Chapter 13

Principles Of Research Design

Research designs can be classified into three broad categories, according to the amount of control the researcher maintains over the conduct of the research study. The three general categories are experimental research, field research, and observational research. Each of these categories varies on two important characteristics: internal validity and external validity.

Research Design and Internal and External Validity

The terms internal and external validity must not be confused with measurement validity, which was discussed in Chapter 7. Instead, these terms refer to the overall validity of a research study as discussed in Chapter 4, not to the measurement of the concepts used in the research.

Internal validity describes the ability of the research design to unambiguously test the research hypothesis. An internally valid design accounts for all factors, including those which are not directly specified in the theory being tested, which might affect the outcome of hypothesis tests. It insures that these factors do not confound the results.

Since it is impossible for any single research design to account for all such potentially confounding factors, we must speak of better or worse internal validity, not of perfect validity. But designs with higher internal validity will, for example, control or account for the actions of variables which might produce spurious relationships. They will use representative samples, so that subject or group differences will not be confused with the action of independent variables. In general, they will eliminate more of the alternative explanations of research findings (those which con-

tradict the theory being tested) than will experimental designs with weak internal validity.

External validity refers to the generalizability of the research, that is, the ability of its conclusions to be validly extended from the specific environment in which the research study is conducted to similar "real world" situations. The results of an externally valid study can be used to predict the behavior of the theoretical constructs outside the laboratory or data center. Externally valid research with generalizable conclusions is obviously more valuable than externally invalid research, whose conclusions are restricted to specific research settings.

Experimental Research

The first category that we will examine is experimental research. In this kind of research study, the researcher controls the setting in which the research is conducted (the "laboratory") and he also manipulates the levels of the independent variable or variables, and follows this by observation of the corresponding changes in the dependent variable or variables.

By controlling the surroundings in which the research is conducted, the researcher can eliminate some environmental conditions that might confuse the results. This control improves the internal validity of the research study. For example, a researcher studying the effects of music on children's learning from educational videotapes would probably want to show the tapes to the experimental subjects in a quiet room. Furthermore, it is likely that she will use the same, or very similar, rooms, equipped with similar furniture, lighting, and potentially distracting items like books and toys. By insuring that all subjects see the tapes under the same conditions, she can eliminate the possibility that learning (or lack of learning) is due to factors other than the experimental videotapes. If the same tapes were shown in uncontrolled settings like individual homes, learning for some children might be disrupted by distracting brothers and sisters, the presence of toys, etc. The effects that these environmental factors have on learning will obscure the effects that are the result of the use of music. These are the effects that the researcher really wants to observe.

By directly manipulating the levels of the independent variables in an experimental design, the researcher can meet all the conditions for establishing a relationship between variables, as outlined in Chapter 4. This manipulative control will also improve the internal validity of the study, as it allows the experimenter to predetermine the time sequence of events, and to insure that the independent variable takes on a wide enough range of values (i.e., has enough variance) that an unambiguous test of the hypothesized relationships can be made.

Suppose the researcher studying children's learning creates two videotapes, one using music at critical points in the presentation, and a second which does not use music, but is otherwise identical. She has manipulated the nominal independent variable (presence or absence of music), while controlling for other possibly confounding factors. By using the same tape for both groups, with only the music track modified, she has insured that the effects of other content features of the tapes, like the narrator, the script, or the illustrative visuals, are constant for viewers of both tapes. These factors will then produce identical effects on viewers of either version of the tape, so the effects of these features will not be confused with the effects of music.

The researcher then selects two different groups of children, using appropriate random sampling techniques, and shows one of the tapes to each group. Several days later, she returns to measure the children's recall of the material on the tape. Using some variant on the basic statistical methods outlined in the previous chapters, she tests a directional comparative hypothesis which states that "Material presented with a musical background will be recalled at higher levels than will the same material presented without music".

In this simple experiment, the researcher has met the basic requirements for testing a hypothesis:

1. The independent variable is present in at least two levels (presence and absence of music).
2. The two groups can be treated as equivalent within the limits of sampling error, since their members were chosen randomly. This eliminates any systematic effect from variables which were not measured as part of the research, like the effect of differing academic abilities or attention spans for different children. Since the groups are randomly chosen, each should contain a similar number of high ability and low ability children, children with long and with short attention spans, etc.
3. This allows the researcher to conclude that any difference seen in the dependent variable

for the two groups must have been produced by the different levels of the independent variable. This establishes covariance.

4. Since the dependent variable is observed after the presentation of the independent variable, temporal priority between the cause variable (the independent variable) and the effect variable (the dependent variable) is established by the researcher.
5. Since the unit of analysis is the individual child, the requirement for spatial contiguity is satisfied.
6. If the researcher has provided a good theoretical linkage which relates the presence/absence of music to recall, the final condition for causality, necessary connection, is established.

The control that an experimental study affords a researcher helps to establish strong evidence for causal connections between the independent and dependent variables. But it can also cause some problems in generalizing the results of the research to the outside world. The very strong control which improves the internal validity of the experiment can sometimes damage its external validity.

Suppose the experiment described above shows that children recall more of the material from the videotape that used music. Most of the factors that could produce a spurious relationship between music use and recall are controlled by the experiment. The program content is constant in both groups, the groups are equivalent because of random assignment, the level of distraction from the environment is constant for both groups, etc. This is an experiment that is strong on internal validity. It is therefore very tempting to generalize its results to all educational videotapes for children, and to prescribe the use of music to enhance learning.

Unfortunately, this experiment happens to be somewhat weak in external validity, so such a prescription may not be warranted. The conditions under which the children actually watch television are very different from the experimental conditions. For example, children often have low levels of attention to television when they are viewing in their homes. In the experimental setting, the attention level may have been much higher, due to the experimental instructions given by a high-authority figure (the researcher says "please watch this tape"), or by the lack of familiar distractions like favorite toys, siblings, etc. As a result, learning from tapes which use music may be very similar to those which do not use it, if overall attention levels are low—that is, not much will be recalled from either kind of tape. If this is the case, adding music to educational tapes will be a waste of money, even though, under the right conditions (like those in the experimental setting), the researcher can show an positive effect of music.

The experiment also uses the tape of a single educational presentation. While the conclusions about the use of music may be correct for this presentation (and probably are, because of the high internal validity of the experiment), the results may not generalize to other teachers, or other topics. Again, the control that can be exerted over the experimental material by making sure it is identical in all experimental conditions carries the cost of limiting the external validity of the conclusions.

The issue here is one of the costs and benefits of controlled observation. A good experimental design will control for potentially confounding factors, whether they are explicitly identified or not. The researcher in the videotape experiment does not have to define all the possible variables that might affect recall (such as attention, distraction, the inherent appeal of the material, etc.), because she can be assured that they are all present in equal amounts in both experimental groups. Since they are, they can't bias the results.

But these variables do affect recall in realistic situations. The control that experimental designs impose over these outside variables may actually obscure the realistic operation of the system of variables in the real world. Figure 13-1 illustrates how this can happen. Variables X and Y are investigated in an experiment, which controls for the effect of an outside variable Z. This variable is negatively related to X and positively related to Y. The numbers represent the strength of relationships (they might be correlation coefficients, for example).

