Chapter 9: Experimental Research

I. Experiments

   a. Experimental designs are often touted as the most "rigorous" of all research designs or, as the standard against which all other designs are judged. In one sense, they probably are. If you can implement an experimental design well (and that is a big "if" indeed), then the experiment is probably the strongest design with respect to internal validity. Why? Recall that internal validity is at the center of all causal or cause-effect inferences. When you want to determine whether a stimulus or treatment causes some predicted outcome or response, then you are interested in having strong internal validity. Essentially, a researcher manipulates the independent variable (exercise, for example) and observes any change in the dependent variable (heart disease, for example).

II. Research Questions Appropriate for an Experiment

   a. Only those problems that let a researcher manipulate conditions in a laboratory setting. Keep in mind that researchers have been very clever in devising ways in which to expand the traditional definition of a laboratory.

III. Random Assignment

   a. Social researchers frequently want to compare. The cliché, "Compare apples to apples, don't compare apples to oranges," is not about fruit; it is about comparisons. It means that a valid comparison depends on comparing things that are fundamentally alike. Random assignment facilitates comparison in experiments by creating similar groups. When making comparisons, researchers want to compare cases that do not differ with regard to variables that offer alternative explanations.

      i. Why Randomly Assign?

         1. Random assignment is a method for assigning cases (e.g., individuals) to groups (e.g., experimental and control) for the purpose of making comparisons in order to increase one's confidence that the groups do not differ in a systematic way.

      ii. How to Randomly Assign

         1. A researcher begins with a collection of cases and then divides the cases into two or more groups using a random mathematical process.

      iii. A Useful Nonrandom Technique Used to Assign Cases Into Groups

         1. Matching

            a. A process whereby a researcher deliberately assigns cases into groups based upon relevant characteristics (a characteristic is considered relevant if in anyway it could affect the dependent variable during the course of the experiment) in order to create similar groups for comparison purposes.

   b. Matching vs. Random Assignment

      i. Matching presents a problem: What are the relevant characteristics to match on, and can one locate exact matches? Individual cases differ in thousands of ways, and the researcher cannot know which might be relevant.

IV. Experimental Logic

   a. The Language of Experiments

      i. Parts of the Experiment

         1. Not all experiments have all these parts, and some have all seven parts plus others.

            a. Independent Variable (stimulus or treatment)

               i. A condition or treatment introduced into the experiment.

            b. Dependent Variable

               i. Dependent variables or outcomes are physical conditions, social behaviors, attitudes, feelings, or beliefs of subjects that change in response to a treatment.

c. Pretest
    i. The measurement of the dependent variable *prior* to introduction of the independent variable.

d. Posttest
    i. The measurement of the dependent variable *after* the introduction of the independent variable.

e. Experimental Group
    i. The group that receives the independent variable.

f. Control Group
    i. The group that does not receive the independent variable.

V. Control in Experiments

a. Control is crucial in experimental research. Aspects of an experimental situation that are not controlled by the researcher are alternatives to the treatment for change in the dependent variable and undermine the ability to establish causality.
    i. Techniques to Establish Control in Experiments
        1. Deception
            a. Occurs when the researcher intentionally misleads subjects through written or verbal instructions, the actions of others (e.g., confederates or stooges), or aspects of the setting.

VI. Design Notation

a. Experiments can be designed in many ways. Design notation is a shorthand system for symbolizing the parts of experimental design. Once you learn design notation, you will find it easier to think about and compare designs. For example, design notation expresses a complex, paragraph-long description of the parts of an experiment in five or six symbols arranged in two lines. It uses the following symbols: O = observation of dependent variable; X = treatment, independent variable; R=random assignment. The Os are numbered with subscripts from left to right based on time order. Pretests are O_1, posttests O_2. When the independent variable has more than two levels, the Xs are numbered with subscripts to distinguish among them. Symbols are in time order from left to right. The R is first, followed by the pretest, the treatment, and then the posttest. Symbols are arranged in rows, with each row representing a group of subjects. For example, an experiment with three groups has an R (if random assignment is used), followed by three rows of Os and Xs. The rows are on top of each other because the pretests, treatment, and posttest occur in each group at about the same time.

