Chapter 15

Semi-Controlled Environments: Field Research

In this chapter, we'll look at experimental designs which are applied outside the controlled environment of the laboratory. Field research uses the designs of experimental research to minimize the threats to internal validity, but applies them in realistic settings to also minimize some threats to external validity. However, as we'll see, there are some tradeoffs between these two types of validity.

Typical Field Research Designs

The basic research designs are identical to those used in experimental research, and you can refer to the previous chapter to review these designs. But the threats to validity differ because of the lack of control over the setting that the researcher must accept. These threats lead to some other strategies for strengthening control that field researchers often employ. Table 15-1 contains some threats to validity for the basic experimental designs. If you compare it to the similar table for experimental research (Table 14-3), you'll see that there are more potential problems with field research designs.

First, there are often sampling problems that may affect the representativeness of the sample. In field settings, the researcher frequently has to use subjects who are available, rather than those chosen by a purely random process. For example, if a researcher is conducting a study in a city street, as in the Ellsworth, Carlsmith and Henson study summarized below, she will have to use

Table 15-1 Threats to Validity: Pre-Manipulation/Post–Manipulation Control Group and Post-Manipulation Only Control Group Designs in Field Research

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	+	_

only the persons present in the street at the time of the study. These may be quite different in many ways from the general population, and thus it is possible that they may be a biased sample. The Ellsworth research shows the thought which a good researcher must use in developing procedures to insure that the sample is as representative as possible.

This problem is amplified if group units of analysis are chosen, like the "community" of the Douglas, Westley and Chaffee research summarized in another example covered below. Probability sampling from these larger units is often impossible.

A related problem in avoiding subject mortality, and thus maintaining the representativeness of the sample, occurs in pre-post designs. Subjects are not on the researcher's "home turf" (the laboratory), and it is easy for them to decline further participation in the project because of more immediate demands on their time. Or they may just disappear. The laboratory setting itself lends a certain amount of social pressure to encourage subjects to participate in research, and when this

setting is missing in field research, the communication scientist must compensate.

The effectiveness of the experimental manipulation is also frequently more questionable in field research than it is in experimental research. Because the environment is not completely under the researcher's control in field research, all conditions in the experimental groups other than the manipulation are not necessarily the same, as they are in experimental research. The field researcher must take extra pains to assure that the manipulation has been effective. Often this involves an explicit measurement of the manipulation (called the manipulation check).

With all these negative threats to validity, why would the researcher choose a field setting? The answer to this question can also be found in Table 15-1. One entry under the External Validity heading refers to "Reactive Effects of Setting". Field research designs are not susceptible to this threat, while experimental designs are. If you remember, this threat was discussed as a serious problem in the previous chapter. Reactive effects may cause the experimental subject to behave differently during the research than he or she would in "real life", and this will lead to researcher to erroneous generalizations of the research findings.

Field settings allow the phenomenon under investigation to operate under realistic conditions. Not only are the subjects reacting to the experimental manipulation in the same way that they would if the experimenter was not making systematic observations, but the operation of other variables in the environment is not impaired.

This point is very important. The virtue of an experimental research setting is in the control of outside variables, so that they do not confound the relationship between the cause and effect variables being observed. But this virtue is also a vice, when the relationship is generalized from the laboratory to the real world (we discussed this problem in detail in Chapter 4). If we use manipulative control of these outside variables, we may eliminate real and important processes in which the cause variable influences the effect variable under realistic conditions. While this will not affect the logic of our experiment (internal validity), it may lead us to the wrong predictions when we extend the results to the world outside the lab (external validity).

This is the problem that field research is designed to attack. For some communication processes, the laboratory setting produces too much reaction, or manipulatively controls too many outside variables, for the conclusions to have much utility. For example, public information or advertising campaigns are often difficult to study in the laboratory. The process of persuading or informing large and diverse audiences involves a huge number of different communication and psychological variables. Manipulative control of all of these variables except the few that are under study may give very different results from those that would be observed if these variables were allowed to operate naturally.

A good example of a type of field research which deals with the external validity problem occurs in advertising research. A common question concerns the comparative ability of two different message formats, or two different set of product appeals to persuade the audience to purchase a product. This question lends itself to a simple postmanipulation control group design, which could be carried out either in an experimental laboratory study or in a field setting.

In this design, two groups from the target audience are randomly selected. Each is exposed to a different version of the advertising message. After exposure to the message, the product purchase amount or probability of future purchase is measured. The message which produces the highest mean value of the purchase variable is then determined to be the most effective, if its difference from the other group's mean is significantly higher.

