concluded that the research does not show whether the children's outgoing curiosity as toddlers was the cause of their better scores on later IQ tests or whether children who have higher IQs are, by nature, more curious and outgoing.

The hypothesis of reverse causality is easier to deal with than the hypothesis of common cause. With the latter, numerous common causes in each case could produce a spurious relationship. With reverse causality, there is only one possibility in each case:  $Y$  caused  $X$ , instead of  $X$  caused  $Y$ .

In a situation in which  $X$  always precedes  $Y$  in time, the very nature of the data rules out the possibility of reverse causality. For example, numerous studies have shown that the average annual income of college graduates is higher than the average annual income of nongraduates. You can rule out the hypothesis of reverse causality because graduation or nongraduation precedes the subsequent annual income. You cannot rule out a variety of possible common causes.

A method of establishing the time order of variables is to obtain measurements of the same subjects at different times. Let us assume that you are interested in the relationship between employees' acceptance of the philosophy of a corporation and job promotion within that corporation. If you merely interviewed a sample of the employees and found that those in higher positions held attitudes and opinions more in line with the company's value system, you would not know whether acceptance of company values and objectives led to promotion or whether promotion increased acceptance of the company value system. To rule out reverse causality as an explanation, you could interview a group of new trainees and obtain—by means of a questionnaire, rating scale, or the like—a measure of their acceptance of the corporation philosophy. Then after a period of time, perhaps 18 months, you could determine from company records which employees had been promoted. If the findings showed that a significantly higher proportion of employees who had expressed attitudes and opinions consistent with corporation philosophy had been promoted, compared with those who had not, you would have better evidence that conformity with company philosophy was conducive to promotion. (You are still left with the possibility that some common cause or causes account for differences in both philosophy and promotion.)

### Other Possible Independent Variables

Independent variables other than the one considered in the ex post facto study may bring about the observed effect on the  $Y$  variable; that is, in addition to  $X_1$ , other variables,  $X_2$  and  $X_3$ , may also be antecedent factors for the variation in the dependent variable. The recorded suicide rate in Sweden is among the highest in the world. Does this mean the Swedish environment causes more people to commit suicide? Does it mean that the Swedish people are more suicide prone than others? Perhaps there is truth in one or both of these hypotheses. It is equally possible, however, that the actual independent variable is the honesty of coroners in Sweden compared with the honesty of coroners in other countries. In countries where great social stigma falls on the families of those who commit suicide, coroners may well use every conceivable means to record a death as accidental rather than suicide. Therefore, the difference between reported suicide rates may be a function of coroner behavior and nothing else.

Let us consider some more examples. At a governors' conference, governor *X* points with pride to the low crime rate in his state. Another governor points out that the police forces in governor *X*'s state are seriously undermanned and the low crime rate may indicate only that very few crimes there are ever reported. An industrialist asks his personnel manager why he does not hire more Old Tuephingen University graduates, asserting that because so many of them are rapidly moving up the promotion ladder, they are obviously more competent than other graduates. The personnel manager tactfully points out that the phenomenon may not be explained by competence but, rather, by the fact that the industrialist is himself an Old Tuephingen University graduate and may be subconsciously favoring his fellow alumni in promotion decisions.

### THINK ABOUT IT 12.2

1. The chief of police finds that the proportion of African Americans who are arrested for traffic violations is much higher than the proportion of whites in his city. State a possible explanation the chief should consider before concluding that African Americans are more reckless drivers than others.
2. Why is reverse causality not a credible explanation in this case?

#### Answers

1. His police officers expect African Americans to be more reckless and thus give them greater surveillance than other drivers. The "hood" expression is "arrested for DWB" (Driving While Black).
2. It is not possible for traffic violations to cause skin color.

Research published by Donahoo and Eckel (2004) reported that the arrival of spring can bring about a significant drop in cholesterol levels of people suffering from high cholesterol. Can we conclude that spring weather causes cholesterol levels to fall? A more likely explanation is that with the arrival of spring people get outside to walk and engage in other physical activity and that it is the increased physical activity that results in the drop in cholesterol. We can rule out common cause and reverse causality in this case, but it is plausible that another independent variable (exercise) could account for the finding.

