Chapter 14

Controlled Environments: Experimental Research

In the previous chapter we outlined the three basic types of research design (experimental, field, and observational), along with the relative strengths and weaknesses of each. In this and the following chapters, we will look at each of the three types more closely. In each chapter, we'll first present some typical research designs, and then we will look at and evaluate some actual research studies.

Typical Experimental Designs

We will first have to introduce some shorthand notation to clarify and speed up our presentation. This notation is outlined in Table 14-1. Familiarize yourself with this notation and proceed.

Case Studies and Single Group Designs

The first two designs we will discuss are not strong, and in fact are not classified as experimental designs by some writers (cf. Campbell and Stanley, 1963), as they have significant threats to internal and external validity. As Table 14-2 shows, there are very few threats to validity for which these designs actually provide benefits, other than those provided by random sampling and single-time measurement.

These designs are extremely weak and should be avoided, as using them is likely to lead to

Table 14-1 Notation for Research Designs

Y	Measurement of the dependent or effect variable
X	Measurement of the independent or cause variable
MX	Manipulation of independent variable
R:	Random assignment or selection of units of analysis
+	The design is more susceptible to this threat to validity
-	The design is less susceptible to this threat to validity
?	This threat to validity is not affected by the design under discussion. The threat may be more or less relevant, depending on other steps taken by the researcher.

Example: R: Y MX Y Single Group Design

(Random assignment or sampling, pre-manipulation measurement of dependent variable, manipulated independent variable, post-manipulation measurement of dependent variable)

Threats to Validity		*Single-Group Design*
History	+	(this design is susceptible to this threat)
Selection	-	(not susceptible)
Instrumentation Obtrusiveness	?	(susceptibility depends on factors other than design)

incorrect or insupportable results. Many factors other than the experimental manipulation may be responsible for change in the dependent variable. For example, results of case studies may be due to subjects' reactions to measurement, or to reactive effects of the experimental setting. Since the effects of the measurement and of the research setting cannot be isolated from the effects of the experimental manipulation of the independent variable, it is impossible to draw strong conclusions from a case study design.

Single group designs are particularly susceptible to invalidity created by over-time measurement. The effects of the manipulation cannot be separated from history and maturation effects, sensitization created by the measurement, etc. If these effects are strong, the researcher will likely completely incorrect conclusions from her research.

Both these designs may have some utility in the early stages of an investigation. Their results may suggest hypotheses which later can be tested under more controlled circumstances. But these designs by themselves are not very useful for establishing relationships which meet the requirements of causality.

Pre-Manipulation/Post-Manipulation Control Group and Post-Manipulation Only Control Group Designs

These designs are the most basic and widely used experimental designs. They provide control for many of the threats to internal validity by providing a comparison or control group. The control group's experiences are identical to the experimental groups, with one exception: they are exposed to a different level of the manipulated independent variable. Often this level is "absence", so that the control group receives no manipulation at all. But as we'll see below, this is not a necessary condition for an experiment. The control group can actually be any comparison group which receives a different level of the independent variable, be that level "absence" or just a different type of manipulation.

Rather than looking for absolute levels or changes in the dependent variable, these designs

Table 14-2 Threats to Validity: Case Study and Single Group Designs

Case Studies:	(a)		*MX*	*Y*	
	(b)	*R:*	*MX*		
Single Group:	(a)		*Y*	*MX*	*Y*
	(b)	*R:*	*Y*	*MX*	*Y*

	Case Study		**Single Group**	
	(a)	(b)	(a)	(b)
INTERNAL VALIDITY				
Instrumentation Reliability and Validity	?	?	?	?
Selection (Sampling)	+	-	+	-
Instrument Obtrusiveness	+	+	+	+
Manipulation Effectiveness	?	?	?	?
History	-	-	+	+
Maturation	-	-	+	+
Measurement Sensitization	-	-	+	+
Measurement Instrument Learning	-	-	+	+
Measurement Instrument Instability	-	-	+	+
Subject Mortality	-	-	?	?
Subject Fatigue	+	+	+	+
Treatment or Measurement Order Effects	?	?	?	?
EXTERNAL VALIDITY				
Representative Sample	+	-	+	-
Reactive Effects of Setting	+	+	+	+
Multiple Treatment Interference	?	?	?	?
Reactive Sensitization	+	+	+	+
Subject Mortality	-	-	+	+

look for comparative differences between the experimental and control groups. Since the control group will be subject to identical effects of all factors except that produced by the different level of the independent variable, any difference in the dependent variable detected when comparing the experimental and control groups after the experimental manipulation should be due to the action of the independent variable alone. Thus the covariance between the independent and dependent variable predicted in the research hypothesis can be separated from the effects of other confounding factors, and it can be tested for statistical significance. This design meets the conditions for a valid test of the existence of a relationship as outlined in Chapter 3. A checklist of these designs' strengths is provided in Table 14-3.

The only difference between these two designs is the presence of a pre-manipulation measurement in the Control Group Design. The benefit of this feature is that it allows "before-after" comparisons, both within and between the experimental and control groups, rather than the single comparison between the experimental and control group provided by the Post-Manipulation Only Con-

Table 14-3 Threats to Validity: Pre-Manipulation/Post-Manipulation Control Group and Post-Manipulation Only Control Group Designs

Pre-Post	*R:*	*Y*	*MX*	*Y*	(Experimental Group)
	R:	*Y*		*Y*	(Control Group)
Post Only	*R:*		*MX*	*Y*	(Experimental Group)
	R:			*Y*	(Control Group)

	Pre-Post (a)	**Post Only** (b)
INTERNAL VALIDITY		
Instrumentation Reliability and Validity	?	?
Selection (Sampling)	-	-
Instrument Obtrusiveness	-	-
Manipulation Effectiveness	?	?
History	-	-
Maturation	-	-
Measurement Sensitization	+	-
Measurement Instrument Learning	+	-
Measurement Instrument Instability	+	-
Subject Mortality	-	-
Subject Fatigue	-	-
Treatment or Measurement Order Effects	?	?
EXTERNAL VALIDITY		
Representative Sample	-	-
Reactive Effects of Setting	?	?
Multiple Treatment Interference	?	?
Reactive Sensitization	+	-
Subject Mortality	+	-

trol Group design. Each design has benefits and disadvantages which makes one or the other more appropriate, depending on the kind of research being conducted.

Suppose we are investigating the effects of the race of a speaker on his ability to change attitudes toward minority hiring practices. To do this, we will develop a measure for the dependent variable which includes a number of statements about minority hiring practices, to which the subject replies on a 9-point Likert scale with "strongly agree/strongly disagree" at the endpoints. These statements will be combined into an single index that indicates the degree of "Favorableness Toward Minority Hiring". This is the dependent variable. This measurement will be taken before and after the subjects receive the communication, in a Pre-Post design.