In the experiment, the direct effect of X on Y is found to be +.25. The variable Z will not enter into this finding, since its effect will be controlled by the experimental design. But in the realistic situation, X will affect Z at a -.80 level, and Z will then affect Y at a +.50 level. So in addition to the direct +.50 effect of X on Y, a 1.0 unit change in X makes a -.80 unit change in Z. Half (.50) of this -.80 unit change is passed on to Y, making the contribution of this path of influence equal to -.40. The total effect of X on Y, in the realistic situation, is then made up of two components: the +.25 direct effect, and the -.40 indirect effect via variable Z. The net effect of X on Y is then -.15 in reality, while
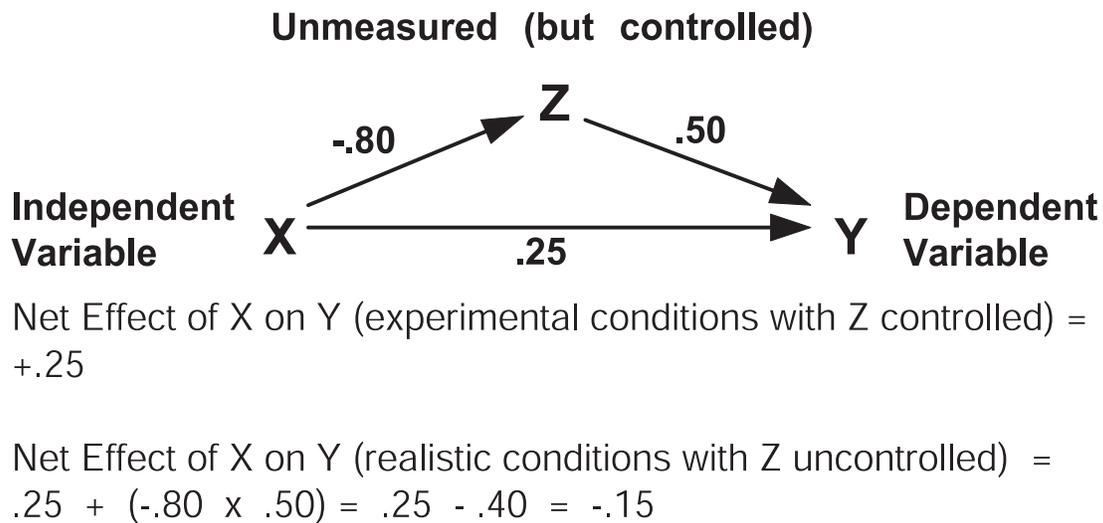
Unmeasured (but controlled)

$$Z$$

-.80   .50

**Independent
Variable**   **X** ————————————→ **Y**   **Dependent
Variable**
                              .25

Net Effect of X on Y (experimental conditions with Z controlled) =
+.25

Net Effect of X on Y (realistic conditions with Z uncontrolled) =
.25 + (-.80 x .50) = .25 - .40 = -.15

FIGURE 13-1   **Effect of Experimental Control on External Validity**

the experimental results will indicate that the effect of *X* on *Y* is +.25! And the experimental conclusion of a +.25 effect is correct.

However, the +.25 effect is not generalizable, and thus the experiment has exhibited poor external validity due to the very control which produces good internal validity.

The solution to this situation is to identify and explicitly include the relevant variables in the experiment. If *Z* is theoretically and operationally defined and is included as part of the experimental design (with the addition of two new hypotheses: *X*—>*Z* and *Z*—>*Y*), then the correct net effect of *X* on *Y* can be found. This solution calls for measuring or manipulating all relevant variables. Of course, this will increase the complexity and cost of the research, once again illustrating the fundamental truth that obtaining higher-quality information requires added costs and added effort.

An alternative approach is described in the next section. In this approach, the variables are observed as they operate in the "real world". External validity is improved without substantially increasing the complexity of the research, but only at the expense of decreasing the internal validity.

Field Research

The second major category of research is field research. In this kind of research setting, the researcher retains control over the independent variables, but conducts the research in a natural setting, without any control over environmental influences.

For instance, suppose that a researcher is interested in the ability of a communication training program to reduce communication anxiety in persons who must give speeches or public presentations. The researcher, who is employed by a large corporation, creates two randomly selected groups of subjects by drawing samples from a sampling frame which is a list of all employees of the organization. Each person in both groups is asked to fill out a questionnaire. The questionnaire contains the information for operationalizing the dependent variable "communication anxiety". It asks for self-reports of the person's apprehension immediately before giving his or her most recent presentation and the discomfort he or she felt while speaking at that time. The independent variable, "training program", is operationalized by creating a program of study and practice in public speaking and use of audio-visual materials. This variable then has two levels (presence or absence of training), and each of these levels is applied to one of the groups. That is, one group receives the training program, while the other does not. This latter group is often called a control group. The directional comparative hypothesis being tested is this: "Those who receive Communication Training will have reduced levels of communication anxiety compared to those who did not receive Communication Training". If the mean anxiety levels for the group which received training is significantly lower than the mean for the control group, the researcher will conclude that the hypoth-

esis was supported.

The researcher waits several months, and then asks each group to again fill out the same questionnaire. He is presuming that both groups will have made some public presentations during the interval. If the training program worked, those in the group which receive the training should have felt more comfortable in speaking than those in the control group.

Note the difference between this kind of research setting and an experimental setting. In field research, the conditions under which the effects of the independent variable are observed are not under the researcher's control. Although the researcher still exerts control over the independent variable (by creating the training program and controlling who is exposed to it), he does not control the setting in which the independent variable exerts an effect on the dependent variable. Different subjects may have had very different public communication experiences. One may have had to give a large number of presentations during the months between the two administrations of the questionnaire, while another may have had few or no opportunities to put the training into practice; some persons may have had to give presentations to large audiences, while others spoke only to small groups, etc. Because of this variation, the researcher must expect that some variation in the dependent variable is due to uncontrolled factors in the field research setting. These variations should not bias the results, however, as the randomly selected groups should both have equivalent numbers of persons with frequent and infrequent presentations, large audience and small audience experiences, etc. But the strength of covariance between the independent and dependent variables will be reduced by the random error that is introduced, and this will make it harder to confidently state that the condition of covariance has been met, i.e., to obtain statistically significant relationships between the independent and dependent variables. This is because there are variables other than the independent variable acting on the dependent variable, and their effects may mask the effect of the independent variable.

Given this penalty, why would a researcher ever choose to do field research, rather than experimental research? The basic reason has to do with the generalizability or external validity of the research. Field research, because it occurs under natural conditions, is often more informative than pure experimental research.

The researcher in our example could have used an experimental design, by requiring that all persons in each of the groups give a presentation on the same subject, to the same audience, in the same room. This control over the research setting would remove the random error due to differences in subjects' public communication experiences, and would enhance the researcher's ability to answer the relatively narrow question posed by the research hypothesis. As we saw in the previous section, this kind of control improves the internal validity of the research.