VII. Types of Designs

a. Researchers combine parts of an experiment (e.g., pretests, control groups, etc.) together into an *experimental design.* For example, some designs lack pretests, some do not have control groups, and others have many experimental groups. Certain widely used standard designs have names. You should learn the standard designs for two reasons.
    i. First, in research reports, researchers give the name of a standard design instead of describing it. When reading reports, you will be able to understand the design of the experiment if you know the standard designs.
    ii. Second, the standard designs illustrate common ways to combine design parts. You can use them for experiments you conduct or create your own variations.
        1. Classical Experimental Design
            a. All designs are variations of the classical experimental design, the type of design discussed so far, which has random assignment, a pretest and a posttest, an experimental group, and a control group.
        2. Pre-Experimental Designs
            a. Some designs lack random assignment and are compromises or shortcuts. These pre-experimental designs are used in situations where it is difficult to use the classical design. They have

weaknesses that make inferring a causal relationship more difficult.

i. Types of Pre-Experimental Designs

1. One-Shot Case Study Design

a. Also called the one group posttest-only design, the one-shot case study design has only one group, a treatment, and a posttest. Because there is only one group, there is no random assignment. A weakness of this design is that it is difficult to say for sure that the treatment caused the dependent variable. If subjects were the same before and after the treatment, the researcher would not know it.

2. One-Group Pretest-Posttest Design

a. This design has one group, a pretest, a treatment, and a posttest. It lacks a control group and random assignment. This is an improvement over the one-shot case study because the researcher measures the dependent variable both before and after the treatment. But it lacks a control group. The researcher cannot know whether something other than the treatment occurred between the pretest and the posttest to cause the outcome.

3. Static Group Comparison (Posttest Only)

a. Also called the posttest only nonequivalent group design, static group comparison has two groups, a posttest, and treatment. It lacks random assignment and a pretest. A weakness is that any posttest outcome difference between the groups could be due to group differences prior to the experiment instead of to the treatment.

3. Quasi-Experimental and Special Designs

a. These designs, like the classical design, make identifying a causal relationship more certain than do pre-experimental designs. Quasi-experimental designs help researchers test for causal relationships in a variety of situations where the classical design is difficult or inappropriate. They are called quasi because they are variations of the classical experimental design. Some have randomization but lack a pretest, some use more than two groups, and others substitute many observations of one group over time for a control group. In general, the researcher has less control over the independent variable than in the classical design.

i. Types of Quasi-Experimental Designs

1. Two-Group Posttest-Only Design

a. This is identical to the static group comparison, with one exception: The groups are randomly assigned. It has all the parts of the classical design except a pretest. The random assignment reduces the chance that the groups differed before the treatment, but without a pretest, a researcher cannot be as certain that the groups began the same on the dependent variable.

2. Interrupted Time Series
    a. In an interrupted time series design, a researcher uses one group and makes multiple pretest measures before and after the treatment.
3. Equivalent Time Series
    a. An equivalent time series is another one-group design that extends over a time period. Instead of one treatment, it has a pretest, then a treatment and posttest, then treatment and posttest, then treatment and posttest, and so on.

4. Latin Square Designs
    a. Researchers interested in how several treatments given in different sequences or time orders affect a dependent variable can use a Latin square design.
        i. Types of Latin Square Designs
            1. Solomon Four-Group Design
                a. A researcher may believe that the pretest measure has an influence on the treatment or dependent variable. A pretest can sometimes sensitize subjects to the treatment or improve their performance on the posttest (see the discussion of testing effect to come). Richard L. Solomon developed the Solomon four-group design to address the issue of pretest effects. It combines the classical experimental design with the two group posttest-only design and randomly assigns subjects to one of four groups.
            2. Factorial Designs
                a. Sometimes, a research question suggests looking at the simultaneous effects of more than one independent variable. A factorial design uses two or more independent variables in combination. Every combination of the categories in variables (sometimes called factors) is examined. When each variable contains several categories, the number of combinations grows very quickly. The treatment or manipulation is not each independent variable; rather, it is each combination of the categories.