An obvious first task for investigators is to attempt to list all the possible alternative independent variables. Then by holding the others constant, you can test in turn each variable to determine if it is related to *Y*. If you can eliminate the alternative independent variables by showing that they are not related to *Y*, you gain support for the original hypothesis of a relationship between *X* and *Y*.

### AN APPLICATION OF ALTERNATIVE EXPLANATIONS

Is there a relationship between having a car and academic achievement among high school students? A high school principal observes that the GPA of automobile owners is noticeably lower than the GPA of other students. He proposes placing restrictions on students' use of cars at school. Is such a conclusion justified?

Let us consider the possible alternative hypotheses.

1. *Common cause.* Are there variables that may influence both auto use and scholarship? A possible common cause is student employment. Having a car may require students to work in order to afford the car. Working results in less time spent studying and thus lower grades. Differences in student lifestyle or values could also account for the apparent relationship. If some students value driving highly and have little interest in scholarship, denying them access to cars would not necessarily increase their scholarship. You could propose a number of credible common cause hypotheses.
2. *Reverse causality.* Is it possible that poor grades are a cause of car use? You could reasonably hypothesize that students who do poorly in school look for other paths to social acceptance and that car use is one possible path.
3. *Other possible independent variables.* Are some teachers assuming that car owners are not interested in their studies and assigning them lower grades than they deserve?

How could this question be investigated? Instead of considering only present grades and auto use, one could record the grades of freshmen and the grades of the same subjects at the end of the first semester during their senior year. If auto use affects scholarship, then the grades of drivers would be expected to drop between the time they were freshmen—and therefore not eligible for driver's licenses—and the time they were seniors. Thus, the dependent variable would be the difference between freshman and senior GPA. One might also co-vary (see Chapter 11) on IQ.

### THINK ABOUT IT 12.3

A supervisor of student teachers notices that those with pierced tongues get lower ratings from their mentor teachers than do other student teachers. The supervisor might hypothesize the following:

- a. The same lifestyle that leads to pierced tongues leads to poor student teaching performance.
- b. Pierced tongues lead to poor health, which leads to poor student teaching performance.
- c. Student teachers who are not doing well decide to have their tongues pierced.
- d. Mentor teachers imagine that pierced tongues indicate antisocial attitudes and give student teachers with pierced tongues lower ratings than they deserve.

Which of the previous hypotheses represent the following possibilities?

1. Reverse causality
2. Cause/effect
3. Common cause
4. Alternate independent variable

### Answers

1. c; 2. b; 3. a; 4. d

## PARTIAL CONTROL IN EX POST FACTO RESEARCH

There are strategies for improving the credibility of ex post facto research, although none can adequately compensate for the inherent weakness of such research—namely, lack of control over the independent variable. These strategies provide **partial control** of the internal validity problems of common cause and other possible independent variables. Among these strategies are matching, homogeneous groups, building extraneous variables into the design, analysis of covariance, and partial correlation.

### MATCHING

A common method of providing partial control in ex post facto investigations is to match the subjects in the experimental and control groups on as many extraneous variables as possible. This **matching** is usually done on a subject-to-subject basis to form matched pairs. For example, if you are interested in the relationship between scouting experiences and delinquency, you could locate two groups of boys classified as delinquent and nondelinquent according to specified criteria. In such a study, it would be wise to select pairs from these groups matched on the basis of socioeconomic status, family structure, and other variables known to be related to both choosing the scouting experience and delinquency. The data from the matched samples could be analyzed to determine whether the proportion of those who participated in scouting is greater among nondelinquents than among delinquents. The matching procedure in ex post facto research presents some of the difficulties described in our discussion of its use in experimentation (see Chapter 10).