But the researcher probably wants to know more than simply whether the hypothesis is supported or not; he also wants to know if the effect which he has hypothesized works under realistic conditions—those conditions outside the rigid control of the experimental laboratory. For this reason, he chose to trade some of the research power of an experiment for the more general test of the hypothesis in the setting to which the results are eventually to be generalized.

In the example, the researcher might find that the training program significantly reduces communication anxiety in the experimental setting. But the experiment only tests the effect of communicator training for a single kind of presentation, to a single kind of audience. To generalize the results of the experiment to all kinds of presentations, with all kinds and sizes of audiences, requires some strong assumptions: 1) all presentations are equivalent to the one required in the experimental procedure; and 2) that all audiences, regardless of size or makeup, are equivalent. The researcher may be quite reluctant to make these assumptions.

Of course, the researcher could modify the experimental design to add different conditions which better represent the complexities of the real setting. He might require the subjects to give different kinds of presentations, to different sizes of audiences, speak in large and small halls and conference rooms, etc. But the research design would then be much more complex, and possibly too expensive to complete. And there would still be no assurance that the researcher had adequately reproduced all the conditions that a large number of public speakers were likely to encounter in the "real world".

The researcher can regain some of the lost sensitivity to the effect of the independent variable in a field experiment by measuring the "outside" variables and using statistical control (this is covered in more detail in Chapter 4). The researcher still manipulates the independent variable, but uses statistical techniques mentioned in Chapter 19 to isolate or control the effects of measured

"outside" variables, and that removes them from the category of unknown error. This is illustrated in Figure 13-2. If the researcher does not measure variable Z, its effect is lumped with all others in the composite group called E. When he explicitly includes Z in the field experimental design, its effect can be isolated from that of X and of all other E variables. This gives a more accurate estimate of the true strength of the $X \rightarrow Y$ relationship.

Measured "Outside" Variable

Z

Manipulated Independent Variable   **X**   **Y**   **Measured Dependent Variable**

E

Net Effect of All Other Unmeasured "Outside" Variables
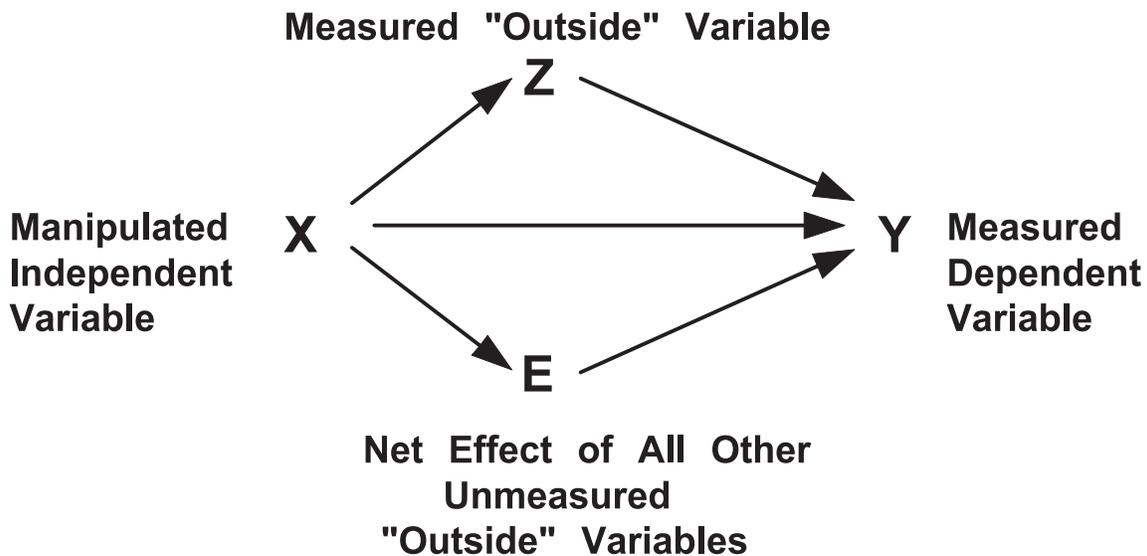
FIGURE 13-2   Manipulated and Statistical Control in Field Research

Observational Research

There are many instances in which the researcher can control neither the independent variable nor the research setting. In this situation, the researcher is limited to measuring, rather than manipulating the independent variable. Like field research, observational research designs exert no control over the setting in which the hypothetical process occurs.

In one class of observational research called retrospective research, this lack of control occurs because the exploration is being carried out sometime after the actual process being researched has actually occurred. For example, a researcher interested in family communication patterns might ask a group of adults to describe their recollections of communications with their parents during their childhood, and then relate the types of communication to the adults' current achievements, relationships with spouses and children, etc.[1] In this case, the independent variable (types of parent-child communications) cannot be manipulated, as the communication occurred many years in the past. And obviously the setting for this process will have been different for each subject, so no control over it can be exerted years later. But it is still quite possible to find covariance between the different types of family communication which took place in the past, and the current amount of achievement, satisfaction with current relationships, etc.

Observational research may also be required when it is impossible to manipulate the independent variable, or when it would be unethical to do so. A researcher studying the impact of newspaper editorial endorsements on voter behavior will not be able to systematically manipulate the endorsements given by newspapers, and even if she could, would probably have ethical qualms about interfering with the political process (even for such a noble purpose as communication research).

A third reason for conducting observational research involves the use of secondary data. This is data collected by some agency other than the researcher, possibly for some purpose other than communication research. For example, a researcher might use census data which includes information about the number of telephones and television sets and radios in different countries to study the effect of the availability of communication technology on national development. Obviously the

researcher can manipulate neither the amount of communication technology (unless he's fabulously wealthy) nor the Gross National Product of countries. He must use an observational design.

Other secondary data sources such as the public opinion polling archives maintained in the Institute for Social Research at the University of Michigan and the Roper Center for Public Opinion Research at the University of Connecticut can be very economical sources of data for observational research. These archives maintain research data from a number of studies done over many years. By selecting a set of poll questions, observational data about many different social phenomena can be obtained. In addition, media content summaries and programs are preserved by the Television Archives at Vanderbilt University, the Presidential Campaign Commercial Archives at Oklahoma University, the New York Times Index, and other sources. We'll cover the use of these sources in more detail in Chapter 18.

Natural Manipulations and Confounding Variables

Quite a few communication phenomena involve concepts and variables which do not lend themselves to being manipulated by the researcher. If this is the case, the researcher must rely on "natural manipulations". In both experimental and field research, variance in the independent variable is deliberately introduced by the researcher. This is the experimental manipulation. But in observational research, variance in the independent variable occurs as a consequence of the natural operation of the "real world".

It is important to recognize that it makes no difference whether the independent variable varies because of experimental manipulation or because of natural manipulations. In either case, statistical methods to detect covariance between the independent and dependent variables are used. But observational research does require that the researcher give up control over the temporal priority of the cause and effect variables. In both experimental and field research, the fact that the researcher manipulates the independent variable, and then observes the dependent variable means that the time ordering between the hypothesized cause and effect is known. This is not the case for observational research. Since both independent and dependent variables are measured, there is nothing to insure that the independent variable (the presumed cause) precedes the dependent variable (the presumed effect) in time. Without time ordering, the conditions of causality cannot be met.