For the experimental group, we will use a speaker of Chinese ancestry, who delivers a prepared address verbatim. For the control group, we will use a speaker of Western European ancestry, who will deliver the same speech. The speakers will be similar in age, physical appearance and

Table 14-4 Example of Pre-Manipulation/Post-Manipulation Results

EXPERIMENTAL GROUP			
Subject	*Pre-Manipulation*	*Post-Manipulation*	*Change*
A	6	7	+1
B	3	4	+1
C	8	8	0
D	4	5	+1
Mean	5.25	6.00	+3
Standard Error	.96	.79	.22
CONTROL GROUP			
Subject	*Pre-Manipulation*	*Post-Manipulation*	*Change*
E	4	3	-1
F	6	7	+1
G	7	5	-2
H	4	4	0
Mean	5.25	4.75	-2
Standard Error	.58	.66	.50
Difference Between Experimental Mean and Control	0.00	1.25	1.25
Group Standard Error	.58	.58	.37

dress.

Note that our control group here receives the same message as the experimental group. This is necessary, as we are investigating the effects of the speaker's race, not of the message. It is not necessary that all exposure to communications be absent from the control group; indeed, to make sure that the control and experimental groups are equivalent in all respects, other than the levels of the independent variable, we will often have to use the same procedures in both groups. Our independent variable is a nominal one, with two levels: presence or absence of visual cues that the speaker is a member of an ethnic minority group. We've chosen, completely arbitrarily, to call the "presence" level of the independent variable the experimental group, and the "absence" level the control group. It really doesn't matter how we label them. All that is necessary is that we have two (or more) levels of the independent variable.

We expect that the initial attitudes of subjects on the topic of minority hiring practices will vary widely. Since we are randomly assigning subjects to experimental and control groups, however, we can assume that the same distribution of attitudes will be present in both groups (i.e., the groups will initially have the similar means and variance on our dependent measure).

Here we see one advantage of the pre manipulation-post manipulation design. It will allow us to test this presumption. By comparing the groups on the premanipulation measurement, we can see the extent of sampling error that was present in the assignment of subjects to the groups. If the groups score similarly on the premanipulation measurement, we can be confident that our random selection procedure has operated properly.

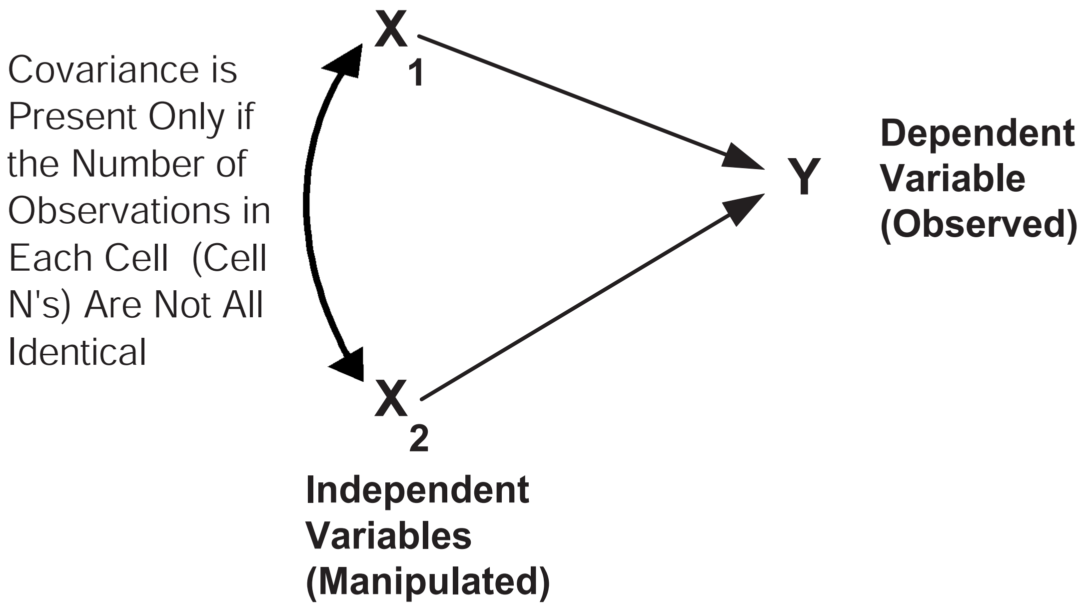


FIGURE 14-1 **Factorial Designs: Multiple Independent Variables**

Table 14-5 Factorial Research Design: Two Independent Variables

		X_2	
		No Product Sample	**Product Sample**
X_1	**Not Exposed to Ad**	*Group 1* (Control Group) [Measured Purchase Intention] *Y*	*Group 2* (Experimental Group) [Measured Purchase Intention] *Y*
	Exposed to Ad	*Group 3* (Experimental Group) [Measured Purchase Intention] *Y*	*Group 4* (Experimental Group) [Measured Purchase Intention] *Y*

We can also compare the difference between the pre- and post-manipulation measurements in the control group. This difference represents the combined effects of all the overtime threats to validity. This comparison is particularly useful when the control group is completely unexposed to any communications. For example, suppose two randomly selected groups of consumers participate in an experiment to test the effects of advertising on the consumer's intent to purchase the product being advertised. The experimental group receives the test advertisements, while the control group receives no advertising at all. The pre-manipulation to post-manipulation difference in the average intent to purchase that we see in the control group will indicate the effects of measurement sensitization, fatigue, etc. These effects are present in both the experimental and control groups.

The pre-manipulation/post-manipulation design can also permit us to control for sampling error and for the effects of the over-time threats. To do this, we may use the change or difference of the dependent variable scores between the pre-manipulation and post-manipulation measurements as a single dependent variable.[1]

Dealing only with change scores will allow us to ignore the initial differences between subjects. This will give a more powerful indication of the effect of the independent variable, as we can focus on relatively small changes that occur within subjects, and ignore relatively large initial differences between subjects. In the example in Table 144, the two groups have the same pre-manipulation mean, indicating a good random sampling procedure, with low sampling error. The post-manipulation means differ by only 1.25 units. But the corresponding change score means differ between the groups by four times as much: 5 units. This larger difference is more easily distinguished from sampling error, and so the statistical test is more powerful.

The Post-Manipulation Only design is less powerful, because it cannot distinguish between variance in the dependent variable that is caused by initial differences in the subjects, and the variance which is caused by the independent variable. But the design is less sensitive to internal validity threats like measurement sensitization and learning, and to external validity threats like reactive sensitization and subject mortality, since measurements are made at only one point in time.

In our example concerning the effect of race on the persuasiveness of a speaker, we would probably not use the Pre-Post design discussed above, but would choose the PostManipulation Only design. Any pre-manipulation measurement of attitudes toward minority hiring would probably sensitize the subjects toward the topic. That would mean that both the experimental and the control groups would listen to the message in ways that are not typical of the way that the population would listen. Also, if only a short time elapsed between the pre-manipulation measurement and the postmanipulation measurement, the subjects' memory of their answers to the pre-manipulation measurement might affect their responses to the same post-manipulation measurement items.