VIII. Internal and External Validity

a. The Logic of Internal Validity

i. Internal validity means the ability to eliminate alternative explanations of the dependant variable. Variables, other than the treatment, that affect the dependent variable are threats to internal validity. They threaten the researcher's ability to say that the treatment was the true causal factor producing change in the dependent variable. Thus, the logic of internal validity is to rule out variables other than the treatment by controlling experimental conditions and through experimental designs.

1. Threats to Internal Validity

a. Selection Bias

i. Selection bias is the threat that subjects will not form equivalent groups. It is a problem in designs without random assignment. It occurs when subjects in one experimental group have a characteristic that affects the dependent variable. For example, in an experiment on physical aggressiveness, the treatment group unintentionally contains subjects who are football, rugby, and hockey players, whereas the control group is made up of musicians, chess players, and painters. Another example is an experiment on the ability of people to dodge heavy traffic. All subjects assigned to one group come from rural areas, and all subjects in the other grew up in large cities. An examination of pretest scores helps a researcher detect this threat, because no group differences are expected.

b. History

i. This is the threat that an event unrelated to the treatment will occur during the experiment and influence the dependent variable. History effects are more likely in experiments that continue over a long time period. For example, halfway through a two-week experiment to evaluate subject attitudes toward space travel, a spacecraft explodes on the launch pad, killing the astronauts.

c. Maturation

i. This is the threat that some biological, psychological, or emotional process within the subjects and separate from the treatment will change over time. Maturation is more common in experiments over long time periods. For example, during an experiment on reasoning ability, subjects become bored and sleepy and, as a result, score lower. Another example is an experiment on the styles of children's play between grades 1 and 6. Play styles are affected by physical, emotional, and maturation changes that occur as the children grow older, instead of or in addition to the effects of a treatment. Designs with a pretest and control group help researchers determine whether maturation or history effects are present, because both experimental and control groups will show similar changes over time.

d. Testing

i. Sometimes, the pretest measure itself affects an experiment. This testing effect threatens internal validity because more than the treatment alone affects the dependent variable. The Solomon four-group design helps a researcher detect testing effects. For example, a researcher gives students an examination on the first day of class. The course is the treatment. He or she tests learning by giving the same exam on the last day of class. If subjects remember the pretest questions and this affects what they learned (i.e., paid attention to) or how they answered questions on the posttest, a testing effect is present. If testing effects occur, a researcher cannot say that the treatment alone has affected the dependent variable.

e. Instrumentation
    i. This threat is related to stability reliability. It occurs when the instrument or dependent variable measure changes during the experiment. For example, in a weight-loss experiment, the springs on the scale weaken during the experiment, giving lower readings in the posttest.

f. Mortality
    i. Mortality or attrition arises when some subjects do not continue throughout the experiment. Although the word mortality means death, it does not necessarily mean that subjects have died. If a subset of subjects leaves partway through an experiment, a researcher cannot know whether the results would have been different had the subjects stayed. For example, a researcher begins a weight-loss program with 50 subjects. At the end of the program, 30 remain, each of who lost 5 pounds with no side effects. The 20 who left could have differed from the 30 who stayed, changing the results. Maybe the program was effective for those who left, and they withdrew after losing 25 pounds. Or perhaps the program made subjects sick and forced them to quit. Researchers should notice and report the number of subjects in each group during pretests and posttests to detect this threat to internal validity.

g. Statistical Regression
    i. Statistical regression *is* not easy to grasp intuitively. It is a problem of extreme values or a tendency for random errors to move group results toward the average. It can occur in two ways
        1. One situation arises when subjects are unusual with regard to the dependent variable. Because they begin as unusual or extreme, subjects are unlikely to respond further in the same direction. For example, a researcher wants to see whether violent films make people act violently. He or she chooses a group of violent criminals from a high-security prison, gives them a pretest, shows violent films, and then administers a posttest. To the researcher's shock, the criminals are slightly less violent after the film, whereas a control group of non-prisoners who did not see the film are slightly more violent than before. Because the violent criminals began at an extreme, it is unlikely that a treatment could make them more violent; by random chance

alone, they appear less extreme when measured a second time.