Some people would argue that this means that causal relationships can only be investigated using experimental or field research designs. But this is not necessarily true. Within family communication research, for example, it does not require any great leap of faith to assert that the independent variable (different types of communication with parents when the respondent was a child) precedes in time the dependent variable (the state of current relationships). Of course, establishing covariance and temporal priority does not rule out the possibility that this time-ordered relationship between the independent and dependent variables might be the spurious result of common relationships with confounding variables, as we mentioned in Chapter 4.

The establishment of scientific relationships in observational research requires that the researcher do two things: first, determine the temporal priority of the independent and dependent variable; and second, account for the effect of all relevant confounding variables.

Establishing temporal priority often can be done by making reasonable assumptions about the time ordering of the variables. The emphasis is on reasonable. Arbitrary time ordering will produce incorrect scientific conclusions. A conservative rule of thumb is this: if you have any doubts about the correct time order of the independent and dependent variables, do not make any assumption. This will mean reducing the relationship from a causal one to the weaker covariance relationship (see Chapter 3), but without a strong temporal ordering of the variables, a covariance relationship may be all that is warranted.

An alternative way of establishing temporal order is to design a study which provides some evidence for the time order of the independent and dependent variables. Except in some special circumstances that we'll not address here, this will involve measurement at two or more points in time. Even then, the evidence for temporal ordering may not be completely unambiguous.

There are a number of ways to test temporal ordering in an observational research study. One typical way is through the use of cross-lagged correlations. In cross-lagged correlations, the independent variable and dependent variable are measured at two or more points in time. The test for

temporal order is made by examining the covariance of the presumed independent variable at Time 1 with the presumed dependent variable at Time 2, and contrasting this value with the covariance between the presumed dependent variable at Time 1 and the presumed independent variable at Time 2 (see Figure 13-3).

If the presumed time ordering is correct, we should observe that the independent variable at an earlier time (Time 1) covaries with the dependent variable at a later time (Time 2). But the temporal asymmetry principle which states that changes in the cause variable will produce later changes in the effect variable, and not vice versa, predicts that the covariance between the dependent variable at an earlier time and the independent variable later should be near zero.

A classic example of analysis by cross-lagged correlation is provided by Lefkowitz, et al. (1972)[2]. At issue was the relationship between children's viewing of violent television programs and their aggressiveness levels. The temporal ordering of these two variables is not clear. Viewing violent programs may be theoretically linked to higher levels of aggression in viewers by processes involving modeling of aggressive acts, by desensitization of the viewer to violence, by legitimization of violence as a solution to conflict, or by some other process in which television viewing precedes aggression. In this case, violent TV viewing is the cause variable and aggressiveness is the effect variable. On the other hand, one can reasonably link the two variables in the reverse time order by stating that naturally aggressive persons will seek to view programming which is consistent with their personal approach to conflict. In this formulation, levels of aggression precede in time the viewing patterns of individuals. Level of aggression is then the cause variable, and violent TV viewing is the effect variable.

A related extended example shown in the next chapter illustrates an experimental approach to the same problem. The Bandura study summarized there uses an experimental design which manipulates the subjects' exposure to communications, and thus controls the temporal order of the independent and dependent variables (children are always exposed to communications before measurement of their behaviors). But many media researchers reject the experimental approach on the basis of external validity. They feel that exposure to communications in an experimental setting is artificial, and so distant from the real way which people are exposed to media messages, that it is not valid to generalize from experimental studies to the real world. In particular, some believe that repeated exposure to messages over a period of years is necessary before meaningful change in
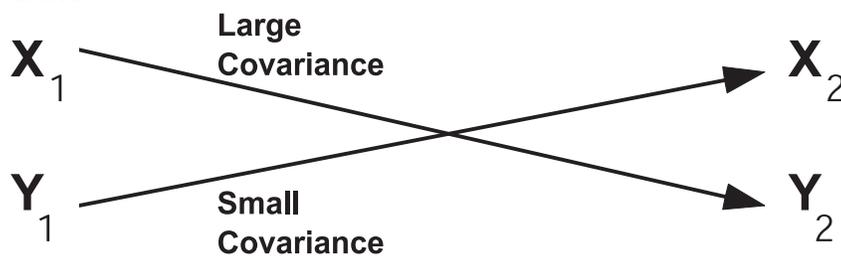
Presumed
Independent
Variable

FIGURE 13-3   Cross-Lagged Research Design

audience behavior can be observed. This means that a retrospective, observational research study is probably going to be required, since it is unreasonable to think that a researcher can control the communication exposure of a sample over a period of months or years. Rather, the researcher must rely on natural manipulations to produce variance in the independent variable, and must also rely on the ability of the research subjects to accurately report the level of the independent variable, after the fact.

However, the Lefkowitz study actually measured viewing habits and aggression levels of the same persons at a 10year time interval. It was not a retrospective study. The variables were first measured when the subjects were children in the third grade. The same variables were measured a second time when the subjects were recent high school graduates. If viewing violent television causes higher levels of aggression, Lefkowitz should have observed a significant correlation between viewing habits in the third grade and the aggressiveness of the same students after they graduated from high school. At the same time, the correlation between the aggressiveness levels of third graders and later television viewing should not have been significant. If the reverse time ordering of cause and effect is true, and aggressive predispositions predict television viewing, the data should show a reversed pattern of significant correlations.

The results of the Lefkowitz study are shown in Figure 13-4. As the diagram shows, this study found evidence for television viewing affecting later levels of aggressiveness, but none for levels of aggressiveness affecting later television viewing. This is very good evidence for the time ordering of these two variables, and helps to establish both the conditions of covariance (the significant correlation between television viewing in third grade and post-high school aggressiveness) and temporal priority (viewing predicts aggression, and not vice-versa).