Multiple Independent Variable Designs

These designs are extensions of the above designs. But instead of a single independent variable (X), there are two or more independent variables (X1, X2, etc.) involved (see Figure 14-1). There are literally a limitless number of different research designs which involve multiple independent variables. We'll mention only one of the most common here: the factorial design.

In the factorial design, the different levels of all independent variables are presented to groups of subjects in all possible combinations. As an example, let's expand the advertising effectiveness study described above. In addition to presence or absence of advertising (X1), we'll add another independent variable (X2) which may cause variation in the subjects' intent to purchase the advertised product (which is the dependent variable, Y). The added independent variable is "provision of a sample of the product". For some experimental groups, we'll provide a sample for inspection; the other groups will not receive a sample.

The resulting design looks like Table 14-5. There are four groups, each of which receives a different combination of levels of the independent variables:

Group 1: No advertising, no sample

Group 2: Sample only, no advertising

Group 3: Advertising only, no sample

Group 4: Both sample and advertising

Group 1 serves as a control group, since no manipulation is done. It can provide a baseline of the purchase intention of the population in the absence of exposure to advertising and the absence of a product sample. Because all levels of the independent variable are present in all their combinations, we can separate the effects of advertising (which should be present in Groups 3 and 4, and not in Groups 1 and 2) from the effects of the product sample (present in Groups 2 and 4, and not in

Groups 1 and 3). We'll provide a guide to the statistical procedures needed to make the covariance tests later in this book.

Let's assume that we decide to use a pre-manipulation/postmanipulation design. Using our shorthand notation, we can diagram this experimental design like this:

Group 1	R: Y_{pre}			Y_{post}
Group 2	R: Y_{pre}	MX_2		Y_{post}
Group 3	R: Y_{pre}		MX_1	Y_{post}
Group 4	R: Y_{pre}	MX_1	MX_2	Y_{post}

This is just a somewhat more complex version of the single independent variable pre- post-manipulation design, with control group. All the threats to internal and external validity are the same.

Other Experimental Research Designs

The designs outlined so far are only the simplest and most common. A number of different designs have been developed to meet the specialized needs of communication and behavioral researchers. Generally they are variations on the basic theme of experimental research, which is manipulation of the independent variable and control of threats to internal validity. For further information on experimental research designs, you should consult a reference like Keppel (1982).

Units of Analysis

The basic unit that is observed and measured is called the unit of analysis. In communication research there are three different classes of units of analysis that are commonly used. The first is the individual. In studies using this unit, the variables are defined to measure characteristics or behaviors of a person: for example, we determine how different levels of the independent variable affect an individual's performance on a task, or the individual's degree of attitude change, aggressive behavior, production of messages, etc. The N's in studies which use this level of analysis are equal to the number of subjects in the experiment.

The second unit of analysis is that of the social system. In this level, variables are defined by the joint action of a number of individuals. These variables might be defined for dyads, small groups, organizations or, for that matter, for societies or countries. Measurement is at the social level: we might measure the duration of time that a dyadic relationship persists, the level of performance of a group, the communication patterns of a family, the adaptability of an organization, the ownership of VCRs in different countries, etc. The N's in this kind of study are determined by the number of dyads, groups, organizations, etc., and not by the number of individual subjects who participated in the experiment.

The third unit of analysis is the level of the message. While the previous units involve study of the senders or receivers of communications, this level examines the communication itself. This level is based in the study of language and symbolic information exchange. Research on media message content or frequency, nonverbal expressions or symbolic behavior, language intensity and choice of vocabulary, and similar studies will use this unit of analysis. The N's in studies using this unit of analysis are the number of stories, the number of utterances, the number of noun phrases, frequency of facial expressions, etc.

Although all three units of analysis are used in communication research, the individual unit is by far the most common. Experimental research also sometimes uses social system units. But it is very rare for experimental research to involve message units of analysis. This is in contrast to observational research, in which message units of analysis are frequently used in media content analyses or interpersonal language use studies.

Measurements that relate to differing units of analysis cannot be directly mixed within a single study. A research study which uses variables based on individual characteristics as the unit of analysis cannot incorporate social system or message variables directly. This point is sometimes confusing, as it seems to imply that we cannot relate messages to individuals, or single individuals to social groups. If it is true, how can we study the communication process, in which individuals, embedded in social organizations, transmit and receive messages?

EXHIBIT 14-1 Experimental Research: The Individual Unit of Analysis: The Effect of Modes of Exposure to Aggressive Behavior on Children's Imitative and Non-imitative Aggressive Play

General Topic

Bandura, Ross and Ross designed an experiment to test a number of hypotheses about the effects of exposure to real-life and mediated aggression on the later behavior of children. They were interested in the degree to which the children would imitate the aggressive media and real-life models, as well as the amount of general aggressive behavior that occurred after exposure.

Hypotheses

The first hypothesis concerned the relationship between the realism of the portrayal and the tendency of the child to imitate the behavior he or she has seen. The researchers argued that aggressive models can be arranged on a reality-fictional dimension, with real-life models at one extreme, cartoon characters on the other extreme and filmed human models in an intermediate position. The theoretical linkage involves the similarity and perceived closeness of the models to the children. The hypothesis predicted that the tendency to imitate the aggressive behaviors would decrease the further the model was from "real".

The second hypothesis related aggression anxiety and aggressive behaviors. They argued that those children who generally inhibit aggressive behaviors would perform fewer imitative and general aggressive acts. An associated hypothesis was generated concerning sex differences: as aggression is generally viewed as particularly inappropriate for females and is consequently likely to be negatively reinforced, males would be more imitative of aggression than females. The fourth hypothesis predicted that exposure to aggressive behavior removes inhibitions and would trigger more general aggressive behaviors. Specifically, frustrated observers of aggressive behaviors would be more aggressive than frustrated non-observers of aggressive behaviors, as their inhibitions against carrying out aggressive actions would be lowered by exposure.

Subjects

The subjects for this study were 48 boys and 48 girls enrolled in the Stanford University Nursery School. Their ages ranged from 39 to 69 months, with a mean of 52 months. An adult male and an adult female served as models of aggression in both the "live" and filmed conditions.

General Procedure

The subjects were divided into three experimental groups and one control group, with 24 subjects per group. One group observed "live" aggressive models and a second group observed the same models on film. The third group viewed a film depicting an aggressive cartoon character carrying out the same behaviors. All groups contained equal proportions of males and females, which meant that half of the subjects in the human model conditions (the "live" and film groups) were exposed to models of the same sex as the subject, and half were exposed to opposite sex models. Subjects in the experimental and control groups were matched individually on the basis of ratings of their aggressiveness in social interactions in the nursery school, made by both the experimenter and the nursery school teacher.

Following the exposure, subjects were tested for the amount of imitative and general aggressive behavior. This test took place in a different setting and in the absence of the models.

EXHIBIT 14-1, cont.

