Bringing Research to Life

"Oh, did I wake you?" Jasōn asks innocently, as he rubs his hand on his very pregnant wife's abdomen.

Coming groggily awake, Dorrie mumbles, "It's not time. Go back to sleep."

"Dorrie, hon'—I've been thinking . . . we shouldn't wait. We should give them names . . . now."

"Names? Now?"

"We'll call them Terry and Robin. Whether boys or girls, either way the names will fit."

They lay side by side in the darkness, in silence. Silence maybe meant he was thinking and would rouse her again. No use fighting to regain sleep. Hear him out. "Why the hurry, Jason?"

"I want you to talk to them. I want you to say, 'Robin, dearest, this is your mommy,' and 'Terry, your mommy and daddy love you terrifically much.'"

"Sweet. May I go back to sleep now?"

"This has possibilities, you know . . ."

"Please, no *possibilities!*" groans Dorrie. "No twilight zone ideas, grasped at in the gray area between conscious and unconscious, wakefulness and sleep, and proposed with utter seriousness."

"Say you read poetry to Robin. You say, 'Robin, this is for you, and it is by a very famous poet, William Blake . . .'"

"'Tyger, tyger, burning bright . . .'"

"That's the idea. Say you read poetry to Robin and maybe sing to Terry. Well, maybe it is not so good an idea letting you sing to little Terry, because you do not sing nearly as well as you read poetry. We would want the differential treatments to be delivered with more or less equal efficacy, wouldn't we? So, maybe you should read Shakespeare to Terry and Blake to Robin, and then, in puberty, we will see if one has a preference for Shakespeare over Blake."

She rolled over on an elbow. "I see. You are suggesting that as the twins share the same genetic makeup they will differentially emphasize contrasting environmental stimuli administered during the earliest development of their central nervous systems."

"Yes."

"You woke me for this?"

"Yes, whatever treatment . . . stimuli . . . remarks . . . you direct to them through internal dialogue, such as poetry by Blake and Shakespeare, will fall on almost identically genetically endowed organisms . . . Of course, I would have to trust you to adhere to the experimental protocols we devise."

"Following your ridiculous line of thought, perhaps Blake and Shakespeare would not be good alternative treatments, since they wrote in different styles and were not contemporaries . . . assuming we take all of this seriously . . . which I don't."

"I am not sure I have heard you express equal enthusiasm for Blake and Shakespeare, so you will surely bias the administration of the treatment in favor of Blake, your favorite."

"Let me understand this, sweetheart." Dorrie was fully awake now and irritated. "I might not stick to the protocols, and you do not trust me to avoid bias in administering the 'treatments' with equal enthusiasm."

"Well, at the conscious level, I have to trust you, because you are a doctor . . .

". . . and my wife and life's companion . . . and the mother of our children . . .

". . . and you understand the importance of nice clean experimentation. But, yes, biases might be present at the subconscious level and remain difficult to control."

She sat up. "Wanting to call the babies by name is a sentimental impulse. My heart says it is sweet of you to want me to talk to them, but my brain says this is a cruel experiment. Nevertheless, let's talk about 'nice clean experimentation,' as you call it.

"Jason, we are talking about our children here, not genetically created identical twin embryos. First, you cannot differentially direct treatment to one or the other. Never—no, not in my wildest dreams—should I believe one of my unborn babies knows he is Robin or Terry. Second, when they reach an age when they attend nursery school, we will totally lose our ability to differentially apply stimuli . . ."

"Would you consider home schooling?"

"I'm a public health doctor, Jason. Doctors don't work at home! That's why we have hospitals!"

"OK, OK. No need to get huffy."

"Jason, I am experiencing the joys and pains of a multiple pregnancy. There are certain hormonal changes . . . which I have heard you indelicately call 'gland things,' but I am a professional person and am well aware of my responsibilities—to the public, to you, to my unborn children, and to myself—to maintain my emotional equilibrium. So lay off, Jason!"

"I'm sorry I woke you. OK?"

"Waking me is only part of the problem. It's what you woke me for—this lame idea that just popped into your head, this notion to manipulate your own offspring, which you forced on me without any critical thinking in advance."

"Well, I was curious, that's all. I was brainstorming."

"If you were doing research in the university or hospital, you would not float such a lame idea. A human subjects committee would roast you for such an ethically questionable idea, starting with . . . I don't know . . . starting with whether unborn children are able to give informed consent, whether I am able to give consent on their behalf, and ending up, maybe, by asking what right you have to try to alter the artistic sensibilities of anyone's offspring. You don't get away with crazy ideas like this in business research, do you?"

A long silence followed. "Swell," he said. "Can we get back to sleep, then?"

"Let's hope so." She pulled his face close, so that by the light shining in from a street lamp he could see how rigid her lips were. "This year I am part of a program to administer treatments to 150 mortally ill subjects. They know that 75 of them are receiving an experimental treatment and 75 receiving colored saline solution. It's a double-blind experiment, so I have no idea who is receiving the placebo.

"Do you have any idea what it means to know you are withholding the promise of life from 75 human beings? Do you grasp what it means to look into their eyes and have them look back and maybe cry, so an unspoken message passes—'Don't let me die'—?

"When I became a doctor, I expected to do my best to save every life. But my fellow doctors are counting on me for proof, Jason, so they may justify the treatment and request funding. I am not playing games with people, Jason. I am letting people die so others may live.

"Experimentation is about needing to know something so badly that you cannot live without an answer, because others cannot live without an answer. What experimentation is *not* about is having a brainstorm or scratching a mental itch. Experimentation is responsibility."