The second requirement that observational research designs must meet is the control of all variables which may cause a spurious relationship between the independent and dependent variables. As we mentioned in Chapter 4, control of these variables may be achieved through manipulation, or through statistical control based on direct measurement of the confounding variables. Experimental designs control confounding variables through manipulation, but still require that they be identified and included in the design if the experiment is to achieve good external validity. Field
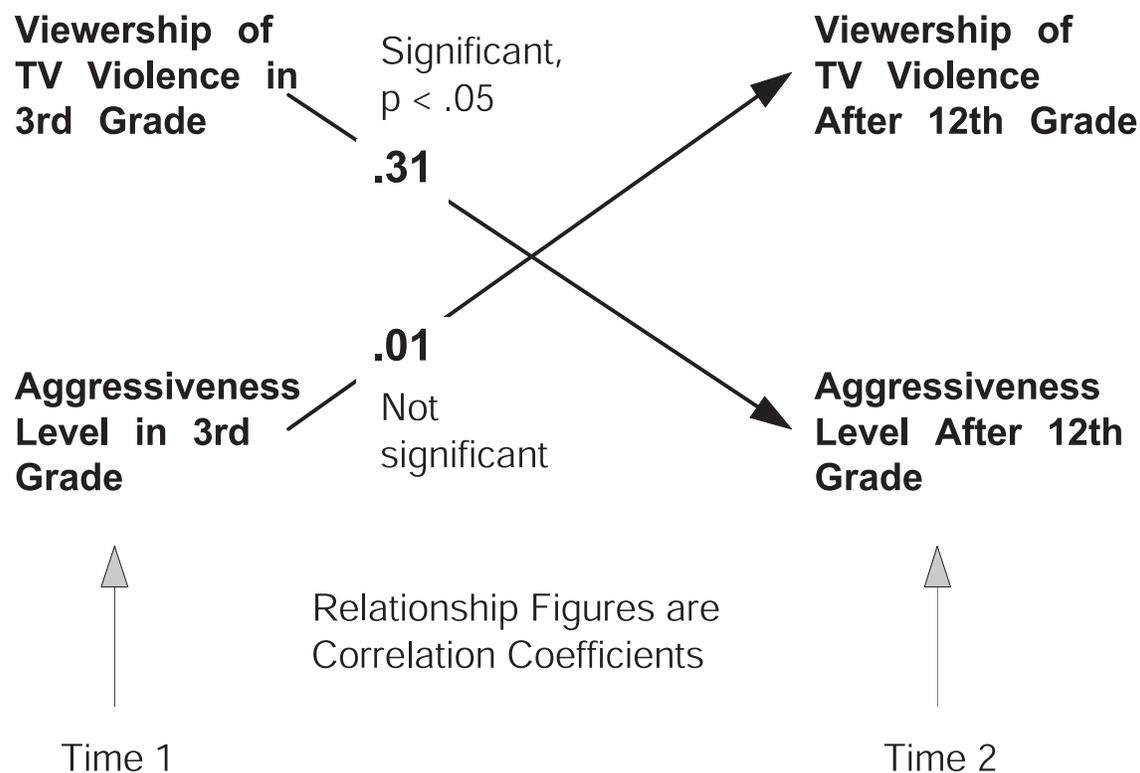

Viewership of TV Violence in 3rd Grade

Significant, p < .05

.31

Viewership of TV Violence After 12th Grade

.01

Not significant

Aggressiveness Level in 3rd Grade

Aggressiveness Level After 12th Grade

Relationship Figures are Correlation Coefficients

Time 1   Time 2

FIGURE 13-4   **Results of the Lefkowitz, et al. Study**

research designs do not require that they be included in the design, but the strength of the statistical tests is improved if they are. In observational designs, the researcher must identify and measure potentially confounding variables, or the internal validity of the study will decrease. And without internal validity, conclusions about relationships are incorrect, and any generalization, regardless of the level of external validity, is meaningless.

Viewed this way, the requirement of identification and measurement of all outside variables which might jointly affect the independent and dependent variables is absolute in the case of observational research, as both internal and external validity will be compromised by the failure to statistically control for these variables. It is almost as important in experimental designs, as failure to identify and include such variables in the research design will limit external validity, although internal validity will not be affected. Identification of outside variables is least important in field research, as both internal and external validity will be maintained. Not surprisingly, field research is usually the most difficult and expensive research setting.

Threats to Internal and External Validity

Although we've identified some general ways that research designs may fail to achieve internal and external validity, we need to talk in more detail about some of the specific problems in the validity of research design. For both internal and external validity, we'll discuss threats that occur when measurement takes place over a time span, and threats that occur at single time points.

This will not be a completely exhaustive list of the threats to validity. Specific research designs, subject populations, or research procedures may be vulnerable to other threats. What follows is a brief discussion of some of the most common threats. Any research design should be critically reviewed by the researcher, looking not only for the following threats to validity, but for any other way by which the action of the independent variable on the dependent variable might be confused with other factors, or by which the nature of the research may fail to generalize to the population being studied.

Single Time Point Issues in Internal Validity

Instrumentation Reliability and Validity

We've already discussed these problems in Chapter 7. Without reliable measurement, we may falsely conclude that the independent and dependent variable do not covary, when in reality our measurements just can't be trusted to be accurate. Likewise, if we are not measuring the theoretical concept that we think we are, the validity of our conclusions will be negligible. The solution to this problem is outlined in the early chapters of this book: pay significant attention to accurate conceptualization and operationalization, and check the reliability of measurement instruments.

Sampling

Again, we've discussed this threat to validity in an earlier chapter. In an experiment, field or observational study, if the subjects or respondents in differing research groups are not randomly chosen, we may confuse differences in the individuals who make up the groups with the effect of the different experimental treatments. The methods of random selection outlined in Chapter 6 provide a way to avoid this threat to internal validity.

Instrument Obtrusiveness

Good internal validity depends upon measurement which does not disrupt or direct the processes being investigated. To the extent that measurement intrudes on the communication process that is being studied, we can expect to be led to incorrect conclusions. A questionnaire which annoys respondents with insensitive or leading questions ("How many hours of mindless television do you watch each week?"), or which is so long that respondents can't fill it out without collapsing with fatigue is simply not going to give the accurate measurement that good internal validity requires. Likewise, an experimental measurement of the satisfaction with interpersonal conversation in which the experimenter interrupts the conversation every 15 seconds to ask the participants to fill out a

scale rating their satisfaction will so disrupt conversation and tip off participants to the nature of the experiment that valid conclusions will be impossible to make. Researchers can avoid this kind of threat to validity by pretesting their procedures. The obtrusiveness of the measurements are directly discussed with pretest subjects who have completed the research procedures, and changes to the procedure are made when it appears that this problem exists.

Manipulation Effectiveness

In experimental and field research, the researcher must assure herself that the intended manipulation of the independent variable actually did produce enough difference in the levels of that variable that good tests of covariance with the dependent variable are possible. Meeting the covariance test to establish a relationship is only possible if both the independent and dependent variable have some real variance. Generally, the greater the variance in the independent variable, the easier it is to observe a significant relationship.

There are three general ways to establish the effectiveness of an experimental manipulation. The first is by observation and assumption: the manipulation is so obvious that anyone can see that it was effective. If a researcher studying the effect of paper color on readership of brochures prints one brochure on blue paper and another on white, it is probably sufficient to say that she has manipulated color successfully.

The second way to establish effectiveness is by a manipulation check. This is a measurement made during or after the primary experimental procedure, to establish that the manipulation had its intended effect. Suppose a researcher was experimentally studying the effects of having negative information about a person prior to interacting with the person. To manipulate this independent variable, the researcher writes two paragraphs, one for each of two experimental groups. In one paragraph, the person's background is described positively and in the other paragraph mainly negative information is included. In this case, it is probably not sufficient to assume that the paragraphs will have the effect desired by the researcher. The researcher should include some measurement of the positive-negative evaluation of the person by the subject. For example, he might use a questionnaire at the end of the experimental procedure which has the question:

Before you began talking to your partner, what was you general feeling about his/her abilities?

      1   2   3   4   5   6   7   8   9
Very negative   *Very Positive*

By checking the means of the responses to this question in each experimental group, the researcher can present some evidence for the effectiveness of the manipulation of the independent variable. If the two groups are not statistically significantly different from one another in their responses to this question, there is no evidence that the manipulation actually worked, and thus the internal validity of the experiment is poor.