By contrasting an experimental and a field study which both use the same basic research design, we can see why field research is often chosen to investigate this type of question. Let's be more specific, and assume that two versions of a television commercial for a headache relief medicine called Molenol are being investigated. The first features an endorsement by a "typical" white-collar worker, who lauds the product's ability to relieve headaches which are produced by his working conditions: demanding bosses, complaining clients, and unresponsive secretaries. The second commercial also features an endorsement by a person who uses the product to relieve work-produced headaches, but this endorser is a professional football player who complains about demanding coaches, violent linebackers, and unprotective linemen. The fundamental research question concerns whether it is better to have an endorser who is closer in social position and has problems similar to the typical audience member, or to have a more glamorous person do the endorsing.

Let us look at various ways in which the research settings produce reactivity. The first major difference between an experimental and field investigation of this question lies in the conditions

under which the subjects are exposed to the message. In the experimental study, subjects would likely be presented with the commercial in a test room. While they might not be instructed to pay close attention to the commercial, the lack of realistic alternatives might mean that the commercials are viewed with very close attention. In contrast, a field study might use a splitcable system, in which half the homes on a cable system would receive one of the test commercials inserted into a normal programming mix, while the other half receives the other commercial in the same spot. (Such split-cable systems are available to commercial researchers in several cities in the U.S.). Subjects in the field test would thus be exposed to the commercials under very realistic conditions, so that their attention levels would likely be much lower than those in an experimental study.

A second critical difference is that the field study subjects might not even be aware that their response is being tested, so that they will not be sensitized to analyze the message to any degree. In contrast, the laboratory experimental subjects will know that they are being tested. This may cause them to analyze the message in much more detail than they would under normal viewing conditions.

A third difference is that the experimental study participants may feel a desire to respond positively to the nice person in the white lab coat who has convinced them to give up some of their time to participate in an important scientific study. They can do this by telling him what they think he wants to hear (this is called a "social demand" response). Clearly, the response of experimental subjects could be very different than the response of the field research subjects.

Finally, the criterion measure of the purchase behavior is likely to be very different in the two settings. It is very difficult to accurately measure purchase behavior in a laboratory setting. Purchasing a product (like many behaviors advocated by communications) is a complex behavior which depends on many other factors other than exposure to communications. One's buying habits and loyalty to competing brands, the cost and the availability of the product, and even the political position of the product's manufacturer can influence a purchase independently of communication effects. In a laboratory study, often the best one can do is ask a hypothetical purchase intention question: "how likely are you to purchase the product the next time you shop?"

This kind of question is very sensitive to "social demand" responses. Social demand refers to the tendency for research participants to respond in ways that they think the researcher wants them to respond, or ways that they believe to be socially appropriate, rather than in the way that they are most likely to respond when no one is looking. Since Molenol appears in the test commercial, subjects in the experiment may tell the researcher that they will purchase the product the next time they go to the drugstore. It's only good manners not to insult the product that the researcher is obviously interested in. But when these research subjects actually get to the drugstore, they buy the same generic aspirin that they've always bought, because it works fine, and it's cheaper (and no one is watching).

In the field study, the researcher may be able to get actual product sales information of Molenol in each of the geographic areas served by portion of the cable which carried each version of the commercial. The researcher can then find out the net impact of each commercial on the final criterion: sales.

Note that the field researcher will have deliberately sacrificed internal validity for external validity. If there are no significant differences in sales in the two regions, this does not necessarily mean that both commercials are equally good (or poor) at convincing consumers that Molenol is a good product. The effects of the uncontrolled outside variables may actually be overwhelming the differential effects of the commercials. For example, suppose that 35% of the subjects who were in the area served by the cable which showed the white-collar endorsement actually purchase Molenol, while 33% in the area in which the pro football endorsement was shown purchase Molenol. This difference is probably not statistically significant unless huge N's are used.

However, an experimental study with all the outside variables controlled for might have shown that white-collar endorsement is actually 200% better at convincing subjects to purchase Molenol, when all other factors are held constant. The experiment will give better information about the communication process being studied, but it will do so under conditions far different from those which actually occur; it will have good internal validity and poor external validity. The field experiment will give less definitive answers about the theoretical process, but will give much better prediction about the performance of the communication under real-world conditions; it has poorer internal validity but better external validity.

It is clear that the choice of research setting must be made with some eye to the use to which

the information from the study is to be put. If the researcher is interested in the more theoretical issue of the effects of endorsements in persuasive communications, a laboratory study would be preferred, because it would give a clearer answer. But if the researcher wants to select the communication which gives the largest increase in sales, she probably will choose a field study.

The internal validity of field studies can be strengthened by using statistical control, rather than manipulative control. With statistical control, outside variables like price and purchase habit in the above example can be measured, and their effects subtracted from the effects of the experimental variable. This allows the outside variables to operate normally while the research is in progress. But the researcher can then separate their effects from that of the manipulated variable, and so strengthen the internal validity of the research. This is the control strategy employed in purely observational research, and we'll discuss it further in the next chapter. We'll also mention some of the statistical procedures to implement this kind of control (like partial correlation and analysis of covariance) in Chapter 19 of this book.

