

Mass Communication Research and Theory

EDITED BY

Guido H. Stempel III

Ohio University

David H. Weaver

Indiana University

G. Cleveland Wilhoit

Indiana University

PUB 2003



*Boston • New York • San Francisco
Mexico City • Montreal • Toronto • London • Madrid • Munich • Paris
Hong Kong • Singapore • Tokyo • Cape Town • Sydney*

15

The Controlled Experiment

Maria Elizabeth Grabe

Bruce H. Westley

The controlled experiment is, when carried out rigorously, the most powerful method of seeking answers to research questions about cause and effect available to the social scientist. Indeed, the controlled experiment is our best—and very nearly only—way of investigating causal processes. In this sense experiments are integral to research aimed at understanding and *predicting* media effects. The survey method (see Chapter 13) also has the potential to produce predictive findings, but often results in *descriptive* research that describes or explains rather than predicts behavior or opinions. On the other hand, it is virtually impossible for an experimental study not to aim at some level of prediction. Also important to keep in mind is that experimenters test theoretical ideas on relatively small groups of people. Unlike survey researchers, experimenters are not interested in making statistical estimates about large populations. Yet, as we will explain later on, the results of an experimental study can be generalized based on logical rather than statistical inferences. A final point of comparison between the two methods: Experiments are used far less often in mass communication research than surveys. About 30 to 34 percent of research studies published in major communication journals employ the survey method; only 9 to 13 percent are experiments.¹

One might wonder how the complicated course of mass communication, which involves variance within every component of the process (channel, message, and audience), could be reduced to the required components of the controlled experiment. The logic of such reductionism is addressed first in this chapter, followed by experimental design matters, and practical information about procedures.

To illustrate the application of the method, let us begin with an example of a published experimental study by Franklin Gilliam and Shanto Iyengar.² As we have seen at

The first author would like to thank Annie Lang and Erik Bucy for their thoughtful suggestions on an earlier version of this chapter.

various points in this book, nothing can happen until a question presents itself. In this case, the question was: Do television viewers adopt the journalistic stereotype of criminals as predominantly African American as a filter for observing daily events? To investigate this question, the researchers first conducted a content analysis of local television news and established the presence of what they termed a "crime script" in which minorities are more likely than Caucasians to be depicted in the role of suspect. While important, this documentation of patterns in mass media content alone does not provide evidence of causal effects. The researchers therefore set out to test the impact of the "crime script" on viewer attitudes about crime suspects.

To isolate the phenomenon of interest, subjects were exposed to one of four different levels of the experimental manipulation. In the first condition subjects viewed a television news story in which the alleged perpetrator of a murder was an African American man. The second level of the experimental manipulation featured exactly the same news story except that the alleged perpetrator was a Caucasian man. The third experimental group viewed the same story but without information about the identity of the perpetrator. The fourth group of subjects comprised a control group—they did not view a crime news story. By introducing such distinct conditions, Gilliam and Iyengar were able to exercise an impressive level of control over the experiment. They kept all visual elements of the news stories constant while only varying the race or appearance of the alleged perpetrator in the first three conditions. Race in the news stories was varied by means of a "mug shot" of the perpetrator, which Gilliam and Iyengar digitally manipulated to change skin color. Thus the "African American" and "Caucasian" perpetrators featured in the stories were identical except for skin tone. In this way the researchers had reason to be confident that if there were differences in how subjects responded to the African American and Caucasian perpetrators, these differences would almost surely be the result of the perpetrator's race.

A cross-section of Los Angeles metropolitan area residents was recruited for the study through flyers and announcements in newsletters offering \$15 for participation in media research. Upon arrival participants filled out a questionnaire gathering information about their demographic status. They then watched a 15-minute newscast with one version of the crime story inserted toward the middle of the newscast. In an attempt to overcome the artificiality of watching the news in a laboratory, the researchers provided a viewing room that was furnished casually and subjects were free to browse through newspapers and magazines, snack on cookies, and talk with other participants. After viewing the newscast, subjects completed a questionnaire probing their attitudes about crime and news as well as asking free recall questions about the crime story.

The results showed that subjects were generally accurate in recalling whether a perpetrator was present in the crime story. Yet, those in the African American perpetrator condition were significantly more accurate in recalling that a perpetrator was mentioned. Most striking is the finding that over 60 percent of subjects in the condition that did *not* feature a perpetrator reported that they could recall having seen one. In 70 percent of these cases, the nonexistent perpetrator was identified as African American. The researchers concluded that what they identified in their content analysis as the "crime script" creates specific expectations about crime, which motivates viewers to fill in the gaps when they don't have information about the perpetrator. These leaps of judgment involve inferences that mirror

what television news has been shown to emphasize in crime reports, namely the portrayal of criminals as predominantly African American.

While certain aspects of this investigation could be criticized, most elements of a carefully controlled experiment are present. By taking measures immediately after the manipulation (the newscast) and by randomly assigning subjects to different treatment conditions (versions of the crime story), it was possible to say with some degree of certainty that the variation in the perpetrator's skin color must have caused the differences in subject responses. But what makes this a "controlled" experiment? In this case control was exercised through random assignment of subjects to different treatments, command over the environment in which the test was carried out, manipulation of the independent variables (skin color and presence of perpetrator), design of the testing instrument (the response questionnaire), and the use of a control group.

We note in passing that this is an **after-only** or posttest-only **design**. That is, no measures were taken before the manipulation as a basis for comparison with the measures taken after the manipulation. Subjects did provide demographic information about themselves before they were exposed to the newscasts but these responses were not used as a pretest of the independent variables. As we shall see, there are problems with after-only designs. In this case, though, the memory questions asked by Gilliam and Iyengar make sense only after the manipulation. Thus, there was no reason to employ a **before-after design**.

The experiment conducted by Gilliam and Iyengar was a *laboratory* experiment. In other words, it was conducted in a controlled environment. But we also make use of other, often somewhat less controlled, experimental designs, including the *field experiment*. The difference can crudely be described as a matter of locus. In the laboratory experiment, the investigation is carried out on the experimenter's own turf; the subjects come to the laboratory. In the field experiment, the experimenter goes to the subjects' turf. In general, the physical controls available in the laboratory are greater than those in the field. For that reason, statistical controls are often substituted for physical controls in the field. This chapter title refers to the *controlled* experiment and assumes that the quality of any experiment must be judged on whether the controls were adequate, not whether they were maintained in the laboratory.

Readers of Chapter 13 might ask how experimental research differs from *panel* studies, which survey the same subjects at two or more times. The difference lies in the manipulation. Panel studies are designed to identify trends. Whatever happens between survey 1 and survey 2 is eligible to be considered a potential "cause" of the differences observed. Consequently, panel studies can deal with causality only speculatively except where special conditions are met. When, in a now classic study, Tannenbaum and Greenberg³ sought to test the effects of watching the 1960 presidential debates between candidates Kennedy and Nixon, they did not entirely control the manipulation (i.e., the first debate) but they knew the debate would take place. So they obtained attitude measures from a sample of voting age adults beforehand, exacted a pledge to view the first debate, and left with the respondents a questionnaire containing further items to be completed immediately after the debate. Their field experiment used a before-after design without a control group and was not a panel study. In their case, the intervention of a media event was what made the difference. Thus, the manipulation in this study was the debates. Due to the lack of controls, it may not qualify as a fully controlled experiment, but it does meet the criteria for a field experiment.