First, using matching in an ex post facto study assumes that you know what the relevant factors are—that is, the factors that may have some correlation with the dependent variable. Even if you are able to match on a few relevant variables, you will leave many other variables unmatched, and these unmatched variables may affect the dependent variable. Furthermore, matching is likely to greatly reduce the number of subjects that can actually be used in the final analysis. Probably there are several variables that need to be controlled. As the number of matching variables increases, it becomes increasingly more difficult to find a match. The loss of cases inherent in the matching process is an even more serious problem in ex post facto research than in experimentation, in which matching precedes the measurement of the independent variable.

A more serious problem than loss of subjects is the role of regression in an ex post facto matched-pairs design. Let us illustrate the point. Principal A has introduced a new reading program in the fifth grade, and after it has been in use for a year he wants to compare its effectiveness with the effectiveness of the reading program it replaced. In the same district, principal B's school is still using the old program. Both schools give the same standardized reading test at the end of each school year. Principal A compares the mean grade-equivalent reading scores for fifth-graders in the two schools. He finds the mean grade level equivalent for his fifth-graders is 6.0, whereas the mean for principal B's fifth-graders is 4.0. Is this dramatic evidence of the effectiveness of the new method? Principal A realizes that the difference between means could be caused by differences between

the pupils when they began the fifth grade in the two schools. He obtains scores for the reading test administered when the pupils were finishing fourth grade. He finds that his students had a mean grade level equivalent of 4.8, whereas principal B's students had a mean grade level equivalent of only 3.2. Therefore, he must make an adjustment for the fact that the two groups were not at the same point when they began fifth grade.

Principal A decided to use matching as a way to control for this difference. He created matched pairs by matching a student from his school with a fourth-grade reading score of 3.1 with a student from principal B's school with a fourth-grade reading score of 3.1, a student from his school with a score of 4.8 with a student from principal B's school with a score of 4.8, and so on. However, because the mean of the B population is lower than the mean of the A population, there will be many low-scoring B students for whom there is no match in the A group and many high-scoring A students for whom there is no match in the B group. The scores of all the unmatchable students will be excluded from the data analysis. For those who could be matched, the mean score for the pretreatment fourth-grade reading test for group A and group B will be identical. Therefore, it appears the researchers have successfully created a group from school B that is the same as the group from school A with regard to reading achievement.

This all sounds very good. Can you now attribute differences in fifth-grade reading scores to a difference in the effectiveness of the old and new methods? Alas, no! The matched pairs are basically those students with poorer fourth-grade reading scores from population A and those students with higher scores from population B. The matched A students' scores will regress *up* toward the total A mean, and the matched B students' scores will regress *down* toward the total B mean. Thus, when you compare the fifth-grade reading scores of the matched groups, you would expect the A mean to be higher than the B mean even if the new method is no more effective than the old method.

Matching looks good because it provides experimental and control groups that are equal on a pretreatment variable or variables. However, when two matched groups are drawn from different populations, regression toward the original population means will be expected to create spurious results whenever the two populations are not equal. Matching pairs from within a *single* population is often a useful strategy. Matching subjects from one population with subjects from another population is a *bad* strategy. As is the case with change scores, matching only partly adjusts for preexisting differences between groups, and this underadjustment can be misleading.

## HOMOGENEOUS GROUPS

You may recall from the discussion of control in experimentation that it is possible to control for the effects of a variable by selecting samples that are as homogeneous as possible on that variable. A similar procedure can be followed in ex post facto research. Instead of taking a heterogeneous sample and comparing matched subgroups within it, an investigator may control a variable by including in the sample only subjects who are homogeneous on that variable. If academic aptitude is a relevant extraneous variable, the investigator could control its effect by using subjects from only one academic aptitude level. Or if gender is a variable to be

controlled, you can use only males or only females in the research. This procedure serves the purpose of disentangling the independent variable in which you may be interested from other variables with which it is commonly associated so that any effects you find can more justifiably be associated with the independent variable. Using **homogeneous groups** restricts the generalizability of the findings only to the specific homogeneous group, thus reducing external validity of the study. Lee and Loeb (2000) used homogeneous groups to study the effect of school size on teacher attitudes and student achievement. They avoided any influence that the dimension of rural, suburban, and urban settings might have on the dependent variables by studying only schools in the city of Chicago. They found that in small schools (those with fewer than 400 students), teachers had a more positive attitude about their responsibility for students' learning and students learned more than in medium-sized or large schools.