Manipulations

Experimental Conditions. In the "Real-Life" aggressive condition, each subject was brought to an experimental room. While the subject was in the room receiving some instructions about an unrelated activity, the real-life model carried out a number of highly distinctive aggressive acts, and made a number of specific verbal aggressions toward an large inflated toy Bobo doll. This doll was approximately the size of the children, and would bounce back upright after being pushed over or struck.

In the "Human Film" condition, a 10-minute motion picture was shown while the subjects were in the experimental room. The models and their behaviors were identical to those in the "Live" condition.

In the "Cartoon-Film" condition, a film of a model costumed as cartoon character, on a cartoon-like set, was used to present the aggressive behaviors. A "cartoon" black cat carried out the same behaviors on the Bobo doll as those that had been exhibited in the other two conditions.

Aggression Instigation

The children who were exposed to the various experimental stimuli as well as the children who were part of the Control group were next brought to a room containing attractive toys. Each child was initially allowed to play with these desirable toys, but once involvement with the toys was established, the child was no longer allowed to continue playing with the toys. Subsequently the child was led to a third room where he or she was allowed to play with a selection of toys which included the Bobo doll, some aggressive toys (a mallet, dart guns, etc.) as well as nonaggressive toys (crayons, a tea set, dolls, bears, etc.). The subject spent 20 minutes in this room during which his/her behavior was observed through a one-way mirror. The subject's behavior was rated on the amount of predetermined behavior in a number of categories every five seconds.

Dependent Variable Measures

The children's behavior was classified into the following six behavioral categories:

- Imitative Aggression: acts imitating the novel and specific physical and verbal aggression exhibited by the Model;
- Partially Imitative Responses: incomplete performance of the aggression or misdirected complete aggressions. Two types were recognized:

1. *Mallet aggression:* striking objects (other than the Bobo doll) with a mallet
2. *Sits on the Bobo doll:* Sitting on the

Table E14-1 Results of the Test for Bandura's Hypothesis 1.

	Comparison of Treatment Conditions					
Dep. Variable	*Live vs. Film*	*Live vs. Cartoon*	*Film vs. Cartoon*	*Live vs. Control*	*Film vs. Control*	*Cartoon vs. Control*
Overall Aggression	ns	ns	ns	[a]	[a]	[a]
Imitative Aggression	ns	[a]	ns	[a]	[a]	[a]
Partial Imitation						
Mallet Aggression						
Sits on Doll	ns	ns	ns	ns	[a]	[a]
Nonimitative Aggression	ns	ns	ns	ns	[a]	[a]
Aggressive Gunplay	[c]	ns	ns	ns	[a]	ns

[a]Indicates that the average value for the first named condition in the contrast was the greater one.
[b] For these dependent variables no significant overall difference was observed; further significance tests are thus inappropriate.
[c]Indicates that the average value for the first named condition in the contrast was the lesser one.
ns indicates "not statistically different"
From Bandura, Ross and Ross (1963). Copyright 1963 by the American Psychological Association. Reprinted (or adapted) by permission.

EXHIBIT 14-1, cont.

doll, but not aggressing toward it;

3. *Nonimitative aggression:* physical acts directed toward objects other than the doll, and hostile remarks other than those uttered by the Model; and
4. *Aggressive Gun Play:* shooting darts or aiming and firing imaginary shots.

Results

A partial listing of results (those relevant to the first hypothesis) is presented here. Based on the information contained in Table E14-1 and other analyses, Bandura et al. drew the following general conclusions.

In terms of overall aggression, exposing the children to aggressive models increases the probability that they will behave aggressively when instigated later. Although the three experimental conditions did not differ from one another, the children in all three experimental groups were significantly more aggressive than those in the Control group. In terms of specific categories of aggression, exposing the children to aggressive models led to increased imitative aggression relative to the Control Group. Live models were observed to elicit significantly more imitative behaviors than cartoon characters, thereby lending partial support to the first hypothesis, the one linking realism of portrayal to imitative aggression. The three experimental groups did not differ from one another in partial imitative behavior. An overall evaluation of the data suggests that humans on film portraying aggression are the most influential in eliciting and shaping aggressive behaviors.

Bandura, A., Ross, D., & Ross, S. (1963). Imitation of film-mediated aggressive models. *Journal of Abnormal and Social Psychology*. 66(1), 3-11.

The answer lies in defining "bridging" concepts. For example, a typical study might involve content analyzing soap opera themes and relating these to the personality characteristics of viewers. This would appear to mix the message unit of analysis (with content categories which quantify themes of sex, violence, amnesia, etc.) with the individual unit of analysis (personal characteristics of dependency, strength of emotional connection to others, etc.). However, on closer examination, we see that the critical variable in relating soap opera content to personality characteristics is really the amount of exposure to the particular themes, not just the amount of the theme present in the message. Exposure applies to individuals, not to the message (television programs), so the study is actually relating two variables at the individual level. Exposure is a bridging concept that links the characteristics of the message to the responses of the individual.

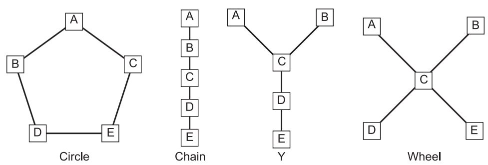
In a similar fashion, individual level concepts can sometimes be aggregated to produce social system units of analysis. For example, we might take the average number of external telephone calls made by individuals in a corporation, and use this single number to represent the amount of "outgoing corporate telecommunications". This would be a measurement appropriate to a social systems unit of analysis, as it would apply to the corporation, and not to any individual.

We mention these distinctions here, as the examples of communication research presented in this chapter and the following chapters were chosen partly to illustrate the use of differing units of analysis. As you read them, be particularly alert to the use, or potential use, of bridging conceptualizations.

Exhibit 14-2 Experimental Research: The Social System Unit of Analysis: How Group Structure affects Group Performance

General Topic

The ways in which members of a group may be linked together to exchange information are numerous. Which of all the different patterns of linkage are "good" patterns from the point of view of performance? Leavitt contrasts the levels of performance on a number of dependent variables of the following communication structures:



Hypotheses

Centrality reflects the extent to which one position is strategically located relative to other positions in a communication pattern such as those shown above. The position at the center of the Wheel has high centrality, while the position at the end of a chain has low centrality. It is a measure of closeness to all other group members and thus it is a measure of the availability of information. Such availability determines one's role in a problem-solving group, specifically one's independence and responsibility. Differences in independence and responsibility should affect group performance variables such as speed and accuracy. Leavitt did not explicitly define a set of hypotheses, so we may infer that he is testing non-directional hypotheses.

Subjects

One hundred male undergraduates from MIT were divided into 20 groups of five men each. These twenty groups were further subdivided so that five groups could be tested on each of the four communication patterns.

General Procedure

Each of the five subjects was assigned an identifying color and was given a card on which there appeared a set of five (out of a possible 6) symbols. Each subject's card was different from all the others: the symbol lacking on a card was different in each case. Thus, in any set of five cards there was only one symbol in common. The problem to be solved by the group was to find the common symbol which appeared on all cards.