This introduction to the use of experiments to test mass communication hypotheses reveals some of the options in experimental design. These nuances will be examined throughout this chapter.

The Logic of the Controlled Experiment

When performed correctly, the controlled experiment tests cause-and-effect relationships within a setting that permits maximum control over extraneous variation. The procedure allows the experimenter to observe the effect of one or more variables on another in such a way as to demonstrate that no other variable could have produced the same effect. It is probably the most sophisticated means of testing causal propositions and, represented by the independent variable, particularly suits efforts to conduct research that puts theoretically grounded propositions to direct test. How such propositions are generated from theory and how measures stand for concepts are discussed in Chapter 8.

The simplest way to show the effect of one variable (called the independent variable) on another (called the dependent variable) is to measure an attribute at time 1, then introduce a manipulation, and then measure the same attribute at time 2. As we shall see, it is not always that simple; but for purposes of illustration, let us imagine testing the proposition that the more a persuasive message arouses fear, the less effective it will be persuasively. The manipulation in this case will be the level of fear (low, medium, and high) contained in the message. "Manipulation" may sound chiropractic, but actually it is a crucial feature of the controlled experiment. The dependent variable, persuasion, will be measured by a set of attitudes toward something. In this case, let us imagine it is strength of belief that speed is the principal cause of death on highways.

To control for individual variation we will randomly assign subjects to treatments so that each subject has an equal chance of receiving one of the three message versions (high, medium, or low fear content). Ideally, the experimenter should make the verbal message identical in all treatments to hold it constant and introduce fear in another way, for example, showing graphic visual images. Take note that the visual fear induction is the manipulation, not the verbally persuasive message that is constant in low, medium, and high fear conditions. Subjects will first be tested for agreement with the statement that speed is the principal cause of death on highways, then exposed to one of the persuasive conditions (low, medium, or high level of fear appeal) and then tested again for agreement with the statement that speed is the principal cause of death on highways.

Inferring Cause

How does an experiment permit us to infer a causal connection between independent and dependent variables? In other words, how can we be sure that the independent variable (varying levels of fear in a persuasive message) caused the dependent variable (persuasion) to appear or change? Common sense would suggest that we search for a single cause for each single event and, when we find it, we can say A causes B or high levels of fear make persuasive messages less persuasive. But scientific endeavor does not assume single causes. Indeed, it searches for the *conditions* under which a particular phenomenon may

be expected to occur. We say, "may be expected" deliberately; in the social sciences we demonstrate the *probability* that a phenomenon will occur under certain conditions. We search for invariance or uniformity in behavior, but what we find are never certainties. Invariance is demonstrated in probability terms. When a set of conditions consistently (i.e., with a high degree of probability) produces the predicted results, we show that such a relationship, while not absolute, could not have been accounted for by chance. The best predictor variables are the ones that account for more variance in the dependent variable than any other set of predictors.

Instead of seeking *causes*, we search for necessary and sufficient conditions for particular behavior or phenomena to occur. If B can occur only in the presence of A, then A is a necessary condition of B. If B occurs whenever A is present, A is both a necessary and a sufficient condition of B. Causality cannot be demonstrated without (1) concomitant variation and (2) precedence or time order. Concomitant variation means that B varies consistently with A. That is what correlation demonstrates. But in showing concomitant variation, we are only showing that two variables vary together. We need more: Does change in A precede a change in B? When A and B vary together, and A precedes B in time (and the reverse is not also true), we have a basis for saying that A is causally related to B. Then, if we can be sure that A and only A produces B, we may say that B is caused by A. In social science it is often difficult to claim that A is the *sole* cause of B but it is important to rule out, as much as possible, any outside influences other than A that might cause the change in B.

This issue is further discussed in the following section. To summarize, there are three conditions necessary to infer causal relationships between variables:

- Cause precedes effect in time.
- Covariation between variables: Change in one correlates with the other.
- There is a high level of probability that nothing but the independent variable explains the change in the dependent variable because other possible causes have been controlled.

Control

At the heart of an experiment's effectiveness for discovering causal relationships is control over the order of events. But in demonstrating *a* causal relationship, we have not shown that the independent variable is *the* cause of the dependent variable. Other variables that share concomitant variation with the dependent variable when they precede it in time may also be shown to be causally related. For example, the pre-existing personal experience of a high-speed car accident might be causally related to the persuasiveness of a message, regardless of the level of fear appeal contained in the message. The goal becomes one of either showing that other explanations are spurious, for example that an extraneous variable's apparent causal relationship to the dependent variable arises out of its correlation with the independent variable, or seeking out the set of conditions that most consistently produces change in the dependent variable. If two independent variables together cause change in the dependent variable, we have a basis for a finding of multiple causality. The goal is then to detect what combination of conditions causes change in the dependent variable.

But just as we may search for multiple causes in the experimental setting, the experiment also permits us to control for extraneous or spurious variables. Control ideally means control over all other possible explanations. A well-designed experiment allows us to show that a set of conditions accounts for a large amount of the variance in the dependent variable and that no other causal agents could have accounted for that variance. In survey research we may be able to show that one variable is more strongly correlated with the dependent variable than are other variables in the survey. We may apply multiple regression, partial correlations, and other sophisticated multivariate techniques to investigate these relationships. But because all these variables occur and are measured together, we have no basis for inferring cause unless, in some way, we can show (or assume) that one preceded the other in time. An experiment must then be devised in a way to ensure that nothing but the independent variable or variables, or an interaction between them, could account for the behavior of the dependent variable, and partial and quasi-experiments may properly be called experiments only when they, too, meet these conditions.

Random Assignment to Experimental Conditions versus Random Sampling. The controlled experiment conjures images of the aseptic confines of a white-walled laboratory. But whether within a research laboratory or in some real-life setting such as a community center, randomization is often the critical means of control. As discussed in Chapter 9, random samples are required for meeting assumptions of probability statistics. In survey research this means that if we were to conduct a test of a hypothesis using such statistics, we must draw a random sample of respondents.

Although experiments and surveys both rely centrally on the principles of probability theory (with random selection as a key ingredient), the application of random selection differs markedly between these two research methods. As discussed in Chapter 13, when random selection is imposed on a carefully defined universe for a survey, researchers are able to make inferences about what to expect if samples, drawn under identical rules, were endlessly repeated. The condition of random selection is critical to the use of probability-based statistics. Yet, in the laboratory we do not rely on random sampling of subjects from a defined universe. Researchers are not fooling themselves to reason that college sophomores, who often participate in experiments, are somehow a *random selection* from the human race. Instead, they are treated as *instances* of the human race. When sophomores are randomly assigned to treatment groups (e.g., low, medium, high levels of fear) all characteristics not involved in the comparison—characteristics that would affect the outcome—are controlled by random variation. This means that characteristics are allowed to vary within groups but are prevented by means of random assignment from varying between groups and becoming sources of unpredicted and uncontrolled variance in whatever is being predicted for the dependent variable.