### **BUILDING EXTRANEIOUS VARIABLES INTO THE DESIGN**

It may be possible to build relevant extraneous independent variables into the ex post facto design and investigate their effect through the use of two-way and higher order analyses of variance. For example, suppose an investigator wants to compare the number of trancies among students who have been through an attendance-promoting program and among students who have not been in such a program. If the investigator thinks that ethnic membership and gender influence the number of trancies, the subjects could be classified as Anglo, Hispanic, Asian American, or black, and also as male or female. An  $F$  test would first be used to assess the main effect of program versus no program on number of trancies. An  $F$  test for gender would assess the truancy differences of males and females. The  $F$  test for ethnicity by program versus no program would assess whether the effect of the program was consistent among the ethnic groups. The  $F$  test for program by gender would assess the consistency of program effectiveness between genders. The gender by ethnicity  $F$  test would assess the consistency of gender differences in truancy among ethnic groups. Finally, the  $2 \times 4 \times 2$  higher order interaction  $F$  test would indicate if the treatment by ethnic group interaction was consistent between males and females. Building other variables into an ex post facto design is a partial solution, but you can never be sure that you have *all* the variables that should have been considered.

### **ANALYSIS OF COVARIANCE**

**Analysis of covariance** (ANCOVA) is sometimes used to partially adjust for pre-existing differences between groups in an ex post facto design. Specifically, it adjusts scores on the dependent variable for any initial differences on the extraneous variable. However, because the adjustment is only partial, ANCOVA does not “solve” the problem of initial differences between groups but only reduces it. When interpreting ex post facto research, it is inappropriate to assume ANCOVA has satisfactorily adjusted for initial differences.

A classic example of the problems inherent in matching and ANCOVA is the Cicirelli (1969) ex post facto investigation of the effects of the Head Start program. This study compared the academic achievement of students who had been in the Head Start program with the achievement of those who had not been in the Head Start program. Children who had been in the program were matched with

non-Head Start children from the same neighborhoods on gender, racial/ethnic groups, and kindergarten attendance. ANCOVA was used to adjust for differences in income per capita, educational level of father, and occupational level of father. The results appeared to indicate that the achievement of the non-Head Start group was greater than that of the Head Start group even when scores were adjusted for initial differences. The authors concluded that the Head Start program was harmful. Researchers have pointed out that because of regression, both matching and ANCOVA underadjust for initial differences between groups. One would therefore expect adjusted posttreatment scores of a disadvantaged group to be less than adjusted posttreatment scores of a less disadvantaged group. Because the extent of the underadjustment is unknown, the Cicirelli study does not determine whether the Head Start experience was harmful or beneficial or had no effect.

#### THINK ABOUT IT 12.4

Read the following abstract from a report of a study using ex post facto research that was published in 2000 and answer the following questions:

### *Academic Achievement and Parental School Involvement as a Function of High School Size*

**Pamela W. Gardner**

**Shulamit N. Ritblatt**

**James R. Beatty**

**San Diego State University**

The purpose of this research was to examine academic achievement, absenteeism, dropout rate, and parental school involvement as a function of high school size. Previous studies indicated that, generally, students of small schools exhibited greater extracurricular participation, greater satisfaction, lower absenteeism, and lower dropout rate. In terms of academic achievement, findings have been mixed.

Academic achievement was measured by scores on the Scholastic Aptitude Test (SAT), while parent participation was measured by membership in the Parent–Teacher Association (PTA) or other parent organization(s). Absenteeism and dropout rate were measured using data provided by the California Department of Education.