The third, and probably the best, way to establish manipulation is to measure the independent variable using some real metric. A researcher studying the effects of violence viewing on children can count the number of acts of violence in the videotapes shown to each experimental group, and possibly weight each act by some "severity" weight (aggressive yelling = 1; slapping = 2; shooting with assault rifle = 10, etc.). To make this measurement will require that the researcher provide an operational definition for the independent variable, something that is sometimes given short shrift in experimental research. But it should not be ignored. Operationally defining the independent variable, even in the simple case where only two experimental groups are involved, will usually improve the researcher's thinking about that concept. And it will surely improve the ability of the researcher to insure that an effective manipulation has been made.

Over-Time Issues in Internal Validity

When measurements are made at two or more points in time, such as in experiments which use before- and aftermanipulation designs, some serious threats to internal validity can appear. The basic presumption in multiple time point measurement is that the only thing that differs between the first and the second or subsequent time points is the level of the independent variable. But this

may not be true, and the researcher must take care not to confuse other factors which may affect later measurements with the effect of the independent variable.

History

Significant social or personal events may intrude between the first measurement and subsequent measurements. If the proper research design is not used, these events can produce changes in the dependent variable which will be confused with the effect of the independent variable. This problem increases in magnitude when there is a longer time span between measurements. A researcher who uses an observational design to study the reaction of the public's image of corporations to corporate advertising over a period of years will have to separate the effects of advertising from the effects produced by ups and downs in the economic climate, the appearance of banking scandals, the jailing of security traders, etc. A research design which uses a comparison group (such as a before manipulation-post manipulation design with control group, described in the next chapter) is often used to account for the effects of history.

Maturation

A related over-time problem is produced by growth and changes that occur within the research subjects. Children and adults change in many ways which are simply due to the passage of time. Children develop new abilities, adolescents expand their intellectual horizons, and the value systems of adults change over time. An internally valid research design must not confuse these changes with the changes produced by the independent variable. A researcher studying the effect of a classroom program to increase the time elementary school children spend reading out of school will have to use a design that accounts for the fact that children's reading ability improves dramatically in their early years. Such designs usually involve the use of a control group made up of equivalent research subjects. Since maturation effects should be identical in both the experimental and control groups, the comparison between them is insensitive to maturation effects.

Measurement Sensitization

There is a danger that the measurement instrument itself, when it is applied at the first time point, may affect the subject in ways that bias subsequent measurement. For example, a researcher who is interested in relating newspaper readership to political knowledge might use a questionnaire that poses a number of questions about the political process, as well as about newspaper readership. But by filling out this questionnaire, the respondent may become self-conscious about his newspaper reading, and particularly about reading political news. In the period of time between the first measurement and subsequent measurements, he may increase readership, pay more attention to political events, etc., in order to "perform better" on the next questionnaire. Just the fact of being involved in a research project may cause the subject to be much more interested in the topic of the research and to modify her behavior accordingly. This difference in performance between the first and the second or subsequent measurements can be confused with the effect of the independent variable.

Research designs which use control groups, can deal with this problem to some degree, as both the experimental group and the control group behavior will be modified to the same extent, and thus comparisons between them will reflect only the effect of the independent variable. But it is often more effective to disguise the intent of the measurement. In the above example, the researcher might "pad" the questionnaire with other questions which do not relate to the political process, and make sure the instructions do not directly mention this as the intent of the research. The respondent might be told only that the questionnaire involves questions about "lifestyle". This kind of disguise can pose some ethical problems. An alternative way to deal with extreme cases of measurement sensitization is to use research designs which do not employ multiple measurements. The post-test only design described below is an example. This decision to use such a design carries some penalty in the power of the statistical tests to detect relationships, as we'll see later.

Measurement Instrument Learning

If the same measurement instrument is used for multiple measurements of the same subject, there is a danger that subsequent performance on the instrument may be affected by simple learning of the experimental task or items on questionnaires. It is a well-known fact that students who repeatedly take general achievement tests like the Scholastic Aptitude Test tend to improve their performances, even though they have probably not learned a substantial body of new material in the intervening time. This improvement in performance can be confused with effects of the independent variable, if the proper design is not used.

Control group designs which use only a post-test may be used to account for this learning effect. Alternatively, posttest only designs may be used to eliminate the possibility of any learning threat to validity.

Another approach to controlling for learning at multiple measurement points is to use equivalent, rather than identical, measurement instruments. However, establishing that two different measurement instruments give reliably equivalent scores is often difficult. Establishing this equivalence usually requires a research study of its own.

Learning can occur within a single measurement procedure, too. For example, a measurement instrument may request a whole series of judgments about communications on a series of scales such as semantic differentials. Initially, these scales are unfamiliar. But as the subject gains more familiarity with them, he may begin to use them in a different fashion. This shift will give a systematic difference between ratings given at the beginning of the procedure and ratings given at the end. And this shift will be unrelated to the actual items being rated.

To guard against this kind of learning, items or experimental tasks must be randomized or arranged in a counterbalanced fashion (more about this below), so that items or tasks appear at the beginning of the procedure for some subjects, in the middle for others, and at the end for still others. Although this will not remove the learning effect, it will diminish its effect so that it is less likely to be confused with the effect of the independent variable.

Measurement Instrument Instability

This is an issue in measurement reliability. If the measurement instrument "drifts" over time, different results will be obtained at different time points. Such drift can be confused with the action of the independent variable over the same time period. Whether such drift in fact exists can be determined by establishing the level of test-retest reliability. Only measures with high test-retest reliability should be used to avoid this threat.

Subject Mortality

Although this phrase conjures up horror movie images of Transylvanian castles and research assistants named Igor, it actually refers to the loss of some subjects from a research study between the first measurement and later measurements. If random selection procedures are used to select subjects or construct the research groups, the resulting sample will initially be representative of the population from which it was drawn. But any loss of subjects from this sample between two measurement points may cause systematic differences in dependent observations that are not due to the independent variable. Subject mortality is rarely random, so this difference can be systematically confused with the effect of the independent variable.

As an example, consider an experiment in which a representative sample of city residents are chosen to study the effects of a health communication program aimed at sickness prevention. Booklets, videotapes, and in-home counseling sessions are provided for the experimental group, while the control group receives none of these communications. Periodically, the subjects are asked to report to a clinic for a check-up, and to report any health problems which have occurred since the last check-up. The results of a check-up are converted, via a complex formula for combining the various measures, into a single index of "healthiness" which is the dependent variable.

To assess the impact of a communication campaign such as this, observations of the dependent variable must be made over long time spans—probably years. During this time, some of the sample will move out of town, some will just stop coming for checkups, and some will really die. None of these events is random. Respondents in the lower economic classes may be more likely to move; those in the higher economic classes may be more likely to ignore the researcher's request to visit

the clinic regularly; older subjects are more likely to die than younger subjects; etc.

This non-random deletion of subjects will result in a loss of internal validity that will bias the results. If subject mortality occurs as speculated above, the researcher will end up comparing a representative sample which contains young and elderly, high and low income subjects at the first measurement point with a sample that is heavily skewed toward younger, middle-class subjects at the end of the experiment. Since both poorer and older subjects can be expected to have more health problems, the final measurement will probably have a higher mean "healthiness" than the first measurement, even if the communication campaign is completely ineffective.