Group members were allowed to communicate with one another by writing messages to those other members of the group with whom he had an open communication channel - that is, a link such as A-B, shown in one of the diagrams above. Each separate written communication was considered a message.

Manipulations

Group members were seated at a round table so that each person was separated from the next by a vertical wooden partition extending from the center to 6 inches beyond the edge of the table.

To allow communication with the other men, a box with color-coded slots was built into the center of the table and the partitions. Only the slots which connected the correct links in the communication network being tested were opened by the experimenter. Members communicated with those to whom they were linked on color-coded cards. In each subject's booth were six switches, on for each of the six different symbols. If a group member thought he had determined the identity of the missing symbol, he pressed the appropriate switch. When all five members had actuated a switch, the experimenter would halt that particular trial. Each group was given 15 consecutive trials on one single communication network. The manipulated independent variable is the centrality of the communication network. Note that this variable can only be defined with collections of two or more individuals. This makes the unit of analysis the social system, rather than the individual. All results are reported by referring to the performance of groups on this variable.

EXHIBIT 14-2, cont.

Dependent Variable Measures

Measurement on a number of dependent variables were obtained. We will restrict our discussion to those variables measured at the group level. These are also conceptualized at the social systems level.

The first dependent variable measured was elapsed time to obtain a solution. Two measures of time were obtained:

- The first time measure was the amount of time required for all members of the network to throw the switch.
- The second time measure was the fastest single trial for each group in a given network configuration.
- The second dependent variable was the number of messages that were transmitted within a group, in a given trial, before the group identified the missing symbol.
- The third dependent variable was the number of errors in identifying the symbol which were made by members of the group. This consisted of corrected errors (those which were corrected before the trial was over) as well as "final errors" at the end of the trial.

Results

Table E14-2 contains a partial listing of the results. From it, we can see the fastest trials occurred in the wheel and Y networks, which have high centrality, and the slowest were in the circle and chain, which have low centrality. More errors also tended to occur in the networks with low centrality.

Leavitt, H. J. (1951). Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology*. 46, 38-50.

Table E14-2 Means For Four Experimental Groups On The Immediacy And Formality Indices

	Communication Network			
Time	*Circle*	*Chain*	*Y*	*Wheel*
Range, 15th trial (seconds)	50-96	28-220	24-52	21-46
Fastest single trial (Mean, in seconds)	50.4*	53.2	35.4	32.0*
Errors				
Mean Total Errors (Last 8 trials)	7.6*	2.8	0*	0.6
Mean Final Errors	6.4*	6.2	1.6*	2.2

* Means marked with asterisk are different at $p < .05$

Examples of Experimental Research in Communication

We'll now move from considering experimental design in an abstract fashion to looking as some real examples of experimental communication science work. We've chosen these examples to illustrate several important similarities and differences among experimental studies, which we will discuss later.

The studies chosen for these examples do not necessarily represent the state-of-the-art in experimental research. We chose them because they are "classic", often cited studies, or because they represent an interesting or novel approach. In some cases, the conclusions have been challenged or elaborated on by later research.

Example of Experimental Research Using the Individual as the Unit of Analysis: The Bandura Study of Media and Child Aggression

Because of space limitations, we will synopsize the descriptions of the research projects in three Exhibits.

The first study, in Exhibit 14-1, involves the individual unit of analysis. In it, the response of individual children to exposure to mediated communications of violence and aggression is studied.

The Bandura experiment uses a post-manipulation control group design, with multiple levels of the nominal independent variable (the type of mediation of the message). Random assignment to the groups is used. This is a very strong design for internal validity. However, the external validity may be somewhat suspect, as the experimental setting, the nature of the communications, and the amount of exposure are not typical of "real life" conditions. We'll cover these in more detail below.

Example of Experimental Research Using the Social System as the Unit of Analysis: The Leavitt Study of Communication Networks and Group Efficiency

The individual unit of analysis that characterizes the Bandura, Ross and Ross study can be contrasted with the level of analysis in a study on group structure and group performance carried out by Harold Leavitt. In this study Leavitt explored the relationship that was thought to exist between the structure of a group (which determines access to the information needed in problem solving) and the speed with which a problem was solved and the number of errors that were committed in the process. In the experiment described in Exhibit 14-2, the structure of group communication patterns was manipulated by the researcher and the decision-making performance of the group as a consequence of these manipulations was measured.

This experiment also uses a post-manipulation only design. There is no control group identified as such, but all results are drawn from comparisons among experimental groups, so the control for over-time threats to internal validity are present. Again we see high internal validity coupled with suspect external validity, due to the artificial nature of the task and the actual communication channels.

The third example illustrates a bridging concept, which transforms message-level measurements into individual units of analysis.

Example of Experimental Research Using Message Measurement and Individual Units of Analysis: The Donohue Study of Conflict and Language Use

In this study the researchers focused on how the presence of a formal procedure for making a decision affects language use, when the decision is made under conditions of conflict. This is related to the use of language to mark relational distances in dyads.

Multiple independent variables are used in this study. One, gender, is measured rather than

Exhibit 14-3 Experimental Research: Message Measures and Individual Units of Analysis: How Problem Solving Structures Affect Utterances

General Topic

This study assessed the effects of providing a problem solving structure to conflicting dyads, on the way in which dyad members use immediate and formal language to mark their relational distance.

Hypotheses

Subjects in dyads who are initially in conflict on a decision making task will use less immediate and more formal relational code choices when they are instructed to follow a set of decision making rules, than will dyads who are free to reach a decision without such rules. Furthermore, females are predicted to use more immediate and informal code choices than men, regardless of the presence or absence of structure.

Subjects

Subjects were selected from 155 students enrolled in an introductory organizational communication course. Subjects were first asked to rate the importance of 10 characteristics of a hypothetical employee, to be employed as an orientation group leader. The subjects were then assigned to dyads so that individuals who maximally disagreed about the importance of the characteristics were paired together. This created 31 male dyads and 32 female dyads. Individuals were told that they were paired together because they had significant disagreement.

General Procedures

Dyads received written instructions explaining that they had 15 minutes to select one of four candidates for a job as an orientation group leader. Eighteen male and 17 female dyads served as a control group: they were instructed to deal in any way they thought appropriate with any disagreements that might arise during their interaction. The remaining 13 male and 14 female dyads were told to arrive at their choice for orientation leader following a four-step decision procedure. These experimental group members were asked to read a description of the four-step procedure. To insure that they understood the procedure, each dyad member was tested on knowledge of the procedure prior to the actual experiment.

All dyad interactions were audio- and videotaped. The tapes were used to produce transcripts of the interactions.

Manipulation

This experiment used the pairing of maximally disagreeing individuals to create dyads in which the members stand in conflict with one another. Conflict is not a variable here as it is assumed to characterize every dyad. The variable manipulated by the experimenters is Decision Making Structure, which has two levels: Presence and Absence. The second independent variable is Gender, which is a measured variable.