The external validity strengths and the care required to address internal validity weaknesses of field research designs can be seen in the examples which follow. These examples, once again, were chosen for their interesting features or classic status. Examples of individual units of analysis and group units of analysis are provided, but no example of a message unit of analysis appears in this section. The use of the message as the basic unit is even rarer in field research than it is in experimental research, so we will restrict our examples to the other two units of analysis.

Examples of Experimental Research in Communication

Example of Experimental Field Using the Individual as the Unit of Analysis: The Douglas, Westley and Chaffee Study of the Effects of a Public Health Information Campaign

The first example shown in Exhibit 15-1 involves the effects of a public health information campaign. Measurements were made on individuals in two communities, one of which was exposed to the information campaign and one of which was not. Although this research was conducted with individuals as the units, it could also be done with a group unit of analysis (the community) if it was replicated in a sufficient number of communities.

The Douglas, Westley and Chaffee study points out some of the problems with representativeness of samples that can occur in field research. The communities to be studied were not chosen randomly, nor were they similar in many respects. But Douglas, et al. tried to match the two communities on the characteristics whose effects they thought might be confused with the effects of the independent variable, like the educational levels, occupations, incomes and the media use of the research subjects.

But even with matching, the two communities differed strikingly on the percentage of the original sample which consented to participate in the research. In the control community, twice as many persons refused to participate in the research. This opens the possibility that some of the effects seen might be due to systematic differences between the samples in each community, introduced because the more "uncooperative" persons were eliminated from the sample in the control community, while similar persons were not eliminated in the experimental community. If these uncooperative persons differed in their knowledge of mental retardation, or their media use, their exclusion will cause systematic differences between the experimental and control community results. These differences will be confused with the effect of the experimental manipulation.

Since the communities initially differed on the level of the dependent variable, Douglas et al. chose a research design which would allow them to look at change in the dependent variables, rather than at absolute levels of the dependent variables. This is a form of statistical control. Unfortunately, the premanipulation-postmanipulation design which was chosen introduced the problem of sample mortality, and worse yet, different levels of sample mortality in each community. This problem reduces the internal validity of the study, since it is possible that some of the effects seen are due to systematic attrition of certain types of research subjects (like those who learned little from

Exhibit 15-1 Experimental Field Research: Individual Units of Analysis: An Information Campaign That Changed Community Attitudes

General Topic

A field study was carried out to determine whether a public information campaign was effective in increasing community information levels regarding mental retardation and in creating more positive community attitudes toward mental retardation. The effectiveness of the campaign was measured by contrasting the knowledge and attitude levels of individuals in an "experimental" community with those in a matched "control" community.

Hypotheses

The authors present a number of assumptions about the characteristics of the topic of mental retardation which form the basis for four hypotheses directly addressing knowledge and attitude levels.

The first characteristic to be assumed is a generally low level of knowledge in the population regarding mental retardation as a community problem. The second assumption is that there is a low probability that hardened attitudes toward mental retardation already exist in the population. The third assumption is that the likelihood of external events reaching the experimental and control communities during the conduct of the research (and thus confounding the results) is low. These assumed characteristics, the authors then reason, constitute optimal conditions for an information campaign to produce both knowledge and attitude change, as predicted by various media effects theories.

The first hypothesis predicted that post-campaign knowledge of information included in the campaign would be greater in the experimental community than it would be in the control community.

The second hypothesis predicted that the pre-to-post campaign increase in knowledge would be greater in the experimental community than in the control community.

The third hypothesis proposed a parallel effect on attitudes toward mental retardation: a greater positive shift from pre-to - post campaign would be observed in the experimental group than in the control group.

The fourth hypothesis proposed a positive relationship between attitude change and information gain in the experimental community.

The final two hypotheses focused on information sources. Hypothesis five predicted that, in the experimental condition, local media would be more often identified as sources of information than non-local media. Furthermore, in the experimental community, local media would be cited more often as sources of information than they would in the control community.

Hypothesis six proposed that "friends" would be more frequently cited as information sources in the experimental community than they would be in the control community.

Subjects

To determine the knowledge level and attitude consequences of the public information campaign, two matching towns in Wisconsin were selected. The two communities were highly comparable in population characteristics, income, occupational distribution and median education. The only differences were that the community to be used as the "control" was higher in median education and had a higher proportion of professionals and managers. The "experimental" community was to be the setting for the information campaign; the control community was not to receive such a manipulation.

Both communities were matched on another set of attributes extremely relevant to the hypotheses to be tested. Both had locally published weekly newspapers, a weekly shopper's guide and had local radio stations. Additionally, both communities were equidistant to the nearest urban center and both depended on that city's two newspapers and four television stations for external news.