The best way to deal with subject mortality is to take every possible step to insure that the minimum number of subjects is lost during the duration of the research project. Research procedures that provide some incentive to continue participation are particularly desirable. The researcher who has funds might pay the research participants, or appeal to their sense of responsibility in contributing to important research, or offer them the valuable results of the study, as incentives to help in completing the project.

Control group designs are useful in avoiding gross errors in inference, as subject mortality in both the experimental and control groups should be the same. While this will allow valid comparisons between groups, subject mortality will still result in inaccurate measurement of the absolute levels of the dependent variable. This improves internal validity, but still leaves problems with external validity, as we'll discuss below.

Subject Fatigue

Any research procedure which requires more than a tiny amount of time or thought may be subject to problems of subject fatigue or boredom. A very long questionnaire, a procedure which requires the subject to write long responses, or an experimental procedure that requires long stretches of focused attention are vulnerable to this threat to internal validity. The basic problem is that the subject's responses at the end of the procedure are not the same as they were at the beginning, and this shift in responses can be confused with the action of the independent variable.

There are two solutions to this problem. The first is self-evident: make the tasks or measurements as simple as possible. But meaningful measurement may require enough effort from the

Original Presentation Order: A   B   C   D

Counterbalanced Order: D   C   B   A
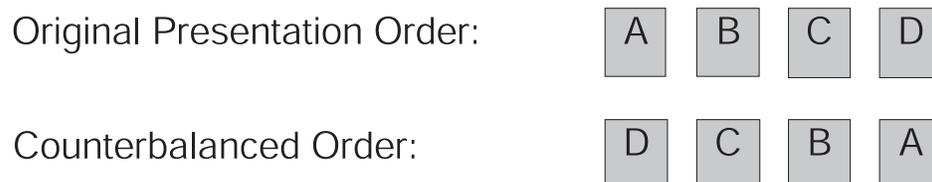
Fatigue/Learning Effect for Commercial

	A	B	C	D
Original Order	1	2	3	4
Counterbalanced Order	4	3	2	1
Average Effect	2.5	2.5	2.5	2.5

FIGURE 13-5   **Reversed Counterbalancing to Counter Fatigue and Learning**

subject that fatigue is inevitable. In that case, steps must be taken to assure that the effects of fatigue do not get confused with the effects of the independent variable.

The problem of fatigue is similar to the problem of measurement instrument learning discussed above. The solution to both these problems lies in the arrangement of measurement tasks or procedures. If measurements are placed at all time points during the research procedure, each measurement will be made under conditions of low, medium, and high fatigue. Since all measurements will occur under all fatigue conditions, fatigue will not be related systematically to measurement of the dependent variable, and so it will not introduce an error in inference.

As a simple example, suppose experimental subjects are rating the emotional content of four magazine advertisements, which we'll call ads A, B, C, and D, on a set of 50 Likert scales. This is a demanding task, and we can expect both fatigue and instrument learning effects. Both these effects are related to the position in the research procedure at which the ad is scored. Ads measured earlier will be rated by subjects who are less fatigued and who have less experience with the scales, while later ads will be rated by subjects who are tired, bored with the procedure, and experienced in using the scales.

The simplest control for learning and fatigue is reverse counterbalancing, in which the order of measurement is simply reversed. If we assume that fatigue and learning increase linearly (at each position the incremental increase in fatigue and learning is identical), we can assign sequential "fatigue/learning" scores to each position in the presentation. As Figure 13-5 shows, the average for each commercial in the reversed counterbalancing is identical, thus removing the effect of fatigue and learning from consideration. However, learning and fatigue are not necessarily linear, and more complex counterbalancing may be required, such as the scheme discussed in the next section.

Treatment or Measurement Order Effects

This threat to internal validity stems from the fact that earlier experimental treatments (manipulations of the independent variable) or earlier measurements of the dependent variable may affect later measurements.

The example described above provides a typical situation. Suppose advertisement B contains a blatant sexual appeal (a cologne ad with two intertwined nude bodies) and ads A and C contain mild sexual appeals (attractive men and women in swimsuits on a beach). The contrast between ad B and the one which follows it (ad C in the original order and ad A in the reversed order) may cause ads A and C to be rated lower on the emotional scales related to sexual appeals, since they are so much tamer than B. Exposure to ad B changes the way that subjects rate the following ads. Without being preceded by ad B, both A and C would score higher on these scales.

This effect may also be produced in experimental or field designs that present multiple manipulations of the independent variable to a single subject. For example, experimental subjects presented with a persuasive message justifying censorship in times of war, followed by one advocating First Amendment freedoms, can be expected to show different amounts of attitude change than subjects who are first presented with a message praising the founding fathers, then one concerning the First Amendment. Experimental treatment effects often persist indefinitely, and the effects of earlier treatments must not be confused with the effects of later treatments.

Counterbalancing of treatments or measurements is prescribed for this threat to internal validity. The simple reversal counterbalancing suggested for fatigue and learning is not sufficient in this situation, as there is still a systematic pattern to the influence of earlier treatments or measurements.

Using Figure 13-5, we can see that ad A will strongly affect the response only to ad B, since it is adjacent to no other ad. But ad B will affect ads C and A, but not D; C will affect only D and B, not A, etc. This unequal balance of effects means that we must use another type of counterbalancing to account for order effects. Specifically, we want one which does a better job of placing each ad next to all the other ads. Figure 13-6 shows such a counterbalancing, called a Latin Square design. Note that this counterbalancing also will control learning and fatigue effects, as each ad appears in each presentation slot once (Orders 1 and 4 are actually the same as the reversed counterbalancing described above). In fact, the requirement that learning and fatigue effects be linear is not present in this arrangement.

While the Latin Square counterbalancing will give complete control for all sequences of two, it will not completely counterbalance sequences of three or higher, as the right-hand columns of Fig-

Presentation Order 1:   A   B   C   D

Presentation Order 2:   B   D   A   C

Presentation Order 3:   C   A   D   B

Presentation Order 4:   D   C   B   A

Sequence	Frequency
A-B	1
A-C	1
A-D	1
B-A	1
B-C	1
B-D	1
C-A	1
C-B	1
C-D	1
D-A	1
D-B	1
D-C	1

This list contains all possible pairs of sequences

Sequence	Frequency
A-B-C	1
A-B-D	0
A-C-B	0
A-C-D	0
A-D-B	1
A-D-C	0
etc.	

This list contains a partial listing of 3-sequences

FIGURE 13-6   Latin Square Counterbalancing

ure 13-6 show. Some higher-order sequence effects are still possible. In fact, to control for all possible sequence effects of K treatments or measurements will require K! (K factorial) sequences. In the case of four measurements, this will require 4 x 3 x 2 x 1 = 24 different orders of presentation.

In many cases, a large number of different presentation orders is not reasonable. For example, to completely counterbalance 8 treatment groups would require 8 or 40,320 sequences of presentation!

The researcher must either choose a lower level of control (such as using a Latin Square design which controls only for the effects of adjacent treatments or measurements) or present the treatments or measurements in a random order to each respondent. This would randomize the error introduced by order effects, but not completely control for it.