Dependent Variable Measures

Utterances produced by the dyads were evaluated on two variables. The first variable is "verbal formality/informality". Four measures were used to compute an overall index of this variable:

- Utterance length: the number of words, not counting interjections and hesitations;
- Structural complexity: this measure is based on the number of independent clauses;
- Syntactical complexity: the ratio of total number of verb forms to total words;
- Gunning's Fog Index: a readability index based on the number of independent clauses.

More formal language is more complex, and it indicates more social distance between the persons conversing.

The second dependent variable was "immediacy-nonimmediacy". This is a measure of the directness or intensity of the interaction. The transcripts of dyadic interaction were divided into units (simple sentences or phrases) and each of the units was rated on each of four language immediacy scale, using a 1 to 3 re sponse for each scale. The scales were combined to form a single immediacy variable.

Results

The availability of a decision making procedure had a significant effect on immediacy, but not on formality. There was no significant effect for gender. Both males and females were more immediate in the unstructured condition than in the structured condition. Table E14-3 shows the major results of the experiment.

Donohue, W. A., Weider-Hatfield, D., Hamilton, M. & Diez, M. A. (1985). Relational distance in managing conflict. *Human Communication Research.* 11(3).

Table E14-3 Communication Network Comparisons On Selected Dependent Variables.

	Immediacy		**Formality**	
Sex	*Structured*	*Unstructured*	*Structured*	*Unstructured*
Male	2.02	2.04	-.18	.58
(Std. Error)	(.01)	(.01)	(.29)	(.29)
Female	2.02	2.09	-.05	-.59
(Std. Error)	(.01)	(.01)	(.36)	(.27)

Donahue, W. A. Weider-Hartfeld, D., Hamilton, M., and Diez, M. A. Relational distance in managing conflict. *Human Communication Research*, 1983, Vol. 11, No. 3, pp. 387-405, copyright © by Sage Publications. Reprinted by permission of Sage Publications, Inc.

manipulated, while the other, decision making structure, is manipulated. The design is factorial, so there are four experimental groups: Male Structured, Male Unstructured, Female Structured, and Female Unstructured. Two dependent variables are used, so analyses are conducted twice: once for language immediacy and once for language formality.

Although the dependent variables are measured on transcripts, and thus appear to be appropriate as variables for research using the message unit of analysis, a bridging concept is actually used to define them as measures at the individual level of analysis. Rather than use utterances or syntactical units as the basic observational unit, the authors computed average scores for each subject on these language variables. This operational definition translates the measurements from being based on observations of the message alone (number of independent clauses, ratio of verbs to total words, etc.) to a basis of observations made of the individual (average number of independent clauses produced by the individual, etc.).

Only one of the hypotheses received support in this research, and this was found for only one of the two dependent variables. The predicted effect of decision making structure was found for the immediacy of the language used, but not for its formality. The factorial design allowed the researchers to separate the effects of gender and decision making structure, and talk about their effects independently.

Similarities and Differences among the Examples

Experimental research is characterized by two major factors, control over the research setting and manipulation of the independent variables, so we'll compare the examples on these characteristics. As these issues primarily affect the internal validity of the research, we'll also compare the studies on external validity concerns.

Control over the Research Setting

Researchers must take great pains to assure that the individuals or groups who participate in the research experience conditions that are as similar as possible, save for the different level of the independent variable that is applied to each experimental group. In the Bandura and the Leavitt examples we find descriptions of the efforts undertaken to exert control over the research setting.

Bandura et al. give a detailed description of the various experimental conditions which emphasizes the great pains which the researchers took to assure that the physical and verbal behaviors of the models would be identical in each of the three experimental conditions. Furthermore, when the subject entered the experimental room and was allowed to play with toys, "the play material was arranged in a fixed order for each of the sessions …..(I)n order to eliminate any variation in the behavior due to mere placement of the toys in the room". The arrangement of the toys left by the preceding child could have predispose a child to play with some toys rather than with others, and this would have affected the measurement of the dependent variable. Keeping the toys in a fixed arrangement for every session eliminated this potential source of variation in behavior.

The Leavitt study also contains references to control over the research setting. For example, the order in which the network configurations were tested was randomized. The experimenter could have chosen to run all the Y patterns first, and then to run all the Circle patterns, etc. But this would expose the research to the threat to validity due to treatment or measurement order effects, measurement instrument learning, and possibly measurement sensitization. Assigning the order of the patterns randomly eliminated the possibility of confusing the effects of circumstances that appeared only early or late in the experiment (such as unfamiliarity with the experimental procedures or fatigue) with the effects of the communication network configuration.

In addition, Leavitt controlled for possible effects of group members' geographical position and workstation color by shifting subjects' positions for each new run. As there were five runs in each pattern and five members in each group, each subject sat in each seat, and was assigned each color once, so any systematic effect was removed.

Control over the research setting can also be used to provide assurances that the various individuals or groups in the experiment are equivalent on all variables other than the independent variable. The requirement of equivalence on all other outside variables can be satisfied in one of two ways: either the researcher can show the groups to be equivalent on a set of relevant variables (matching); or he can select participants in such a way that he can assume that the groups will be equivalent (random selection).

Bandura and colleagues individually matched subjects in the experimental and control conditions on the basis of ratings of their aggressive behavior in social interaction in the nursery school. The experimenter and a nursery school teacher rated the children on four five-point rating scales which measured the extent to which the subjects exhibited physical aggression, verbal aggression toward inanimate objects and aggression inhibition. By matching the children in the four conditions, Bandura could show that for every child with a particular pattern of aggressiveness scores in, say, the "live model" group, there was a child with a similar pattern of scores in each of the other groups. The groups are therefore identical in the extent to which group members possess traits of general aggressiveness and predispositions toward aggression. In addition to matching subjects on the aggression variables, Bandura et al. also matched the groups on gender, so that the experimental groups were evenly split between males and females. If these variables exhaust the universe of relevant variables whose effects might be confused with the action of the independent variable, then, by controlling for their effect, Bandura and colleagues can confidently argue that any differences observed between the groups cannot be due to differences in general aggressive tendencies or gender, and must be due to the action of the independent variable.

In the Leavitt research, the references to control of individual subject variables are less clear. There are no mentions of matching subjects across groups or of random assignment of individuals to groups. However, since all subjects participated in solving a preliminary series of four problems, group differences in basic abilities should show up if the results of these preliminary tests are compared across groups. This kind of post hoc matching is sometimes used to assure that groups that were created with non-random procedures are actually equivalent on relevant outside variables.

In the Donohue, et al. article there is also very little reference to control. Although there are two main conditions (structured decision making versus non-structured decision making), there is no description of how dyads were assigned to these experimental conditions. Control was exerted on the creation of these dyads, by pairing subjects with maximally different scores on the employee characteristics. This is non-random pairing of the subjects, but if the resulting dyads were randomly assigned to experimental conditions (a procedure which was probably followed), then there is no systematic relationship between subjects and the experimental groups to which they were assigned, so this non-random pairing procedure will not confound the results.