This methodological procedure offers more insight into what we meant in the opening paragraphs of this chapter when we said that experiments are effective in testing causal relationships but not effective in making population estimates. If the goal is to estimate parameters in an adult population (as nationwide surveys allow), it would be foolish to sample college sophomores, obviously. But in the laboratory we are not estimating parameters in a defined population. Instead, our goal is to carry out controlled experiments that test for relationships between variables. Experiments produce subject responses to stimuli, and

the insights derived from these responses are stable and generalizable under the conditions specified in the experimental design. In other words, what can be statistically inferred and generalized is that under the same experimental conditions the same cause will have the same effect. We cannot make statistically accurate generalizations of the findings to a larger population. Statistical generalization depends on random sampling that is typically not practiced in experimental research. Yet, that does not mean we are completely unable to make generalizations about experimental findings. In fact, by replicating experiments with different groups of subjects over time and finding the same results, experimental researchers are empowered to make nonstatistical generalizations, or logical inferences, about their findings.⁴

Because experimental researchers are usually less concerned about generalizing to large populations than survey researchers, *random sampling* from a population is not important to experimentalists. What is important is *random assignment* of subjects to experimental conditions.⁵ In fact, random sampling from a population is often impractical and counterproductive to the goals of an experimental study. Consider a study testing the impact of public service announcements about condom use on subject attitudes about the practice of safe sex. First, unlike survey participants who could respond to a questionnaire via phone, the Internet, or by mail, interpersonal contact between experimental subjects and the experimenter is unavoidable. Experimenters administer subject exposure and responses to stimuli in a relatively controlled physical environment. This makes the possibility of drawing a random sample from a national population impractical. Second, even if a random national sample were practically feasible, the range in subject demography and behavior captured in such a sample might not serve the goal of the inquiry. Including subjects who are in monogamous relationships or perhaps in age groups associated with little sexual activity is clearly not a reliable way to test the effectiveness of a public service announcement in shaping attitudes about safe sex.

A more productive approach is to sacrifice sweeping generalizability empowered by random sampling and purposefully select subjects who are not in committed relationships and who are sexually active. In fact, what would be gained by generalizing the findings in this case to the entire U.S. population? The pressing question is if the public announcement has the potential to impact the spread of a deadly sexually transmitted disease among a specific group of sexually active people. This discussion of the trade-off between wholesale generalizability and sample relevancy further unpacks what we meant in the opening paragraphs of this chapter by arguing that experiments are quite potent in testing causal relationships but not in making large population estimates.

Random Order of Experimental Stimuli. In addition to randomly assigning subjects to experimental conditions, experimenters also randomly order exposure to stimuli. Practically, this means that the series of media messages used as stimuli is organized into different sequences so that all subjects are not exposed to the stimuli in the same order. The specific goal of random order is to control for unknown consequences in the succession of tests and tasks in the experimental setting, namely primary and recency effects. Miller and Campbell⁶ have demonstrated that when there is a time delay in measuring attitudes after exposure to two persuasive messages, the first message is generally more effective (primacy effect). Yet, the second message becomes more effective when there is a time delay between

exposure to the two messages and the attitude measure is taken directly after the second message (recency effect). In this case the first message is forgotten while the second is still fresh in memory.

When ordering is arbitrary we can control for order effects by deliberate variation and randomization. For example, if subjects will be exposed to multiple television messages about speed-related accidents on highways, the experimenter could create at least two random orders of the messages. Half the subjects will view one order and the other half the other order, while making sure that who gets what order is determined purely on the basis of random assignment.

Purposive Sampling. There are additional means of obtaining control over variables that may produce sources of uncontrolled variation. Often demographic differences, even when experimenters do not hypothesize differences, are controlled to avoid their becoming artifacts of the study (i.e., uncontrolled factors that may offer alternative explanations of the outcome). For example, say we are concerned about (but do not wish to vary deliberately) the possibility that gender may be a significant variable in a predicted result. **Purposive sampling** will give us an equal number of men and women from the subject pool. What remains important is that the women and men are *randomly assigned* to experimental groups.

In studies using schoolchildren, researchers often group them by grade so that early, middle, and later grades can be compared. But in combining, let us say, first, second, and third graders into the early treatment group, it would be desirable to purposefully select an even number of students from the three grades to represent the early treatment group.

It should be added that by purposefully controlling demographic variables in this way the ideal distribution is obtained for studying the effect of the variable because it is equally distributed within the subject pool. Which brings us to the next method of control.

Including a Source of Potentially Uncontrolled Variance as a Variable in the Experimental Design. Another method of controlling the influence of variables is to treat them as explicit factors in the experimental design rather than controlling them. For example, the experimenter has the option to include gender as a factor in the experimental design. This enables observation of how gender might influence the dependent variables as well as interact with other independent variables in the study. This technique of control might make more sense after reading the section on factorial designs. Do take note, though, that adding a factor to a factorial experimental design complicates the design and generally requires collecting data from more subjects.

Validity

The unknown effect of pretreatment influences, the treatment itself, and dependent measures are potential threats to experimental validity. Campbell and Stanley⁷ have listed twelve of them. The first six sources of invalidity impede on the soundness of the experiment, referred to as *internal* validity. Researchers have to assess if they have indeed measured what they intended or claimed to have measured. Thus, when a study has internal validity there is sufficient reason to argue that the changes in the dependent variable are due

to the influence of the independent variable(s) and not some other extraneous influences. The last six items refer to *external* validity and pertain to whether the results of an otherwise valid experiment could be generalized (through logical inference) to human populations within the scope of the experimental design.

- *History* refers to uncontrolled events occurring between initial and posttreatment measurements. This is a problem particularly in conducting field experiments.
- *Maturation* is what happens to a subject, group, or community as a direct consequence of the passage of time that influences the study outcome without the researcher's knowledge or intention.
- *Testing* subjects before and after exposure to a treatment could have an unintended and undesirable impact on the dependent variable. For example, the effect of a pretest on the scores of the subsequent posttest is a concern in studies where memory is measured, especially when the same questions are used in the pretest and posttest.
- *Instrumentation* pertains to changes in either the technical instruments of measurement or the researcher conducting the experiment. When technically advanced equipment—such as physiology sensors or computer-controlled stimuli presentations—is used, there is the chance that equipment could become faulty over the course of data collection without the researcher's realizing it. This will introduce error variance into measuring the impact of independent variables and threatens validity. It is also quite likely that over time, the experimenter will change. Fatigue and bias, or even a more experienced outlook in conducting the experiment and making observations about subject behavior, could cause unintended variance. These inconsistencies in either equipment or researchers cause concern about experimental validity.
- *Statistical regression* of extreme scores toward the mean is a well-known concept to measurement specialists. In some experimental designs potential subjects are subjected to a pretest to help researchers identify a target subject group with extreme scores (either very good or bad) to participate in the experiment. Yet, chance factors play an important role in test taking that might artificially inflate or deflate a subject's performance. Because of chance the good or bad luck that might have played a role in the first test is not likely to be present in following tests and subjects will perform closer to their average capacity. Yet, the researcher might mistake the regression to the mean as the effect of the experimental treatment.
- *Bias in respondent selection* results when respondents are assigned to treatments or controls on any other basis than random assignment.
- *Experimental mortality*, or the loss of subjects on some basis other than random deletion, is often a problem in panel studies.
- *Interaction between maturation and selection factors* arises when, based on the selection of subject groups, subjects mature at different rates. Observed differences between the groups might then be attributed to this interaction between maturation and selection rather than exposure to the treatment. Comparing a group of freshman college students with high school graduates who are in their first-year of full-time employment clearly presents maturation differences. The changes that occur

in first-year college students might be very different from the transformation that takes place during the first year of full-time employment after high school. These two groups might appear quite comparable and homogeneous in that they both are comprised of, say, 19-year-old subjects who consume more than 7 hours of television per day. Yet, the maturation differences might account for more variance in dependent measures than the treatment.