Counterbalancing may also be necessary within measurement instruments. A very long questionnaire may introduce fatigue effects that affect items appearing nearer the end of the questionnaire, or may contain sensitive items which might affect subsequent responses. In these cases, counterbalancing of items within the questionnaire is a good practice.

The subject of counterbalancing is a complex one, and the interested communication researcher

should consult one of the many textbooks and handbooks on research design to find more details about the alternatives.

Single Time Point Issues in External Validity

Representativeness of the Sample

Sampling errors can cause problems in external validity as well as in internal validity. We covered many of the problems of a non-representative sample in Chapter 6. In particular, groups which are self-selected can cause problems. Persons who volunteer for research projects can be expected to be very different from the typical person (who does not usually volunteer). Convenience samples pose the same problems. For example, much communication research is done on college and university undergraduates. The results of this research is open to questions of external validity, unless the phenomenon being investigated is not related to the social background, age, intelligence, economic status, or race of the research subject. But these are variables on which undergraduates are substantially different from the general population, and unfortunately, most communication processes involve one or more of these variables.

To account for this threat, the researcher must either justify the generalizability of the sample (for example, physiological responses of students to communications are not likely to differ from those of the rest of the population, because of any differences in social variables), or limit generalization (the results apply only to white, upper income, educated young adults).

Reactive Effects of Setting

The research setting itself can produce responses in subjects that limit their generalizability. Participants in communication research are often exposed to communications in an artificial setting which enhances their attention to messages, their motivation to process and/or act on the contents of the messages, etc. The effects of these deviations from "real world" conditions limit the generalizability of the results.

To limit reactive setting effects, which affect both internal and external validity, the researcher must try to simulate the real environment to which the research is being generalized, and to be as unobtrusive as possible. For instance, laboratory television viewing should be done in as natural a setting as is possible. This may mean providing the laboratory with couches and chairs and subdued lighting, removing laboratory equipment from sight and introducing alternative targets of attention such as magazines. Or it might mean viewing with family members or friends, rather than alone. An interpersonal communication study of conversations should use a lounge-like setting, rather than a sterile classroom.

Observation and measurement should be hidden to the extent that is possible. Nonverbal measurement might be done with concealed video cameras, observations of group interaction could be made from behind one-way mirrors, etc. The researcher must critically examine the physical research setting, and use all creative means to make it as natural as possible.

Multiple Treatment Interference

Just as sequencing, fatigue and learning from multiple treatments or multiple measurements can affect the internal validity of a research study, they can also affect the external validity. Counterbalanced designs can improve the internal validity, but they do little to counter the multiple treatment effects on external validity. Counterbalancing controls for systematic effects by spreading them over all treatment conditions equally, but it does not remove the effects. As a result, the "unrealistic" treatment effects may produce findings in research settings that are not reproduced in "real world" settings.

If this threat to external validity appears to be substantial, the researcher must use a research design which does not involve multiple treatments or measurements taken from a single subject. This increases the number of subjects necessary, but will remove this threat to validity.

Over-Time Issues in External Validity

Reactive Sensitization (to externals)

Behavioral changes can be introduced by measurement and experimental manipulations, as discussed above in the measurement sensitization sections. Subjects who leave controlled settings between measurement sessions may react differently to communications and other environmental stimuli, as a result of their participation in the research.

A control group design can help with the internal validity problem. But the external validity problem remains, since both the experimental group and the control group will change in unpredictable ways. Looking again at the political communication example, if we find that a structured program of newspaper reading improves the political knowledge of the experimental group by 15%, compared to the control group, we must temper our conclusions with the knowledge that both the experimental and the control group's newspaper reading behavior have been modified. The experimental group's behavior was modified by the initial measurement, which many have increased their receptiveness to political news, and also by the manipulation of their reading habits; the control group's behavior was modified only by the initial measurement. The difference between the groups is due to the structured program. But we must be careful in concluding that we will see this difference if we introduce the program to the general public without the sensitizing effect of the initial measurement.

A research design which uses only after-the-fact (postmanipulation) measurement may be required to answer this threat to external validity.

Subject Mortality

The loss of subjects over time introduces a similar problem in external validity. Since the beginning and the finishing samples are different in makeup, due to mortality, it is difficult to determine exactly how much of the difference which occurs over time was due to mortality, and how much was due to the independent variable. It is thus difficult to generalize the effect observed in the research to the unmeasured population. This situation is much worse in observational designs which do not have a control or comparison group. In this case, subject mortality can be fatal to external validity.

The solutions to this problem are the same as those described in the section on subject mortality as a threat to internal validity. Keep as many subjects within the research program as is possible.

Summary

In this chapter we have distinguished among three major types of research: experimental research, in which the independent variable or variables are manipulated and the environmental conditions or the setting of the research is controlled; field research, in which the independent variable is manipulated, but the setting is uncontrolled; and observational research, in which the independent variable is measured under different levels that are the result of natural manipulations, and the setting is uncontrolled.

Each of these types of research have problems with both internal and external validity. Internal validity is the ability of the research design to provide support for claims about the presence of a relationship between the independent and dependent variable. External validity is the generalizability of the results to non-research or "real world" settings. In general, experimental research is high on internal validity and low on external validity, observational research is low on internal validity and high on external validity, and field research has medium levels of both types of validity.

Factors which pose threats to internal and external validity can occur at single time points, when measurements or experimental manipulations are made, and they can occur over time, when multiple measurements are part of the research design. Many of the threats can be answered with appropriate sampling, research design, treatment and measurement counterbalancing, and forethought in preparing the research setting, manipulations, and measurement instruments. But all research designs involve some compromise between validity and practicality, so no single design is free from all threats to validity. As a communication scientist, you must weigh the options, and design your research so that the fewest and least damaging threats are present.

Notes

(1) For an example of retrospective research in family communication patterns, see Chaffee, S.H., McLeod, J.M., & Wackman, D.B. (1973). Family communication patterns and adolescent political participation. In J. Dennis (Ed.), Socialization to politics. New York: Holt, Rinehart & Winston.

(2) Lefkowitz, M. M., Eron, L. D., Walder, L. O., & Huesmann, L. R. (1972). Television violence and child aggression: A follow up study. In G. Comstock & E. A. Rubenstein, (Eds.) Television and social behavior: Television and adolescent aggressiveness (Volume 3). Washington, D.C.: Government Printing Office, U.S. Dept. of Health, Education, and Welfare.

References and Additional Readings

Campbell, D.T. & Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: RandMcNally.

Drew, C.J. & Hardman, M.L. (1985). *Designing and conducting behavioral research*. New York: Pergamon Press. (Part II, "Basic Design Considerations").

Kerlinger, F.N. (1986). *Foundations of behavioral research* (3rd ed.) New York: Holt, Rinehart and Winston. (Chapter 17, "Research design: Purpose and principles"; Chapter 18, "Inadequate designs and design criteria"; and Chapter 19 "General designs of research")

Kidder, L. H. (1981) *Selltiz, Wrightsman and Cook's research methods in social relations*. New York: Holt, Rinehart and Winston. (Chapter 2: "Causal analysis and true experiments")