Within each community a systematic random sample of households was drawn by choosing, after a random start, every tenth household on these communities' electric utility lists. This yielded 134 homes in the experimental community and 169 in the control. A single adult respondent was chosen randomly in each household.

Exhibit 15-1 cont.

General Procedure

Because the two communities were not equivalent in their initial involvement in mental retardation work, this field experiment was designed as a "before-after" or "premanipulation-postmanipulation" study.

The before and after measures were separated by a period of eight months. During this period the information campaign was carried out in the "experimental" community; the "control" community was left alone. The pre-manipulation observation consisted of two sets of measures: one set contained items designed to measure individuals' knowledge of mental retardation, while the other measured attitudes toward mental retardation. Cooperation rates were 81% in the experimental community and 44% in the control community. The post-manipulation test contained items to measure general as well as specific knowledge levels and attitudes after the campaign had been concluded. The post-test cooperation rate was 78% in the control community (22% subject mortality) versus 85% (15% subject mortality) in the experimental community, resulting in a final N of 85 in the experimental community and 60 in the control community.

Manipulation

A public information campaign of six months duration was presented to the experimental community between the pre-manipulation and post-manipulation measurement dates. During this period, local channels of communication were utilized to present information about mental retardation in the experimental community. Twenty news stories, five feature stories and a Mental Retardation Week advertisement were published in the local paper. News items were broadcast over local radio "an uncounted number of times" and were also inserted in church bulletins. Further information efforts consisted of presentations to church, 4H and service clubs' meetings. Finally, the local Junior Chamber of Commerce conducted a year-long project focusing on the subject.

Dependent Variable Measures

As was mentioned above, the dependent variables were measured at two points in time: pre-campaign and postcampaign. The researchers identified two dependent variables: knowledge about mental retardation (specific as well as general knowledge) and attitudes towards mental retardation.

Pretest Measurement

Knowledge Measurement. In the pretest, the knowledge variable was measured using a set of six items designed to measure general levels of information about retardation. The scores on the individual items were used to compute mean correct scores for each respondent and each community.

Attitude Measurement. The pretest also contained a set of 21 Likert-type items designed to measure attitudes toward retardation. Some examples:

Mentally retarded people should never get married

1 2 3 4 5 6 7
Disagree Agree

Our community should establish special community activities for the retarded.

1 2 3 4 5 6 7
Disagree Agree

Here, too, an individual respondent's attitude was determined by computing his or her mean across all 21 attitude items, and means for each community were computed.

Post-Manipulation Measurement

Knowledge Measurement. In the posttest, the set of six general knowledge questions used in the pretest was readministered. Again, individual respondent means and community means were computed.

In addition, the posttest also contained an additional set of ten knowledge items covering specific information which had been presented in the information campaign. Some examples:

- How many retarded people would ordinarily be in a city of about 5000 people?

Exhibit 15-1 cont.

- What percent of the mentally retarded can be trained to go out and make a living?

The answers to these questions were scored as "correct-incorrect" and for each community the percentage of respondents giving correct answers to these questions was determined.

Information Sources. For each of these ten specific information questions, respondents were asked to identify the source from which that specific information was obtained. They were asked to check as many responses as were appropriate from this list:

local newspaper
local radio
state newspaper
club meeting
friends
other _____ (what was it?).

Attitude Measurement. The posttest contained the same set of 21 Likert-type items designed to measure attitudes toward retardation that were on the pretest. Once again an individual respondent's attitude was determined by computing the mean across all 21 items, and an overall mean for each community was also computed.

Results

Table E15-1 contains a partial listing of the results of this field experiment. This table contains results gleaned from various tables in the original article and does not report results relevant to all the hypotheses which were tested in this field research.

From these results the authors draw the following conclusions: The information campaign had a significant effect on specific information levels. On the average, almost one half of the respondents in the experimental community correctly answered questions about information provided in the campaign. This proportion is almost twice as large as the proportion in the control community, where only slightly more than one-fourth of the respondents were able to answer these questions correctly.

A different conclusion is drawn about general information levels. The means for the general information items in the two communities after the campaign are not significantly different from one another, nor are they different from the pretest means. It is concluded that the information campaign did not affect knowledge levels beyond those items specifically covered in the campaign.

A final conclusion that can be drawn is that the information campaign positively affected attitudes toward mental retardation. In the experimental community the difference between the pretest and posttest was significant; in the control condition no significant shift was detected.

Douglas, D. F., Westley, B. H., and Chaffee, S. H. (1970). Response to an information campaign that changed community attitudes. *Journalism Quarterly.* 47(3), 479-487.