- *Effects of testing*, both reactive and interactive, are potential threats to experimental validity. Reactive outcomes are effects on the dependent variable attributable to a premanipulation test; interactive effects relate to interactions between selection biases and the dependent variable. In these cases the testing or selection effects are said to “confound” the dependent variable, which means that the experimenter cannot know whether the results are the product of the predicted relationship or some uncontrolled cause.
- *Uncontrolled interaction between selection biases and the dependent variable* creates concern about validity. For example, failure to control gender differences in assigning subjects to treatments where gender is related to the behavior being predicted is likely to result in invalid findings.
- *Ecological validity* is threatened when the experimental arrangements are in some way unrelated to the general behavior being predicted. In some cases the laboratory situation is so complex or artificial that the results bear no similarity to the relationships the experimenter wishes to study.
- *Confounding caused by multiple treatments* can arise, of course, only in particular designs that call for more than one treatment for a subject or group and the effects of the first are not successfully erased.

Experimental Designs

This section will elaborate on key issues of experimental design. Three different types of designs will be discussed: “true” or classical designs, pre-experimental designs, and factorial designs. Before we get going on the discussion of experimental designs, it might be helpful to review five concepts central to most experimental designs:

- *Treatment*. This refers to the experimental stimuli that contain manipulations of the independent variable(s) of the study. The central goal of any experiment is to test the effect of a treatment on the dependent variables of the study.
- *Pretest*. This is a measurement of dependent variables administered before subjects receive a treatment.⁸
- *Posttest*. This is a measurement of dependent variables administered after subjects receive a treatment.
- *Experimental group*. The group of subjects who receive the experimental treatment.
- *Control group*. The group of subjects who do not receive an experimental treatment. Control groups are often subjected to pretests and posttests even though they don’t receive the treatment, to act as a point of comparison.

"True" Experimental Designs

A true experiment is marked by certainty that the experimental treatment influenced subjects. In this way the experimenter is able to gather evidence that the experimental treatment, and nothing else, is responsible for change in the dependent variables of the investigation. The three most widely used experimental designs are the pretest-posttest design with control group, the posttest-only design with control group, and the pretest-posttest design with additional control groups to deal with the effects of testing.

In a nutshell, the *pretest-posttest design with control group* assigns subjects randomly to at least two groups, administers a pretest, exposes one but not the other group to an experimental manipulation (treatment), and then posttests both groups (see Figure 15.1).

The *posttest-only design with control group* randomly assigns subjects to an experimental group and a control group and tests both groups after the treatment but not before. Only the experimental group is subjected to the experimental manipulation (see Figure 15.2).

The *four-group pretest-posttest design with controls* combines the procedures of the pretest-posttest control group and posttest-only control group designs. Solomon⁹ is often credited for creating this experimental design. Subjects are randomly assigned to four groups. Group 1 is tested before and after the manipulation. Group 2 is subjected to a pretest and posttest but receives no treatment. Group 3 is exposed to the experimental manipulation and posttested only. And Group 4 receives the same posttest as the other groups but is not subjected to either a pretest or the experimental manipulation (see Figure 15.3). This is the most elegant of the three most frequently used classical experimental designs because it provides evidence not only that the experimental group gained significantly more than did the control group (the one not manipulated) in whatever direction the theory predicted, but that this gain could not be accounted for by any interaction between pretest experience and the manipulation.

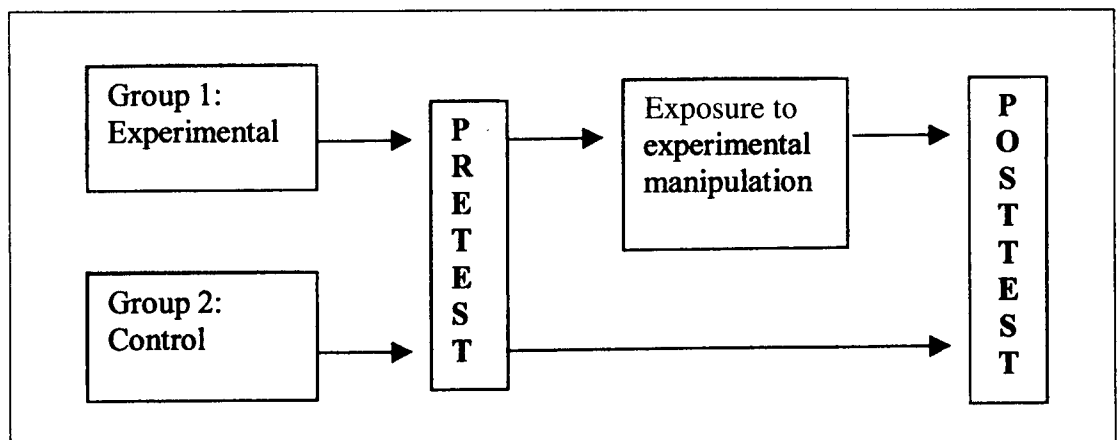


FIGURE 15.1. Pretest-Posttest Design with Control Group.

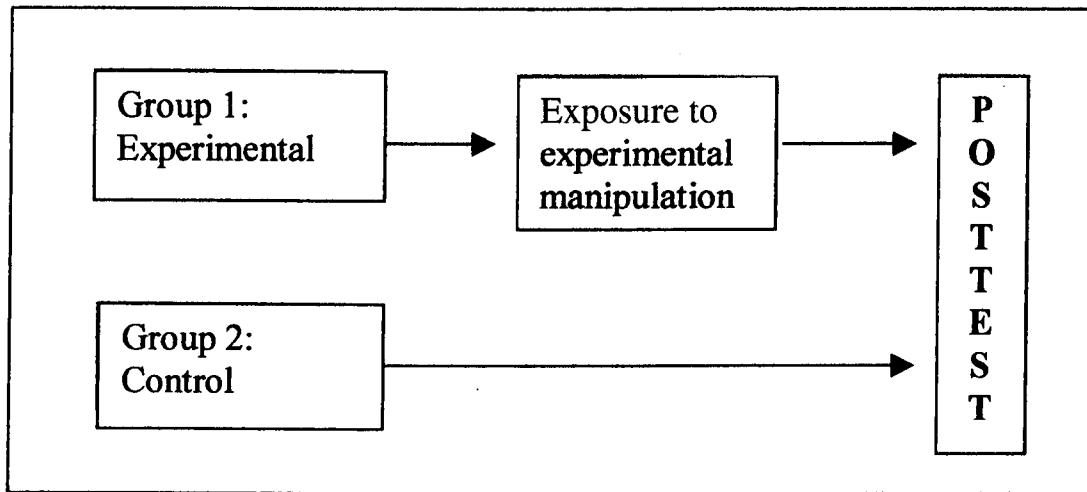


FIGURE 15.2. Posttest-Only Design with Control Group.