2. A second situation involves a problem with the measurement instrument. If many subjects score very high (at the ceiling) or very low (at the floor) on a variable, random chance alone will produce a change between the pretest and the posttest. For example, a researcher gives 80 subjects a test, and 75 get perfect scores. He or she then gives a treatment to raise scores. Because so many subjects already had perfect scores, random errors will reduce the group average because those who got perfect scores can randomly move in only one direction-to get some answers wrong. An examination of scores on pretests will help researchers detect this threat to internal validity.

h. Diffusion of Treatment or Contamination

   i. Diffusion of treatment or contamination is the threat that subjects in different groups will communicate to each other and learn about the other's treatment. Researchers avoid it by isolating groups or having subjects promise not to reveal anything to others who will become subjects. For example, subjects participate in a daylong experiment on a new way to memorize words. During a break, treatment group subjects tell those in the control group about the new way to memorize, which control group subjects then use. A researcher needs outside information such as post-experiment interviews with subjects to detect this threat.

i. Compensatory Behavior

   i. Some experiments provide something of value to one group, of subjects but not to another, and the difference becomes known. The inequality may produce pressure to reduce differences, competitive rivalry between groups, or resentful demoralization. All these types of compensatory behavior can affect the dependent variable in addition to the treatment. For example, one school system receives a treatment (longer lunch breaks) to produce gains in learning. Once the inequality is known, subjects in the control group demand equal treatment and work extra hard to learn and overcome the inequality. Another group becomes demoralized by the unequal treatment and withdraws from learning. It is difficult to detect this threat unless outside information is used (see the earlier discussion of diffusion of treatment).

j. Experimenter Expectancy

   i. Although it is not always considered a traditional internal validity problem, the experimenter's behavior, too, can threaten causal logic. A researcher may threaten internal validity, not by purposefully unethical behavior but by indirectly communicating experimenter expectancy to subjects. Researchers may be highly committed to the hypothesis and indirectly communicate desired findings to subjects. For example, a researcher studying reactions toward the disabled deeply believes that females are more sensitive toward the disabled than males are. Through eye contact, tone of voice, pauses, and other nonverbal communication, the researcher unconsciously encourages female subjects to report positive feelings toward the disabled; the researcher's nonverbal behavior is the opposite for male subjects. Here is a way to detect experimenter expectancy. A researcher hires assistants and teaches them experimental techniques. The assistants train subjects and test their learning ability. The researcher gives the assistants fake transcripts and records showing that subjects in one group are honor students and the others are failing, although in fact the subjects are identical. Experimenter expectancy is present if the fake honor students, as a group, do much better than the fake failing students. A commonly used technique by researchers in order to reduce the effects of experimenter expectancy is a double-blind experiment.

1. Double-Blind Experiment
   a. The double-blind experiment is designed to control researcher expectancy. In it, people who have direct contact with subjects do not know the details of the hypothesis or the treatment. It is double blind because both the subjects and those in contact with them are blind to details of the experiment.

b. External Validity and Field Experiments
   i. Even if an experimenter eliminates all concerns about internal validity, external validity remains a potential problem. External validity is the ability to generalize experimental findings to events and settings outside the experiment itself. If a study lacks external validity, its findings hold true only in experiments, making them useless to both basic and applied science.