Table E15-1 Knowledge and Attitude Differences between an Experimental and a Control Community Following a Public Information Campaign

			Experimental Community	*Control Community*
A	*General information levels*	*Before*	1.89	2.10
		After	1.66	2.33[a]
B	Specific information levels (% correct)		48.4%[b]	28.7%
C	*Attitude levels*	*Before*	2.01	2.00
		After	2.16[c]	2.04

a: Differences between Before and After means are not significant.
b: 7 of 10 items showed significant differences between Experimental and Control communities
c: Significant increase in positive attitudes from Before to After

information campaigns). This reduction in internal validity is the price the researchers paid for conducting their research under conditions that insured excellent external validity.

A Second Example of Field Research Using the Individual as the Unit of Analysis: The Ellsworth, Carlsmith and Henson study of the Effect of Staring on Flight

In the next example, we see how researchers go to great lengths to deal with the problems of reduced internal validity by conducting a series of studies to eliminate possible alternative explanations for their results. These alternative explanations, which are symptomatic of poor internal validity, arise because of the inherent lack of control over the setting that field researchers must endure.

Table E15-2 shows typical results. This table gives the essential results to Experiment 4. The results of the other experiments were reported in the original article in a similar fashion.

Compared to a laboratory experiment, the Ellsworth, et al. field experiment must deal with many threats to internal validity. First, there are problems with the sample and the assignment of subjects to experimental conditions.

Although painstaking randomizing procedures were used to select the subjects and to assign them to experimental or control conditions, the basic sample is still not representative of the general population. Only motorists or pedestrians who were on the street at non-rush hours could be used in the field research.

This eliminates most of the working population, who are likely to be in offices or factories at the times when the research was conducted. So while the researchers can state that there was no bias in assignment of subjects to experimental conditions, they cannot state that the subjects are representative of the general population, as there were probably many more students, housewives, retired persons, or unemployed workers used in the research than would appear in a probability sample of the general population.

But the researchers are probably not too worried about this, because it is difficult to see how overrepresentation of these kinds of subjects would change the results. Why would one expect that students or unemployed workers would react differently to a stare than other members of the general population? Although the unrepresentativeness of the sample could be a threat to validity, its hard to see how it does obscure the results in this case. If we have no reason to suspect the sample, even though we know it's biased in some way, we can feel a little more comfortable about trusting the results. What's critical is that the bias does not appear to affect the operation of the theoretical process, or the measurement of the dependent variable.

A second noteworthy aspect of this research is the effort taken by the researchers to eliminate alternative explanations for the observed differences between the experimental and control groups. The researchers examined the possibilities that misinterpretation of the communication might be taking place (with the stare interpreted as a racing challenge or as having some sexual implication), that subjects might simply be reacting to an unusual situation, or that the reaction might occur only when automobiles were involved. By being able to reject each of these alternative explanations, the researchers strengthened the internal validity of the research.

By explicitly including variables like sex of research assistant and subject and incongruity of the situation, the researchers exerted some manipulative control over possibly confounding variables. Although they were not able to do so, if they could have observed other variables, they could have used observational or statistical control procedures to remove their effects, further strengthening the internal validity of the research. In fact, in Experiments 1 and 2, the age of the subject was estimated by the research assistant, and the effect of age on flight time was tested. As the researchers found no consistent relationship of age with the dependent variable, they did not consider it in later experiments. But by measuring the variable of age and including it in the analysis, the strengthened their conclusions by eliminating another possible confounding variable.

Exhibit 15-2 The Stare as a Stimulus to Flight in Human Subjects: A Series of Field Experiments.

General Topic

A series of field experiments were conducted to test the idea that avoidance behavior can be created by the nonverbal communicative act of staring. It was hypothesized that a stare has negative or threatening communicative properties, and that persons who are the object of the stare will try to escape from it. If escape is temporarily impossible, tension may build up and the escape will be more dramatic.

Hypotheses

The general hypothesis tested was that the speed of withdrawal from a situation would be affected by the presence or absence of a stare by a stranger. Specifically, the speed with which a motorist or pedestrian subject crossed an intersection after a red light would be greater when the person had been the object of a stare by another motorist in an adjoining lane, or by a pedestrian standing on the curb next to the subject.

Subjects

Subjects were male and female Caucasians, ranging in age from about 16 to 70, who were driving at a busy intersection in a suburban town in northern California. Subjects were selected in both morning and afternoon hours, but not during rush hours.

General Procedure

The basic procedure involved an assistant who either drove up next to the motorist on a motor scooter, or who stood on the curb next to the driver's side on a one-way street. The assistant selected the motorist to be tested, and either stared at him or her for the duration of the red light (experimental condition), or did not stare (control condition). The assistant timed the duration of the stare and the time which the motorist took to cross the intersection after the light changed with a stopwatch hidden in the assistant's pocket. The assistant also recorded the sex and estimated the age of the driver.