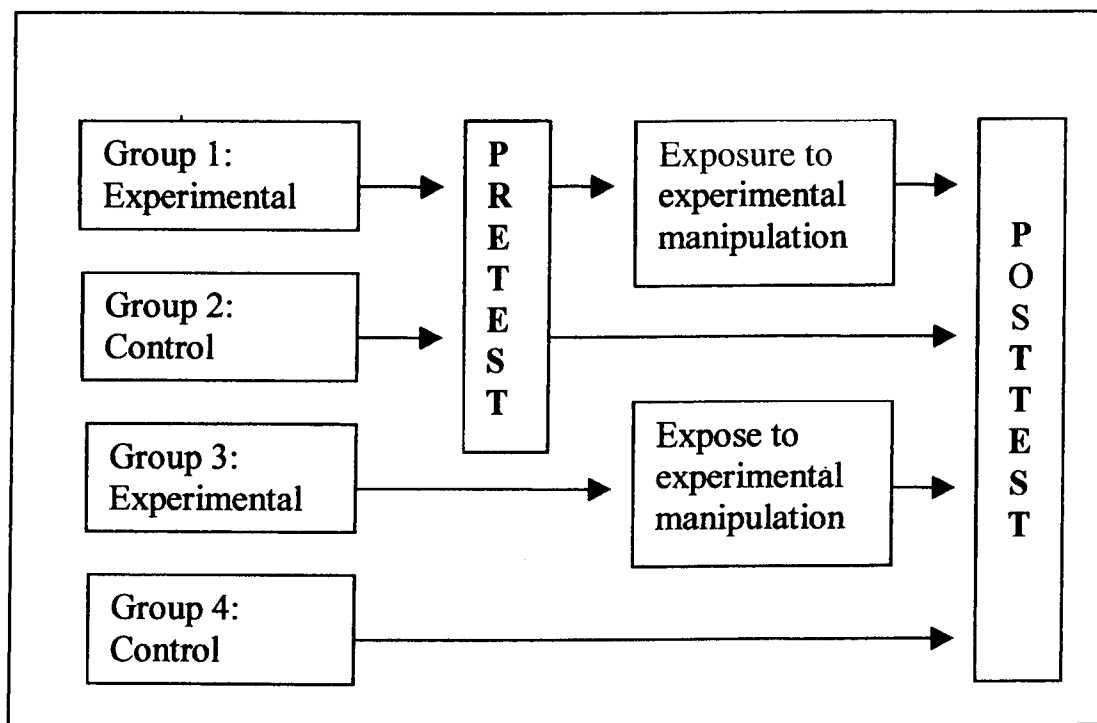


FIGURE 15.3. Solomon Four-Group Pretest-Posttest Design.

Pre-Experimental Designs

Pre-experimental designs are alternatives to classical experimental designs in situations where laboratory control is not feasible. They differ most prominently from “true” or classical experimental designs in that subjects are not randomly assigned to experimental and/or control groups. It is important to note that pre-experimental designs are not fully reliable for detecting causal relationships. While some experimentalists recommend against their use, they are still employed, most commonly in evaluation research. In other words, rather than refraining from evaluating the effectiveness of, say, a media campaign to make women aware of the importance of mammograms, pre-experimental designs can be employed to form some evaluation of the campaign.

In the *one-shot case study* (Figure 15.4) no pretest is administered and control groups are not used. Random assignment of subjects to groups is irrelevant because there is only one group. Because pretests or control groups are not included in this procedure, the experimenter cannot be sure if the treatment is the cause of what is measured in dependent variables.

The *one-group pretest-posttest* (Figure 15.5) design is a slight improvement on the one-shot case study. As the term suggests, one group is subjected to a pretest and a posttest. Yet, the absence of a control group and random assignment of subjects keep the researcher from having certainty about causal relationships. Without a control group there is no way of knowing if something other than the treatment that might have occurred between the pretest and posttest is responsible for the change in the dependent variables.

The *static-group comparison* (Figure 15.6) design consists of two subject groups and a posttest. It lacks a pretest and subjects are not randomly assigned to the two groups. The first group is exposed to the experimental manipulation, then receives the posttest. The second (control) group receives the posttest without the manipulation. Because there is no random assignment and pretest, the experimenter cannot be certain that posttest differences between two groups are due to the treatment because the group differences could have existed prior to the treatment.

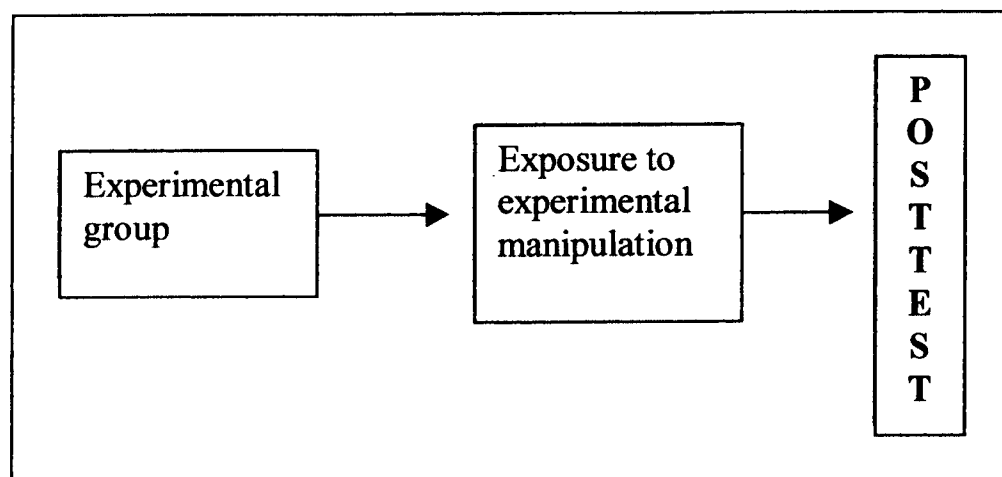


FIGURE 15.4. One-Shot Case Study.

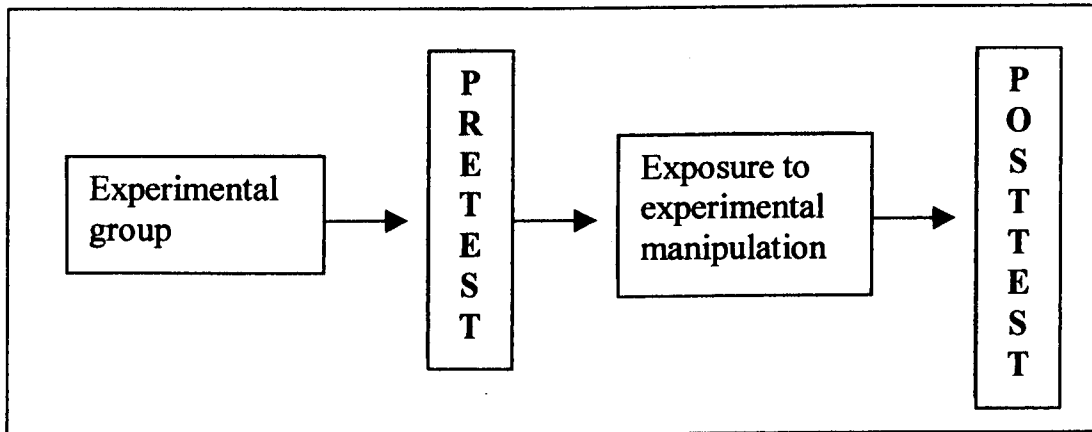


FIGURE 15.5. One-Group Pretest-Posttest Design.