c. Realism in Experiments
   i. Are experiments realistic? If not, will the affects be replicated outside the laboratory? Two forms of realism can help us answer some of these questions.
      1. Experimental Realism
         a. The degree or impact of an experimental treatment or setting on subjects; it occurs when subjects are caught up in the experiment and are truly influenced by it.
      2. Mundane Realism
         a. Asks: Is the experiment like the real world? Mundane realism mostly answers questions raised about external validity. Two aspects of experiments can be generalized. One is from the subjects: Are the subjects similar to the general population? Another aspect is generalizing from an artificial treatment to everyday life: Is watching a violent horror movie in a classroom similar to watching similar shows over the course of many years?
            i. Reactivity
               1. Subjects may react differently in an experiment than they would in real life because they know that they are in a study.
                  a. Types of Reactivity
                     i. Hawthorne Effect – is a specific kind of reactivity. The name comes from a series of experiments by

Elton Mayo at the Hawthorne, Illinois, plant of Westinghouse Electric during the 1920's. He serendipitously discovered that the act of monitoring an individual may produce changes in the dependent variable.

ii. Novelty Effect – another kind of reactivity that produces changes in the dependent variable as a result of something new being introduced to the subjects.

iii. Demand Characteristics – subjects may pickup clues about the hypothesis or goal of an experiment and they may change their behavior to what they think is demanded of them.

d. Field Experiments

i. This section has focused on experiments conducted under the controlled conditions of a laboratory. Experiments are also conducted in "real life" or field settings where a researcher has less control over the experimental conditions. The amount of control varies on a continuum. At one end is the highly controlled laboratory experiment, which takes place in a specialized setting or laboratory; at the opposite end is the field experiment, which takes place in the "field"-in natural settings such as a subway car, a liquor store, or a public sidewalk. Subjects in field experiments are usually unaware that they are involved in an experiment and react in a more natural way.

e. Practical Considerations

i. Every research technique has informal tricks of the trade. They are pragmatic and based on common sense but account for the difference between the successful research projects of an experienced researcher and the difficulties a novice researcher faces. Three are discussed here:

1. Planning and Pilot Tests

a. All social research requires planning, and most quantitative researchers use pilot tests. During the planning phase of experimental research, a researcher thinks of alternative explanations or threats to internal validity and how to avoid them. The researcher also develops a neat and well-organized system for recording data. In addition, he or she should devote serious effort to pilot testing any apparatus (e.g., computers, video cameras, tape recorders, etc.) that will be used in the treatment situation, and he or she must train and pilot test confederates. After the pilot tests, the researcher should interview the pilot subjects to uncover aspects of the experiment that need refinement.

2. Instructions to Subjects

a. Most experiments involve giving instructions to subjects to set the stage. A researcher should word instructions carefully and follow a prepared script so that all subjects hear the same thing. This ensures reliability. The instructions are also important in creating a realistic cover story when deception is used.

3. Post-Experiment Interview
    a. At the end of an experiment, the researcher should interview subjects, for three reasons:
        i. First, if deception was used, the researcher needs to debrief the subjects, telling them the true purpose of the experiment and answering questions.
        ii. Second, he or she can learn what the subjects thought and how their definitions of the situation affected their behavior.
        iii. Finally, he or she can explain the importance of not revealing the true nature of the experiment to other potential subjects.

f. A Word on Ethics
    i. Ethical considerations are a significant issue in experimental research because experimental research is intrusive (i.e., it interferes). Treatments may involve placing people in contrived social settings and manipulating their feelings or behaviors. Dependent variables may be what subjects say or do. The amount and type of intrusion is limited by ethical standards. Researchers must be very careful if they place subjects in physical danger or in embarrassing or anxiety-inducing situations. They must painstakingly monitor events and control what occurs. Deception is common in social experiments, but it involves misleading or lying to subjects. Such dishonesty is not condoned as acceptable and is acceptable only as the means to achieve a goal that cannot be achieved otherwise. Even for a worthy goal, deception can be used only with restrictions. The amount and type of deception should not go beyond what is minimally necessary, and subjects should be debriefed.