Five experiments were conducted to test the basic hypothesis and to eliminate alternative explanations for any increased flight speeds in the experimental condition. Experiment 1 tested the basic hypothesis that the time to cross the intersection would be less in the experimental condition than in the control condition. In this experiment, the assistant rode up next to the driver on a motor scooter, and either stared at the motorist (experimental condition), or did not stare (control condition). Experiment 2 was conducted to test the idea that decreased time was the result of a perception by the motorist that the stare was a challenge to a race. In this experiment, the person doing the staring was on foot on the curb, but other procedures were the same. Experiment 3 was conducted to test for effects and interactions between the sex of the assistant and the sex of the motorist, as a possible confounding factor affecting the speed of withdrawal. In Experiment 4, the timing procedure was applied to pedestrians crossing the street, rather than motorists, to further eliminate any possibility that the effects were specific to automotive situations. Pedestrians stopped at the same red light were either stared at, or were not, and the time which they took to cross the street after the light change was recorded. In Experiment 5, another possible alternative explanation for withdrawal behavior was tested. The alternative explanation rested on the idea that people were withdrawing from an unknown or incongruous type of behavior, rather than reacting to the communicative properties of a stare. In this experiment, the assistant sat on the sidewalk at the intersection and began to pound the pavement with a hammer when a motorist in the experimental condition approached, but did not stare at the motorist. The motorist's transit time across the intersection was then measured. In this experiment, two additional conditions were used. One was the control condition, where no action was taken by the assistant. In the other condition, the assistant stared at the motorist, rather than pounding concrete.

Manipulation

The independent variable present in all experiments was the presence or absence of a stare, a nominal variable. In Experiment 3, the additional nominal independent variables of Sex of Assistant and Sex of Motorist were added and were systematically manipulated to test for differences between males and females for motorists and for experimental as-

Exhibit 15-2 cont.

Table E15-2 Mean Crossing Time for Experiment IV: Eight Experimenters, Pedestrian Subjects

Stare				No Stare			
Male Experimenter		*Female Experimenter*		*Male Experimenter*		*Female Experimenter*	
Subject		Subject		Subject		Subject	
Male	Female	Male	Female	Male	Female	Male	Female
11.1	10.9	11.6	10.9	12.1	11.0	12.8	12.9
(23)	(10)	(16)	(17)	(17)	(14)	(17)	(14)
11.0		11.2		11.6		12.9	
(33)		(33)		(31)		(31)	
11.1*				12.2*			
(66)				(62)			

* Means are significantly different from each other. All other differences are not statistically significant. Copyright 1972 by the American Psychological Association. Reprinted (or adapted) by permission.

sistants, and for differences due to mixed sex conditions (the assistant and the motorist were of different sexes). In Experiment 5, a new nominal independent variable was introduced, the presence or absence of a novel or incongruous behavior (sitting on the curb and hitting the pavement with a hammer).

To assure nonsystematic assignment of subjects to conditions, the experimental assistants relied on several randomizing procedures. In Experiment 1, the assistant was given a paper with the sequence in which she was to assign the subjects to each condition. Motorists were counted each time the assistant pulled up to the traffic light. The paper, which was constructed with a random numbers table, might have said something like "Stare at the second, third, sixth, eighth, ninth…." These motorists were the subjects in the experimental condition. The remaining subjects were not stared at, and made up the control condition.

A more elaborate procedure was used in Experiment 4:

> *Each experimenter had a supply of 16 jelly beans in his/her pocket, 8 each of two different colors. One color signified the stare condition, the other, the no-stare condition. When a potential subject arrived at the crosswalk, the experimenter removed 1 of the jellybeans, noted its color, and ate it. In this manner, a quasi-random assignment of subject to conditions was achieved without the experimenter attracting the attention of the subject and other by-*

Dependent Variable Measures

In each experiment, the amount of time for a motorist or pedestrian to cross the intersection was used. This variable is defined at a ratio level of measurement (Time in Seconds). The research design was post-manipulation only, so only a single measurement of the dependent variable was made.

Results

In all experiments, the time to cross the intersection was shorter in the "Stare" conditions than it was in the control conditions. This is evidence for the ability of a nonverbal communicative act to produce a flight behavior. The ability of the stare to produce this effect was present in pedestrians, as well as motorists, so it is unlikely that the stare was interpreted as an invitation to race, or in any other way connected to automobiles or the driving situation. Effects for the sex of the motorist and experimental assistant were not significant, indicating that increased speed of departure was probably not due to a desire to impress members of the same or of the opposite sex. In Experiment 5, the flight speed was significantly higher for persons stared at than it was for persons who saw the incongruous behavior of pavement pounding.

Ellsworth, P. C., Carlsmith, J. M., & Henson, A. (1972). *Journal of Personality and Social Psychology*. 21(3), 302 –311.

Example of Field Research Using the Social System as the Unit of Analysis: The Fredin Study of the Effect of Interactive Telecommunication in Schools

The final example of field research illustrates a field experiment with a social system as the unit of analysis. In this research, the basic observations are made of entire school faculties. Data is collected from individuals in the schools, but this data is aggregated to provide variables that apply to schools, and not to the individuals.