Manipulation of the Independent Variables

The Bandura et al. study employed three independent variables: the degree of realism of the aggressive model, which consisted of three levels (Live Model, Filmed Model and Cartoon); Subject Gender; and Model Gender. Only the first of these independent variables was manipulated. The remaining two were measured independent variables which were used in matching subjects and experimental conditions.

The Leavitt study included only one independent variable and it was manipulated. The variable was represented by four different configurations of group structure which are contrasted to one another: the Circle, the Chain, the Y and the Wheel. Although these categories appear to be nominal in nature, they reflect the extent to which individuals in the group have "centrality" or access to information. This centrality is greatest in the Circle and decreases as we go toward the Wheel. Accordingly, the independent variable should be considered to be ordinal.

The study by Donohue and his associates reveals some interesting procedures for manipulation of the main independent variable of Structure and for the establishment of the conflict condition in all groups. Conflict is operationally defined by the pairing of two individuals who have maximally different rankings on employee characteristics, and it is presumed constant for all experimental and control groups. Once a dyad was constituted following this definition, conflict was said to exist. However, the researchers had to assume that because a priori disagreement existed, conflict had to exist as well. Note that the creation of conflict is not a manipulation of the independent variable, as all experimental conditions presumably had equal levels of conflict. But the conflict was necessary to test the hypotheses, so the researchers used their control of the experimental setting to create it.

The creation of conflict was a critical control, so the researchers carried out a manipulation check. Participants were asked to rate on a scale from 1 to 5 (with 5 representing "a great amount of disagreement") the amount of disagreement they perceived in the negotiation process. The unstructured group's mean was 2.18; the structured group's mean was 2.32. This can be interpreted as low to average levels of disagreement, and it brings one of the conditions of the hypothesis test into question. If the manipulation did not truly create a high state of conflict, how confidently can it be said that this experiment constituted a fair test of the hypothesis? The extent to which the manipulation did not really produce conflict constitutes a threat to internal validity.

In sharp contrast is the manipulation check for the structure manipulation. Not only were participants tested prior to the experiment on their knowledge of the four-step decision making procedure, but they were audio- and videotaped so that compliance to the structure could be assessed. The analysis of the tapes provides evidence that the experimental manipulation indeed was present — dyads that were supposed to use the rules indeed did use them. This knowledge of the manipulation effectiveness bolsters the internal validity.

Factors Affecting External Validity

To assess external validity, we must look at the extent to which we can generalize the results obtained in the experiment to the external world. We have encountered the topic of generalization before when we discussed sampling in Chapters 5 and 7. We stated then that the process of generalization only makes sense when we are reasonably sure that the sample that we observed is representative of the population to which we wish to generalize.

We can add to this statement another condition for experimental research: the process of generalization from the conditions of an experiment to the "real world" only makes sense when we are reasonably sure that the conditions under which we conducted the research are representative of the conditions which naturally occur in the "real world" to which we wish to generalize. The findings of an experiment will be only as externally valid as the degree to which the selections of subjects, experimental manipulations and experimental conditions can be considered to be representative samples of real conditions from the general population.

Sample Representativeness

The way in which we usually obtain a representative sample is by choosing randomly from the population. Robert Plutchik, in a book called Foundations of Experimental Research (1983) was only partly joking when he stated that "(T)he criticism is sometimes made that a good deal of what passes as psychological fact is based upon either the white rat or the college sophomore" (p. 47). The implication of this barb is that sometimes we substantially stretch the limits of generalization when we move from experimental subjects to wider populations. And, as Plutchik also states, we too frequently overgeneralize from other aspects of experiments.

Let's look at the subjects used in the three extended examples in light of their external validity. In the Bandura study, the children participating in the research were boys and girls enrolled in the Stanford University Nursery School. There is no reference to the method of selection of these participants from the enrollment of that school. Furthermore, the extent to which we can generalize from a nursery school such as this one would greatly depend on the criteria that the children met in order to be enrolled in this school.

If parental affiliation with Stanford University is required for attendance, then the students in this school are likely to be systematically different from the population-atlarge. They would probably score higher on variables such as verbal ability, income, intolerance of aggressive behaviors, etc., than would students at a preschool which used a policy of open enrollment. If such affiliation was required, then generalization of the results to nursery school students in Klamath Falls, Oregon, or Eagle, Wisconsin, or South Bronx, New York may be questionable.

The subjects in the other two extended examples were undergraduates in major research universities. An issue not discussed in either article was whether these students volunteered their participation, whether participation fulfilled a course research requirement; whether participation was monetarily rewarded, etc. If these conditions were present, the subjects may have been motivated to a far different degree or in a different way than persons in the general population might be. The subjects in both the Leavitt and Donohue studies were also all within the same general, but narrow, age bracket. Obvious hazards are involved in generalizing from students to middle-aged or older individuals.

Using samples of subjects who are selected for their convenience and accessibility may reduce to a very significant degree our ability to generalize. Of course, this situation is not unique to experimental research, as it can occur in field and observational research, too. But it is frequently a larger problem in experimental research, since the controlled settings of experimental research (the "laboratories") are usually not movable. Subjects must be induced to come to the research setting, and frequently this causes problems. In a sense, this problem in external validity is a consequence of the improvement in internal validity that controlled settings provides experimental research.

So when are we justified in generalizing from our samples which were drawn from restricted populations? Generally, whenever we can say that there is no reason to think that persons outside the restricted population will react differently than persons inside this population. To generalize from the Bandura study, we must feel that "children are children", at least with regard to their learning of behaviors from models. Likewise, the Leavitt and Donohue studies can be generalized to the extent that we believe that college students use the same strategies and social rules of communication as the older population.

Representativeness of Experimental Manipulations

Experimenters

In the Bandura et al. study a female experimenter was used in all experimental conditions. Would a male experimenter have created the same level of aggression when he told the children that they could not play with the attractive toys any more? We can expect that the sex, prestige, race, age and personality characteristics of the experimenter will influence the compliance of participants with experimental procedures.

In the Donohue et al. and Leavitt studies the identity of the experimenter(s) is not identified. It is likely, however, that faculty members or graduate students acted as experimenters. For an average undergraduate student, instructions given by such an individual will elicit different levels of compliance with experimental manipulations than instructions coming from a peer.

Communication Stimuli and Experimental Tasks

The representativeness of communication stimuli used in an experiment is a major factor in determining the level of external validity. The Bandura et al. study used the degree of realism of the model exhibiting aggressive behavior as the main independent variable. For this purpose three stimuli were defined: the actions of a live model; a filmed version of those same actions by the same live model, and a cartoon-like version of these actions. These reasonably represent real-world exposure conditions. Children will often be in situations where they can observe aggressive behaviors from adults and in the media.