Sixty-seven randomly selected large California public high schools (enrollments over 2000) were compared to 60 randomly selected small California public high school (enrollments between 200 and 600 students). The t-tests and analyses of covariance were employed to determine the differences between the large and small high schools. Contrary to several hypotheses, the large schools exhibited higher academic achievement on total SAT score, verbal SAT score, and math SAT score. In addition, this cohort had a higher proportion of students taking the SAT. Confirming other hypotheses, the small school cohort displayed lower absenteeism, lower dropout rate, and higher parental school involvement.

The results of this investigation indicate the need for further research, both in size and scope. In addition, the findings carry implications for educational policy decisions. Standard procedures for reporting absenteeism and parental school involvement should be established.

*Source:* Gardner, P. W., Ritblatt, S. N., & Beatty, J. R. (2000). Academic achievement and parental school involvement as a function of high school size. *High School Journal*, 83(2), 21–27.

### Questions

1. What is the independent variable?
2. What are the dependent variables?
3. How are the dependent variables operationally defined?
4. Is this study proactive or retroactive?
5. What is the population of interest?
6. How was the sample selected?
7. Is reverse causality credible?
8. Is common cause credible?
9. Which partial control strategy was used?

### Answers

1. High school size
  2. a. Academic achievement
    - b. Parental participation
    - c. Absenteeism
    - d. Dropout rate
  3. a. Scholastic Aptitude Test scores
    - b. Membership in PTA or other parent organization(s)
    - c. Data from the California Department of Education
  4. Proactive
  5. Students and parents of California high schools
  6. Random selection of large (over 2000 students) and small (200 to 600 students) California high schools
  7. No
  8. No
  9. Analyses of covariance
- 

## THE ROLE OF EX POST FACTO RESEARCH

Given the hazards involved in ex post facto research, many educational researchers believe they should not engage in this type of research at all. Basically, they contend that it is better to admit ignorance than to risk reaching incorrect conclusions. However, others point out that many variables of great interest are not amenable to experimental research. Researchers cannot randomly assign children to broken or intact homes, to high or low social class, to achievement-oriented or non-achievement-oriented peer groups, to high or low self-esteem groups, and so forth. Therefore, if they want to learn anything about relationships between such attribute variables and other variables, the ex post facto method is their only recourse. An ex post facto study is better than no study at all. If researchers use appropriate methods of partial control and consider alternative hypotheses, perhaps they can be correct more often than wrong.

Certainly, there have been many highly credible ex post facto studies. The U.S. Surgeon General's 1964 study of the relationship between smoking and lung cancer is a well-known example. It is not possible to randomly designate

a group of human subjects who are to smoke for years and a group who are not to smoke, so the study had to be done as an ex post facto investigation. The reversed-causality hypothesis that lung cancer causes people to smoke is not plausible. None of the common-cause hypotheses offered seem very likely: Nervous people are prone to both smoking and lung cancer, some genetic predisposition leads to both, and so forth. The Surgeon General controlled for many alternative independent variables—for example, by analyzing separately samples from areas of high air pollution and low air pollution. Experimental results from studies of animals that were made to inhale or not inhale cigarette smoke produced evidence of a cause–effect relationship. Given all this, despite the dangers inherent in ex post facto research, most would conclude that it is better to bet there *is* a cause-to-effect relationship between smoking and lung cancer among humans than to bet there is *not* such a relationship. In fact, much medical research is ex post facto in design. Researchers study diseases and try to determine the influence of factors such as eating habits, lifestyle, and genes. Ex post facto research in education has permitted investigations of the effects of variables such as home background, father absence, early experiences, disabilities, teacher competence, and others that are beyond the control of educators. In some instances, ex post facto research has discovered relationships or raised questions that can later be investigated more systematically in well-controlled experimental studies. Appropriately used and cautiously interpreted, ex post facto research will continue to provide a valuable methodology for the acquisition of knowledge.

## SUMMARY

---

Ex post facto research is used to investigate relationships when the researcher cannot randomly assign subjects to different conditions or directly manipulate the independent variable. Ex post facto research begins with subjects who differ on an observed dependent variable and tries to determine the antecedents (cause) of the difference. Or the researcher begins with subjects who differ on an independent variable and tries to determine the consequences of the difference.