This example also shows the basic limitations and strengths of field research. The researcher was very constrained in the sample he could use. To try to make the experimental conditions comparable, he matched the units in each group on variables that he thought might affect the relationships that he was testing. The experimental groups are thus neither representative of all work groups, or even of all schools, since there was no random selection from either of these universes. Even assignment to the experimental groups was systematic, rather than random. This was done deliberately, to reduce the chances of random sampling error. This kind of sample raises some potentially troubling questions about the internal validity of the research. Could the results be simply due to the nature of the schools assigned to each group, or to peculiarities of the schools in the single town in which the research was conducted? To the degree to which we are unsure of the answer to this question, we are questioning the internal validity of the study.

Limitations on the sample were amplified by the choice of the social system unit of analysis. By using whole schools as the basic unit, the researcher needed to exert much more effort to obtain a single observation, and the number of possible observations (schools) was limited by geographic and cost considerations. But since the basic theory dealt with the response of social systems (work groups) this choice of unit of analysis was warranted.

On the positive side, the experimental setting is very generalizable (at least to other schools). Teachers used the communication system in ways that were completely natural. There were no restraints on the operation of other relevant processes—teachers still had to budget their time to meet classes, grade papers, and do all their other work, so they gave no excessive amount of attention to the communication system, as they might have done in a laboratory experiment. Interpersonal communication with other teachers was also done under natural conditions.

Another typical difficulty associated with field research is illustrated in this example. Although interpersonal communication is central to the theory being tested, it is not directly observed. The researcher directly measured only the frequency of communication contacts between persons in the groups, and not content of communication. This forced him to speculate about the nature of the interpersonal communication that occurred, rather than being able to observe it and describe it.

But one of the things which gives field research good external validity is the absence of an experimenter peering over the subjects' shoulders while the research is in progress. This has the negative effect of limiting the detail with which the researcher can describe the communication process, while at the same time insuring that the process occurs naturally.

Summary

Field research uses the designs of experimental research, but applies them outside the laboratory. Changing the setting of the research generally increases the external validity (generalizability) of the results, but often at the expense of internal validity. A major problem with field research lies in the representativeness of the sample. Field research subjects are frequently chosen for pragmatic reasons, rather than with probability sampling procedures. Field researchers must often work with subjects that are available, rather than those who are clearly representative of the population. This problem is intensified when social system units of analysis are used.

Another problem lies in the lack of control over factors which may confound the results. Communication processes are allowed to occur under natural conditions, and sometimes these messy conditions can obscure the real relationships among research variables. The researcher cannot observe the communication process in as much detail in a field study, either. Part of the reason for providing a natural setting for the research lies in removing any intrusive observations or measurement.

These limitations are accepted by the field researcher, in return for the increased confidence in

Exhibit 15-3 Experimental Field Research: Social System Unit of Analysis: Telecommunications in Schools

General Topic

The effects of introducing an interactive cable television system into work groups was investigated. In particular, the role of interpersonal communication on the regulation of diversity of new ideas within the groups is examined. Interpersonal communication can either suppress or enhance the diversity of new ideas, depending upon the amount of shared information and ideas within the interpersonal communication network. Interactive telecommunication has been expected to increase the diversity of new ideas in societies, but the author argues that effects of interpersonal communication in regulating diversity of ideas make this effect more complex.

In this research, the introduction of specialized two-way interactive and one-way broadcast cable technology in elementary school systems was studied. Television programs which presented new ideas or innovations in elementary school teaching were produced by the teachers themselves, and broadcast to participating schools. These broadcasts then probably served as gist for interpersonal discussion of teaching methods and materials.

Hypotheses

Because of the conflicting predictions about the ways in which interpersonal communication might regulate diversity, the researchers used a general, nondirectional hypothesis: Work groups (in this case, elementary school faculties) using an interactive cable communication system will show a different relationship between "intermediation" (a characteristic of an interpersonal communication network which is defined below) than will groups not using the system.

Units of Analysis

Instead of looking at individual faculty members within each school, the researchers defined their variables at a social system or group level of analysis. The basic unit was the school. Each school was characterized on its amount of intermediation and the diversity of new ideas about teaching which were discussed within the school.

Definition and Measurement of Variables

Diversity of new ideas was measured by a simple procedure: individual faculty members in the schools were asked to fill out a questionnaire which contained the question:

> *Are there any ideas or methods you've seen or heard about during the past school year about different ways elementary school teachers might do their work?*

The responses were content analyzed (see Chapter 18 for a description of this measurement procedure), and the total number of different responses give by each faculty member was recorded.

To move this individual-level measurement to the social system level, the researchers took the average value of the individual diversity scores for each school. This aggregated value then represented the diversity of new ideas score for the basic unit of analysis, the school.