Factorial Designs

When a researcher is interested in investigating the simultaneous effects of more than one independent variable on dependent variables, a **factorial design** is necessary. Let us use a specific example to explain a number of issues related to factorial designs. Say a researcher is interested in knowing which specific mass communication channels (television, radio, newspaper, or World Wide Web) are most effective in conveying memorable news information to people in two different age groups (Generation Y versus early Baby Boomers). The independent variables, known as factors, are channel and age group; the dependent variable is likely to be some measure of memory.

The design for such a study will be described in a research report as a 4×2 (four-by-two) factorial design. The channel factor has four levels (television, radio, newspaper,

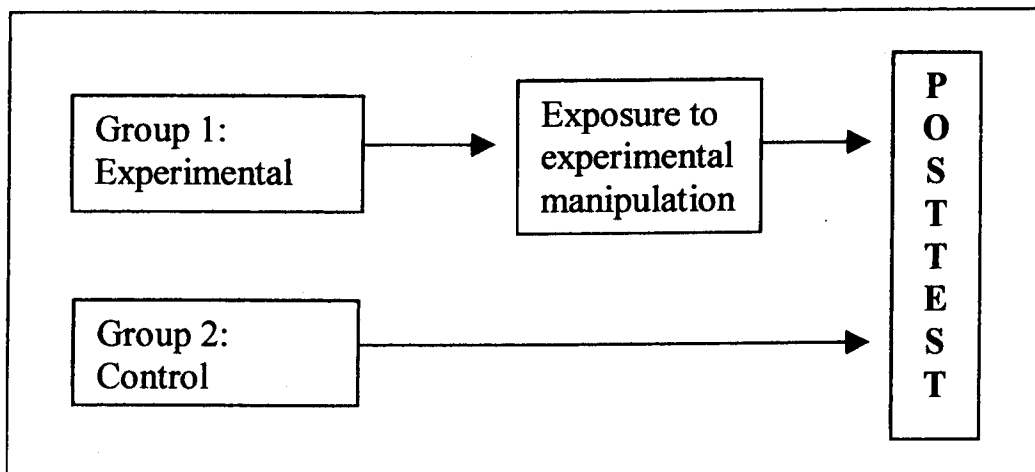


FIGURE 15.6. Static-Group Comparison Design.

World Wide Web). The age factor has two levels: Generation Y (young people between 15 and 20) and early Baby Boomers (people in their fifties).

The treatment consists of the different combinations of the levels *within* each independent variable. In this case eight different treatment combinations or cells (literally, 4 channel levels x 2 age levels) will be considered during data analysis (see Figure 15.7).

Independent Variables

Characteristics. There are a few dichotomous characteristics associated with independent variables that influence experimental design decisions. Independent variables could either be discrete or continuous.¹⁰ *Discrete* variables change in distinct categorical leaps. For example, in our sample study above, both the age and channel factors are discrete variables. Subjects are either Baby Boomers *or* part of Generation Y, and media channels are distinct: radio, television, newspaper, or the Web. Many demographic variables such as political party affiliation, gender, race, and marital status are discrete in that they contain distinct categorical subunits: Democrat, Republican, Reform Party, politically unaffiliated or male, female, transgendered, and so on. *Continuous* variables, on the other hand, change in steady increments on a continuum from high to low. For example, a television message could range from emotionally arousing to dull, or from emotionally positive to negative. In experimental research continuous independent variables are usually transformed into discrete variables with fixed subcategories. Indeed, the age factor in our sample study was operationalized to contain two categorical levels (Baby Boomer and Generation Y) instead of the natural

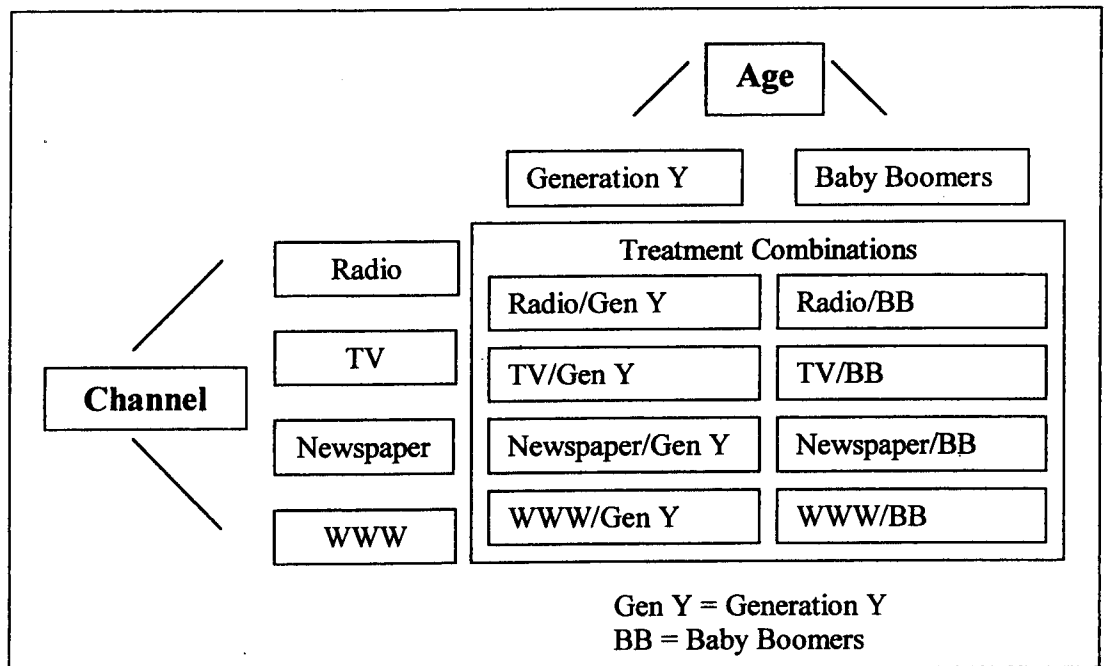


FIGURE 15.7. Factorial Design.

continuous range in age (e.g., 0 to 100). Experimental researchers do this because discrete independent variables are required for factorial designs. Every independent variable must have distinct levels that can be tabulated as shown in Figure 15.7.

Independent variables can be either internal or external to subjects. *Internal* variables originate from subject characteristics. Researchers cannot create these variables; they emerge from observations about subjects. In our hypothetical study of memory for news stories, the age factor is an example of a variable internal to subjects. Not all variables internal to subjects are easily observable. Cognitive and emotional dimensions such as sensation seeking, political sophistication, and fear of crime cannot be assessed based on the physical appearance of subjects. When cognitive or emotional attributes are used as independent variables, researchers must measure potential subjects' ranking on these variables to determine how much or little they have of these characteristics before selecting them to represent a specific level of an independent variable. *External* variables originate from things outside subjects. In mass communication research external variables usually comprise some dimensions of media messages that are manifested in stimuli subjects are exposed to. The media channel variable is an example of a factor external to subjects.

After deciding on the *experimental factors* (e.g., age and media channel), researchers also have to consider adding *control factors* to the design. Clearly, in the above example the two experimental variables, age and channel, are the two factors that drive this investigation. Their influence on memory is at the center of *theoretical* interest in our hypothetical study. On the other hand, control variables are *methodological* tools to improve the accuracy of observations about the experimental variables. Two control variables in particular could be added to most factorial designs as a check on systematic error. These two control variables are message repetition and order. These are also important issues to consider in other experimental and pre-experimental designs discussed earlier in this chapter. The following discussion focuses on how to implement control variables in factorial designs.