Although there are many disadvantages of ex post facto design, it nevertheless is frequently the only method by which educational researchers can obtain necessary information about characteristics of defined groups of students or information needed for the intelligent formulation of programs in the school. It permits researchers to investigate situations in which

controlled variation is impossible to introduce. Attributes such as academic aptitude, creativity, self-esteem, socioeconomic status, and teacher personality cannot be manipulated and hence must be investigated through ex post facto research rather than through the more rigorous experimental approach.

The possibility of spurious relationships is always present in ex post facto research. Considering the possibilities of common cause, reversed causality, and possible alternate independent variables can help educators evaluate such research more realistically. Several partial control strategies can help researchers avoid gross errors in ex post facto designs, but none can entirely solve the problems inherent in those designs. Always exercise caution when interpreting ex post facto results.

**KEY CONCEPTS**

active independent variable  
 analysis of covariance (ANCOVA)  
 attribute independent variable  
 causal relationship  
 common cause  
 ex post facto research  
 extraneous variables

homogeneous group  
 matching  
 partial control  
 post hoc fallacy  
 proactive ex post facto  
 research

retroactive ex post facto  
 research  
 reverse causality  
 spurious relationship

**EXERCISES**

1. What are the two major limitations of ex post facto or causal-comparative designs?
2. Classify the following independent variables as active or attribute:
  - a. Kindergarten entrance age
  - b. Method of teaching reading
  - c. Montessori school attendance versus attendance at public school
  - d. Teaching experience
  - e. Use of drug A in treating diabetes versus drug B
  - f. Learning style
3. What method would you use to investigate the following research hypotheses?
  - a. High school students who work during the school year will have lower grades and lower standardized test scores than students who do not work.
  - b. First-grade students who are taught reading using phonics will read at a higher level than first-grade students taught by another method.
  - c. Beginning teachers who are mentored will be more effective and express more satisfaction with teaching than beginning teachers who are not mentored.
  - d. Students who are in small classes (less than 15) in grades kindergarten to 3 will perform at a higher level in grades 4 to 6 than will students in regular-sized classes (more than 20).
4. Interpret this statement: "A statistical relationship between two variables is a necessary but not a sufficient condition for inferring a causal relationship between the variables."
5. Show the design you would use to test the hypothesis stated in Exercise 3(a).
6. What alternative hypotheses would you want to explore if your causal-comparative research indicates that there is a relationship between self-concept and achievement in school?
7. What steps would you take to avoid the post hoc fallacy in your ex post facto research?
8. For the following examples, suggest an alternative explanation for the finding:
  - a. A researcher finds a relationship between use of alcohol and unemployment and concludes that using alcohol causes workers to lose their jobs.
  - b. A researcher finds that the proportion of the population confined to mental hospitals for manic-depression is greater in the United States than it is in England. He concludes that the American environment is more conducive to the development of depression.
  - c. A researcher finds that the increase in the use of cell phones in recent years has been accompanied by an increase in the incidence of brain tumors. He concludes that using a cell phone can result in a brain tumor.
9. Find an example of an ex post facto study in the current literature. Identify (a) the problem, (b) the operational definitions of the variables, (c) the design used, (d) the steps taken to control extraneous variables, (e) the findings, and (f) the conclusions. Did the author include a caveat about the conclusion?
10. A recent study reported that of 4500 women who underwent dental X-rays during pregnancy, 1117 had low-birth-weight babies. The researchers concluded that women who have dental X-rays during pregnancy face an increased risk of having underweight babies. What questions might you want to

- ask about this research before accepting their conclusion?
11. The following are examples of published studies. Which do you think involve post facto research methodology?
    - a. Do After-School Programs Help Students Succeed?
    - b. The Impact of Hope and Social Activity on Academic Performance of Midwestern College Students
    - c. The Importance of Middle School Math on High School Mathematics Achievement
    - d. The Effect of Using Classroom Lap Top Computers on the Writing of Middle School Students
  12. Answer the following questions about the Gardner, Ritblatt, and Beatty (2000) study.
    - a. What is the independent variable?
    - b. What are the dependent variables?
    - c. How are the dependent variables operationally defined?
    - d. What is the population of interest?
    - e. How was the sample selected?
    - f. Is reverse causality credible?
    - g. Is common cause credible?