Intermediation was defined as a group property, and so it was operationally defined directly at the social system level. In groups with high levels of intermediation, most group members are involved in talking to most other group members. In groups with low levels of intermediation, only a few group members communicate with all other group members, or communication is limited to discussion within small subgroups.

The basic measurement used to detect intermediation was a variation of the sociogram. This technique asks members of a group to rate all other members of a group on some characteristic. In this case, it was the amount of communication that occurred between the person filling out the sociogram, and other members of the group:

> *I am going to hand you a list of people who work at your school. Please go through the list and put an "X" by each person you discuss teaching ideas with a least once a week.*

This data was used to compute an intermediation score for each school by a fairly complex operational definition. We'll reproduce it here to show the way that complex

Exhibit 15-3 cont.

ideas like intermediation can be defined by mathematical operational definitions:

$$I = ((H - L) + C) * M$$

Where

H is the number of individuals in a school who were named by others 2 or more times more frequently than the median faculty member in the school.

L is the number of individuals in a school who were named by others 2 or fewer times less frequently than the median faculty member in the school.

C is a constant chosen so that all (H - L) values are positive.

M is the maximum value found in a school for the ratio of the number of times an individual was named by others to the total number of time he or she could have been named.

Manipulation

There were 41 schools included in this study. The schools were split into three groups: 14 schools which were provided with two-way interactive cable systems which teachers could use to view programs, respond to them, and interact with other viewers; a second group of 12 schools which received only broadcast versions of the same programs in which no interaction was possible; and the remaining 15 schools which received neither, and were the control group. Schools were matched across groups for racial makeup and for median adult education levels. Each group of schools represented a class of the nominal independent factor: the type of communication system.

Results

Support was found for the basic hypothesis. Intermediation was found to significantly decrease the diversity of new ideas in the control schools. That is, increased levels of discussion apparently produced some type of consensus which reduced the number of ideas which were actually discussed. But in both the one-way and two-way cable schools, the relationship between interpersonal communication and the diversity of ideas discussed was not statistically significant. Further tests showed that the control group correlation was significantly different from both of the cable groups' correlations, and the cable groups' correlations were not significantly different from each other. In the schools in which cable was introduced, the author speculates, the effect of introducing new ideas via television programs spurred discussion, which counteracted the negative effect of higher levels of intermediation. This produced correlations which were not significantly different from zero, as shown in Table E15-3.

According to the theoretical linkage provided by the researchers, the effects should have been greater for two-way cable than for one-way broadcasts. The correlations for these two groups were in the predicted direction, but the difference was not large enough to be significant. This may have been due to the low power of the test caused by the small number of units of analysis (small N). If you remember the discussion of power in Chapter 12, you'll recall that tests with small numbers of observations are prone to have high Type II (beta) errors, where the researcher finds no evidence for a relationship, when one actually exists. If the researcher could have included more schools in this study, his hypothesis might have been supported more fully.

Fredin, E. S. (1983). The context of communication: Interactive telecommunication, interpersonal communication, and their effect on ideas. *Communication Research*. 10(4), 553-581.

Table E15-3 Correlations between Intermediation and Diversity of New Ideas

Type of Communication System					
Two-Way		*One-Way*		*Control*	
(N = 14)		(N = 12)		(N = 15)	
Correlation	Significance	Correlation	Significance	Correlation	Significance
.32	.14	-.25	.31	-.53	.006

the results that the natural settings provide. This improved external validity permits the researcher to state that any relationships observed in the research study are probably true reflections of the "real world".

A good field research design will attempt to minimize the problems with internal validity, while retaining the benefits of the improved external validity. Sample matching, use of random assignment procedures for experimental groups, unobtrusive observation, and measurement of relevant unmanipulated variables (for example, subject's age and sex) are all important to good field research design.

In fact, the points soon to be made about statistical control in unmanipulated designs can also be applied to field research. In the next chapter, we'll see that measurement and statistical control of relevant variables can be used as a substitute for manipulative control. Use of these techniques in field research can alleviate to some extent the loss of internal validity which a researcher must accept in order to obtain the improved external validity provided by a field research design.

References and Additional Readings

Achen, C. H. (1986). *The statistical analysis of quasi-experiments*. Berkeley, CA: University of California Press. (Chapter 1, "Experiments and Quasi-experiments").

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

Cook, T. D. and Campbell, D.T. (1979). *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand-McNally.

Fiedler, J. (1978). *Field research*. San Francisco: Jossey Bass. (Chapter 1 "Planning for field operations"; Chapter 2 "Research sites and settings").

Wimmer, R.D. & Dominick, J.R. (1987). *Mass media research: An introduction* (2nd Edition). Belmont, CA: Wadsworth. (Chapter 7, "Field Research and Related Research Methods")