The *message repetition* factor refers to how many messages, or examples of the experimental manipulation, subjects ought to be exposed to. In other words, in our memory study, how many news stories should subjects view, read, and listen to in four different channels (radio, TV, newspaper, and WWW)? Related to this issue, should each story be presented in all four media channels or should different news stories be selected for each of the four media channels?

There are no quick answers to these questions. Stanford Professor Byron Reeves¹¹ points out that media messages rarely, if ever, exemplify only one thing. It is therefore virtually impossible to completely isolate the single feature of interest to a researcher in an experimental stimulus. At first, the implication of this insight appears devastating: the natural variation between media messages introduces severe threats to internal and external validity of experiments. Internal validity is compromised because we are unable to completely isolate the experimental manipulation, and this makes us unsure about which part of the media message is affecting the dependent variable. In the case of our sample study we have reason to question whether an uncontrolled message characteristic such as violent content, rather than media channel, might be responsible for affecting memory for news.

This dilemma offers a clue as to why it is more desirable to use the same story across versions. Put differently, it is preferable to use radio, television, newspaper, and Web versions of the same story—for example, a lawsuit about the patent for the drug Prozac. If we

test the same message across media channels, natural variation between messages is less of a problem than when we test different messages across different media channels. Yet, it is not always possible to manipulate the same message to represent different versions of the independent variable. For example, we may not be able to find news reports of the Prozac story, or any other stories, in all four media channels. Even if we are able to find reports of the story in all four media channels, they may differ so substantially that we are not able to do what we set out to achieve: reduce variance across messages.

If we truly want to control natural variation between messages, the solution then seems to create four channel versions of the Prozac story, using exactly the same information and making the stories roughly equal in length. Even this apparent solution poses problems, specifically related to ecological validity. Broadcast writing differs dramatically from print writing. Thus do we take the newspaper story and add visual material to it to create the television story? Or do we take the television or radio story scripts to represent the newspaper story? After all, the four versions of the Prozac story should each be typical of the four different media channels that they represent if we want to be able to generalize our findings to radio, television, newspaper, and the Web.

In other studies, searching for or creating the same story across versions might not be an option. For example, if we want to compare the effects of "real" and fictional television violence on physiological arousal, it would be impossible to manipulate the same message to represent both fiction and nonfiction versions. Instead we might represent "real" violence by selecting material from Fox network shows such as "The World's Scariest Police Chases" and "When Animals Attack" and choose fictional violence from cop shows such as "NYPD Blue" or "Nash Bridges."

This discussion should reveal that often there are not clear-cut answers to design issues. That is precisely why the process of designing an experiment requires creativity combined with logical thinking and an understanding of the method. Our advice is thus to carefully contemplate the options and justify design decisions in the methods section of the research report—and be able to defend your decisions on scientific grounds.

Let us return to the discussion about the threats of natural message variance to experimental validity. We have just elaborated on the threats to internal validity. External validity in such situations is also compromised because variance between messages makes generalizability of findings difficult. Indeed, these threats to internal and external validity hold true if subjects are exposed to only one example of the experimental manipulation. If subjects are exposed to multiple messages, on the other hand, each containing some uncontrolled differences as well as the controlled experimental manipulation, the variance *between* messages is reduced to random error, which is less damaging than systematic error. Using multiple messages benefits experiments in which the same message is used across versions and when different messages are chosen to represent different versions of the independent variable. It cannot be emphasized enough that using multiple messages is key to rigorous experimental investigations of mass media effects. In fact, some experimentalists argue that we should be more concerned about sampling messages than subjects.¹²

The response, then, to the question about how many messages subjects should be exposed to is: as many good examples of the experimental manipulation as the researcher can find and the subjects can reasonably attend to. Typically, fatigue and boredom set in after about one hour of viewing and actively responding to dependent measures.¹³

In a factorial design multiple messages are treated as a control factor, usually referred to as the *message repetition factor*. In the case of our memory study, let us assume that after considering fatigue and availability of suitable messages, we select three different stories to represent each of the four levels of the independent variable, media channel. Thus the four channels will be represented by a total of twelve different stories, which will be collapsed (three per channel) during data analysis. Their impact—in combination with the other independent variables—will be investigated on the dependent variable. The previous experimental design for our hypothetical study only featured experimental factors. Now we will add the first control variable, message repetition, to the design: age (2) x channel (4) x message repetition (3).

The second control factor to consider is order. To assess the impact of primacy and recency effects as discussed earlier in this chapter, multiple experimental messages must be presented in a different sequence to different subjects. If we expose all subjects to the same order of experimental messages we cannot be certain that the independent variables, and not the sequence in which stories were viewed, is affecting the dependent variables. How many different message orders should be created? Again, there is no convenient formula to determine this, but clearly more than one order is necessary if we want to control order effects. Some researchers are comfortable using as many orders as the number of levels in the message repetition factor. Sometimes it is too expensive or practically impossible to meet that standard. How are the message sequences determined? Researchers typically use what is referred to as semi-random procedures to compile message sequences for each order. By semi-random we mean that rules are set to ensure the suitability of each order before standard random assignment procedures are followed. True random assignment often results in orders that resemble each other to a small degree, thereby introducing bias and diminishing the control over order effects. For example, two orders could have the same messages in positions three and four, which might create an uncontrolled primacy or recency effect in the experiment. If random procedures produce sequences that will not be effective in controlling order effects, sequences are purposefully altered to enhance control, making this a semi-random procedure. In the case of our memory study, let us assume we decided to create three (same number as the levels in the message repetition factor) message sequences. By including the order factor in the experimental design we could statistically assess order effects during data analysis. The full design with both control factors will look like this: age (2) x channel (4) x message repetition (3) x order (3). Subjects from both age groups will be randomly assigned to one of three stimuli orders. Figure 15.8 shows how three orders could be created in which the twelve stories are semi-randomly assigned. Note how each order has all twelve stories in unique sequences: (1) subjects will not be exposed to the same channel in two consecutive stories; (2) each order starts and finishes with different stories and different channels, and (3) looking across the order rows it is clear that stories are not duplicated in any position.

With the conceptualization of the independent variables for our memory study now completed, including the experimental and control factors, there is another design issue that deserves attention: who will receive what level(s) of the independent variable. The most typically considered options are between subjects, within subjects, and mixed experimental designs. Although experiments are often thought of as either between or within subject designs, mixed designs are quite common and often the only choice.

Order 1	Order 2	Order 3
WWW story 3	Radio story 1	TV story 2
Newspaper story 1	WWW story 2	Radio story 3
Radio story 2	Newspaper story 3	WWW story 2
TV story 1	WWW story 3	Newspaper story 3
WWW story 2	TV story 2	Radio story 1
TV story 3	Radio story 3	WWW story 1
Newspaper story 2	Newspaper story 1	TV story 3
WWW story 1	TV story 3	Newspaper story 1
Radio story 1	Newspaper story 2	Radio story 2
Newspaper story 3	TV story 1	WWW story 3
TV story 2	Radio story 2	Newspaper story 2
Radio story 3	WWW story 1	TV story 1

FIGURE 15.8. Example of Three Random Stimuli Orders.