However, all three versions of the stimuli shared one characteristic: the model exhibited novel and unusual aggressive behaviors. To the extent that these behaviors are novel and unusual, they are not representative of the population of aggressive acts. A consequence is that these stimuli will probably attract the attention of the research subjects. It is therefore possible that the imitative aggressive behaviors that result from stimuli of this type might not result from observing more mundane (and thus typical) acts.

The representativeness of the stimuli in the Leavitt study also affects the generalizability of findings. The independent variable, the degree of structuring of communication in groups, was manipulated by means of a rather complex arrangement of color-coded seating positions around a table divided by upright partitions which allowed passing only written messages through slots between dividers. The resulting arrangements were hardly representative of the kinds of conditions under which people generally work or solve problems. Additionally, the experimental task (determining a missing symbol from a set of cards) is not representative of activities normally carried out by people in problem-solving groups.

In addition to determining how representative the selected communication stimuli or experimental tasks are of the universe of such stimuli or tasks, the representativeness of the number of levels of stimulus and their intensity should also be considered. Put into statistical terms, the variance of the independent variable must be representative of the real range of levels and of variance that the independent variable takes on in the real world.

An independent variable can take on a range of values, of which at least two must be selected in order to test whether the variable is related to some other dependent variable. The variable "structure of decision making", as in the Donohue study, is a variable which can take on a fairly big range of values in reality (it has lots of variance). By using only two values out of this range, the Donohue experiment leaves open the question about the generalizability of its results to all structured decision making situations.

Contrast the selection of levels in that study with the selection in the Bandura et al. and the Leavitt studies. Bandura's "degree of realism" was represented by three levels; Leavitt's structuring or patterning variable was represented by five levels. It is apparent that increasing the number of levels of the independent variable is more likely to represent the full range of such a variable in the real world.

The researcher must consider not only the number of different levels, but also the intensity of these levels. To take an extreme example, suppose a researcher is interested in the effects of depictions of sexual activity on television. He sets up an experiment with three levels of the independent variable: an episode of "Little House on the Prairie"; an episode of "Days of Our Lives", featuring a steamy fade-out scene; and an episode from the EXXXSTACY adult cable network, featuring some fairly improbable adventures of coeds. This range of communications certainly has a wide variance, and it is likely to produce the best conditions for detecting significant relationships with dependent variables (the wider the variance of the independent variable, the easier it is to detect covariance). But it really does not accurately represent the range of sexual depictions normally used by real world viewers, most of whom do not view soaps, and virtually none of whom subscribe to a triple-X rated cable service. Thus the results have dubious external validity.

Dependent Variable Measures

This factor is closely related to problems associated with defining the dependent variable. In Chapter 2 we discussed the process of defining variables. From that chapter you may recall that simple concepts are easy to define conceptually and have straightforward operational definitions. Age, for instance, can be defined as "how long someone or something has been in existence" and a reasonable operational definition might be "number of years in existence". More complex concepts often have a range of indices that claim to measure the meaning associated with the concept. Complex concepts such as group performance (as in the Leavitt study) can be measured in a number of different, and sometimes inconsistent, ways.

Leavitt measured group performance using at least three indices: the number of errors in identifying the missing symbol, the amount of time needed to arrive at a decision, and member's satisfaction with group membership. Because these indices all tap into different aspects of performance, measurement of just one of these indices would give very little justification for generalizing to the concept of performance as applied in the general population. For example, the measure of "errors in making a decision" is only one of these aspects. We would not be justified in concluding that increases in centrality, which are found to be associated with decreased probability of errors, improves "performance" in general. A narrow definition will not generalize sufficiently.

The experiment by Donohue et al. uses a wide range of measurements to get at the concept of language use. Language use is defined in terms of its "formalitynonformality", for which four different responses are coded, and "immediacy-nonimmediacy", which uses another four different response categories. The range of these measures allows the authors to make wider claims about the generality of their results.

Summary

In this chapter we have outlined some basic designs for experimental research. True experimental designs include some comparison or control group which makes the design less susceptible to various threats to validity.

Two designs are particularly useful and are in widespread use: the Pre-Manipulation/Post-Manipulation Control Group design, and the Post-Manipulation Only Control Group design. The former design uses two measurement points, one before and one after the experimental manipulation. The latter uses only one measurement point, after the experimental manipulation. The Pre-Post design can give powerful results, as it can be used to make before-after comparisons of the values of the dependent variable within individuals. This allows the researcher to eliminate initial differences between subjects, and focus on the change that is produced by the independent variable. But the design is also susceptible to over-time threats to validity, like measurement sensitization and learning effects and subject mortality. The Post Only design is not as statistically powerful, but it is not susceptible to these threats.

Three units of analysis used in communication research are introduced: the individual unit of analysis, the social system unit, and the message unit. The individual unit of analysis predominates in experimental research, although some social system units can be found in interpersonal (dyads), group (collections of several members) and organizational (work groups or corporations) research. Message units of analysis are rarely found in experimental research in communication.

Variables appropriate to one level of analysis are often used at a different level by redefining them with "bridging" concepts like exposure (e.g., exposure to political news, which converts the message level of analysis to the individual level), or by aggregating them (averaging the responses of individuals at the individual level of analysis, to create a group score at the social system level of analysis). Variables defined at one level cannot be used at another without this redefinition.

We provided several in-depth examples of experiments on communication processes using different units of analysis. We used these experiments to illustrate the application of experimental design. Within the context of these experiments we have shown how the experimental control, manipulation and measurement decisions made by the researchers affected the experiment's internal validity and thus our confidence that the conclusions are correct. We also used these experiments to indicate how external validity can be affected by the research design and procedures.

While this chapter touches on some basic issues in experimental research, the serious communication researcher should consult one of the number of excellent books on experimental design we have listed below. The books will provide more detail about the issues presented here, and will also discuss other important issues which we are prevented from raising by space limitations.

Notes

(1) The use of change scores is recommended only for variables with high measurement reliability, as the reliability of change scores is the square of the measured variable's reliability. Thus a dependent variable with .80 reliability, which is usually considered very good, will produce a pre-post change score with a reliability of .802 or .64, which is poor. Other, more robust, ways to analyze change that occurs within a subject over time will be mentioned in Chapter 17.

References and Additional Readings

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

Huck, S.W., Cormier, W.H. & Bounds, W.G. (1974). *Reading statistics and research*. New York: Harper and Row. (Part 111: "Research design", containing Chapters 11 - 14)

Keppel, G. (1982). *Design and analysis: A researcher's handbook* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall. (Chapter 1, Design of experiments").

Kirk, R. E. (1982) *Experimental design: procedures for the behavioral sciences*. Monterey, CA: Brooks/Cole. (Chapter 1 "Introduction to Basic Concepts in Experimental Design").

Wimmer, R.D. & Dominick, J.R. (1987). *Mass media research: An introduction* (2nd Edition). Belmont, CA: Wadsworth. (Chapter 5, "Laboratory Research and Experimental Design")