## ANSWERS

1. The two major limitations are lack of randomization and lack of manipulation of the independent variable.
2.
  - a. Attribute
  - b. Active
  - c. Attribute
  - d. Attribute
  - e. Active
  - f. Attribute
3.
  - a. Ex post facto
  - b. Experimental
  - c. Experimental
  - d. Ex post facto (it is generally not feasible to manipulate class size, although it has been done)
4. To report that one variable ( $X$ ) is the cause of another variable ( $Y$ ), the researcher must first show that there is a statistical relationship between the variables. However, this finding by itself is not sufficient to establish a causal relationship. The researcher must examine the time ( $X$  must precede  $Y$ ) and must show that no other variables could cause  $Y$ .
5. Use the proactive design. That is, you would select two groups of high school students, one of whom works a certain number of hours each week ( $X$ ), and would compare their grades and scores on standardized tests ( $Y$ ) with those of a comparable group of students who do not work. Or, instead of studying the effect of working or not working, you could study the effect of the extent of work as measured by the number of hours worked each week. One group could work less than 10 hours a week and the other group could work 20 or more hours. You would have to control a number of variables, including aptitude scores, previous grades, socioeconomic status, and gender. If you used only one class level (only juniors or only seniors), you could control for the variable of class standing (if relevant).
6. A more positive self-concept could lead to higher achievement in school. Or the reverse causality hypothesis could account for the finding. That is, higher achievement in school could lead to a more positive self-concept. You would also want to explore a common-cause hypothesis. Perhaps another variable, such as intelligence, causes both achievement and self-concept.
7. Try to control any relevant extraneous variables that could affect the measured variable. Methods of control include (a) matching, (b) using homogeneous groups, (c) building extraneous variables into the design, and (d) using ANCOVA or other statistical means to partially control unwanted variables.
8.
  - a. Perhaps unemployment leads to the use of alcohol. Or perhaps profound depression causes people to lose their jobs and also take to drink.
  - b. Perhaps depressed people in England receive a different type of treatment and are less likely to be hospitalized than people in the United States. Perhaps English psychologists are less inclined to make a diagnosis of depression.

- c. Perhaps there is a common cause. The stress and rapid pace of contemporary life are conducive to the use of cell phones and may also influence the development of brain tumors.
9. Answers will vary.
10. One would first want to ask why the women received X-rays because, normally, dentists are cautious about using X-rays during pregnancy. Perhaps there was some disease or infection that necessitated the use of X-rays, and it was the disease or infection that caused the low-birth-weight babies. The study's lead author stated that more research is needed to determine if dental X-rays really were the culprit.
11. b and c are ex post facto research.

## REFERENCES

- Cicirelli, V. (1969). *The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development* (report presented to the Office of Economic Opportunity pursuant to contract B89-4536). Westinghouse Learning Corporation and Ohio University. (Distributed by the Clearinghouse for Federal Scientific and Technical Information, U.S. Department of Commerce, National Bureau of Standards, Institute for Applied Technology, PB 184 328.)
- Donahoo, W. T., & Eckel, R. H. (2004). Physiological mechanisms for the seasonal changes in serum cholesterol level. *Archives of Internal Medicine*, *164*(22), 2506–2507.
- Gardner, P. W., Ritblatt, S. N., & Beatty, J. R. (2000). Academic achievement and parental school involvement as a function of high school size. *High School Journal*, *83*(2), 21–27.
- Hotz, R. (2002, April 15). Outgoing toddlers have higher IQs as preteens. *Indianapolis Star*, p. 1.
- Lee, V. E., & Loeb, S. (2000). School size in Chicago elementary schools: Effects on teachers' attitudes and students' achievement. *American Educational Research Journal*, *37*, 1–31.