Using the *between subjects* design, also called the *independent groups design* (Figure 15-9), subjects are assigned to different versions of the independent variables. In our memory study, Baby Boomers would be divided into four groups and each group would be exposed to only one of the media channels: radio, television, newspaper, or the World Wide Web. The same procedure would be followed for the Generation Y subjects. Thus, each subject would see three news stories in only one channel. Both the age and the channel factors are treated as between subject factors; in other words, they do not overlap or repeat for any subject.

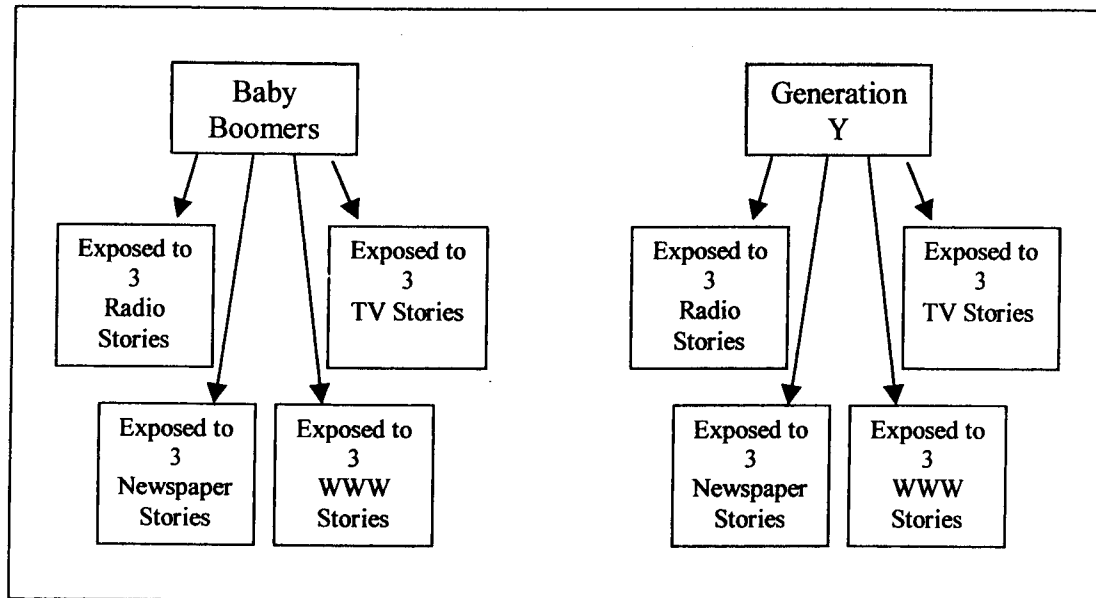


FIGURE 15.9. Between Subjects Design.

Using **within subjects designs**, also known as *repeated measures designs*, the same group of subjects is exposed to all versions of the independent variable. In other words, the responses of the same group of subjects are measured repeatedly. Whereas the independent groups design compares separate experimental groups, the repeated measures design exposes all subjects to all levels of all independent variables. In this design each subject serves as his or her own control. Put differently, each subject is compared to him- or herself in the various experimental conditions. Yet, some variables simply cannot be tested using a within subjects design. Demographic variables, such as the age factor in our sample study, are good examples of variables that can only be investigated using between subject comparisons. The channel factor, on the other hand, could be subjected to either between or within design analyses. Clearly, in a repeated measures investigation of memory, subjects cannot respond to a test first as a Baby Boomer and then as a Generation Y member. On the other hand, subjects can respond to four different media channels in a repeated measures design. We can thus apply a repeated measure to the channel factor, but not the age factor, in our memory study by exposing subjects in the two different age groups (Baby Boomers and the Generation Y group) to the twelve news stories in all four media: radio, television, newspaper, and the World Wide Web. This combination of repeated and independent group measures is referred to as a *mixed design* (Figure 15.10).¹⁴

There are three benefits to using the within subjects design. First, this design requires fewer subjects than independent groups designs, which makes it ideal in situations where the subject pool is small or difficult to recruit. Second, using relatively few subjects makes within subjects studies more efficient and cost effective than between subjects designs. Third, within subjects designs are more sensitive than between subjects designs in detecting significant differences caused by the independent on dependent variables. By

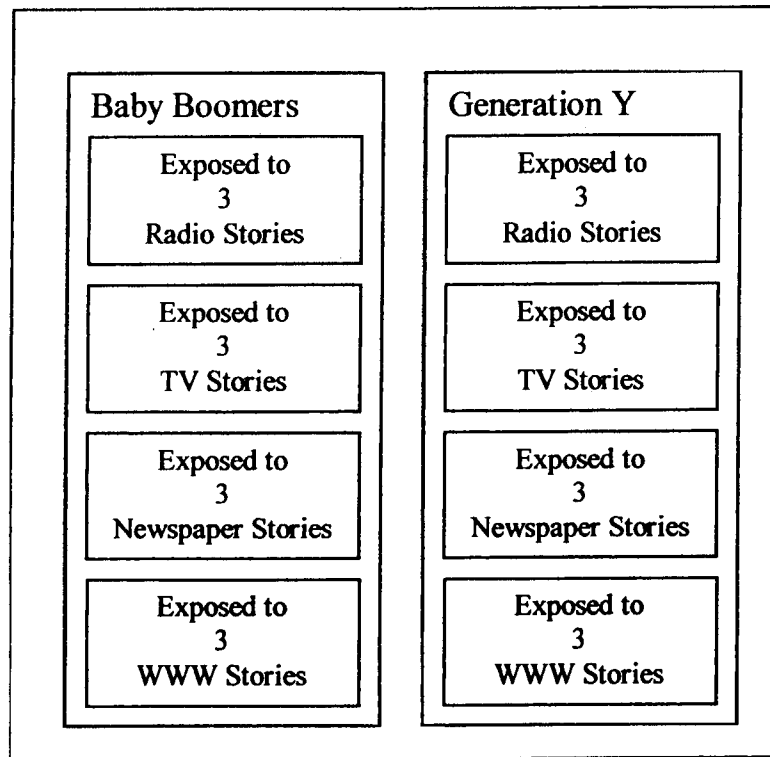


FIGURE 15.10. A Mixed Design.

sensitive we mean there is less error variation in repeated measures designs. Consider the nature of the repeated measures design: The experimenter measures the impact of all the experimental conditions on *the same group of subjects*. In the independent groups design, researchers measure the impact of experimental conditions on *different groups of subjects*. Logically, there is more variation *between different people* than *within the same subjects*. Thus the repeated measures design has less error variation and is particularly useful in studies where the independent variables might have subtle or hard-to-observe effects on the dependent variables.

Yet, the repeated measures design is not always an appropriate choice. Repeated exposure to messages could produce sensitization effects that are especially problematic when memory measures are administered. Let us consider our hypothetical experiment again. If we had produced three news stories in four media channels instead of using twelve unique stories to present the four different media channel versions, subjects would have been repeatedly exposed to the same information presented in different channels. Being repeatedly exposed to the same experimental stimuli increases rehearsal of the information and is likely to improve memory. Thus a measurement of memory for news may not be attributed to the manipulation of the independent variable but rather to repeated exposure to the experimental stimuli.

With the knowledge of what between, within, and mixed designs entail, let us reconsider the status of each factor in our sample study: age (2) x channel (4) x message repetition (3) x order (3). Age is, and can only be, a between subjects factor because it is a