What Is Experimentation?

Why do events occur under some conditions and not under others? Research methods that answer such questions are called *causal* methods. (Recall the discussion of causality in Chapter 6.) Ex post facto research designs, where a researcher interviews respondents or observes what is or what has been, also have the potential for discovering causality. The distinction between these methods and experimentation is that the researcher is required to accept the world as it is found, whereas an experiment allows the researcher to alter systematically the variables of interest and observe what changes follow.

In this chapter we define experimentation and discuss its advantages and disadvantages. An outline for the conduct of an experiment is presented as a vehicle to introduce important concepts. The questions of internal and external validity are also examined: Does the experimental treatment determine the observed difference, or was some extraneous variable responsible? And, how can one generalize the results of the study across times, settings, and persons? The chapter concludes with a review of the most widely accepted designs and a "Close-Up" example.

You may wish to revisit our discussion of causality in Chapter 6.

Experiments are studies involving intervention by the researcher beyond that required for measurement. The usual intervention is to manipulate some variable in a setting and observe how it affects the subjects being studied (e.g., people or physical entities). The researcher manipulates the independent or explanatory variable and then observes whether the hypothesized dependent variable is affected by the intervention.

An example of such an intervention is the study of bystanders and thieves.[1] In this experiment, students were asked to come to an office where they had an opportunity to see a fellow student steal some money from a receptionist's desk. A confederate of the experimenter, of course, did the stealing. The major hypothesis concerned whether people observing a theft would be more likely to report it (1) if they observed the crime alone or (2) if they were in the company of someone else.

There is at least one **independent variable (IV)** and one **dependent variable (DV)** in a causal relationship. We hypothesize that in some way the IV "causes" the DV to occur. The independent or explanatory variable in our example was the state of either being alone when observing the theft or being in the company of another person. The dependent variable was whether the subjects reported observing the crime. The results suggested that bystanders were more likely to report the theft if they observed it alone rather than in another person's company.

On what grounds did the researchers conclude that people who were alone were more likely to report crimes observed than people in the company of others? Three types of evidence form the basis for this conclusion. First, there must be an agreement between independent and dependent variables. The presence or absence of one is associated with the presence or absence of the other. Thus, more reports of the theft (DV) came from lone observers (IV_1) than from paired observers (IV_2).

Second, beyond the correlation of independent and dependent variables, the time order of the occurrence of the variables must be considered. The dependent variable should not precede the independent variable. They may occur almost simultaneously, or the independent variable should occur before the dependent variable. This requirement is of little concern since it is unlikely that people could report a theft before observing it.

The third important support for the conclusion comes when researchers are confident that other extraneous variables did not influence the dependent variable. To ensure that these other variables are not the source of influence, researchers control their ability to confound the planned comparison. Under laboratory conditions, standardized conditions for control can be arranged. The crime observation experiment was carried out in a laboratory set up as an office. The entire event was staged without the observers' knowledge. The receptionist whose money was to be stolen was instructed to speak and act in a specific way. Only the receptionist, the observers, and the "criminal" were in the office. The same process was repeated with each trial of the experiment.

While such controls are important, further precautions are needed so that the results achieved reflect only the influence of the independent variable on the dependent variable.

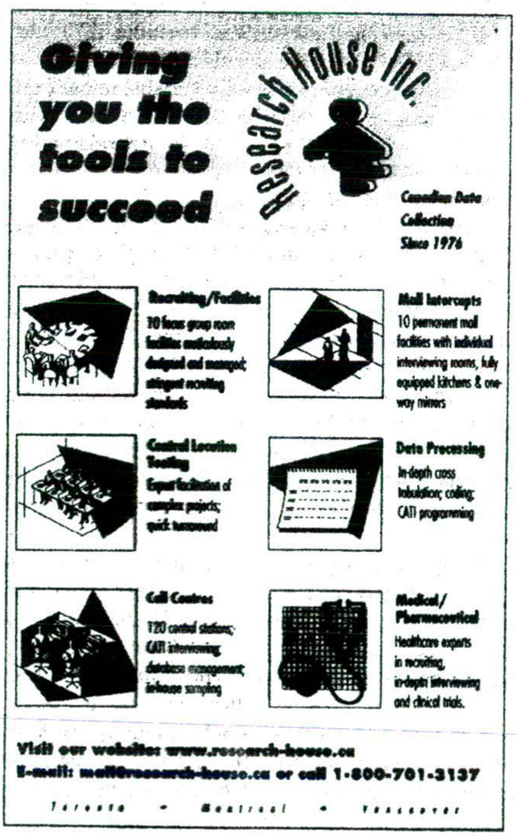
An Evaluation of Experiments

Advantages

When we elaborated on the concept of cause in Chapter 6, we said causality could not be proved with certainty but the probability of one variable being linked to another could be established convincingly. The experiment comes closer than any primary data collection method to accomplishing this goal. The foremost advantage is the researcher's ability to manipulate the independent variable. Consequently, the probability that changes in the dependent variable are a function of that manipulation increases. Also, a control group serves as a comparison to assess the existence and potency of the manipulation.

The second advantage of the experiment is that contamination from extraneous variables can be controlled more effectively than in other designs. This helps the researcher

Experimental research frequently requires special expertise and facilities that can accommodate multistage designs, like the facilities offered by the Canadian firm Research House Inc. www.research-house.ca



isolate experimental variables and evaluate their impact over time. Third, the convenience and cost of experimentation are superior to other methods. These benefits allow the experimenter opportunistic scheduling of data collection and the flexibility to adjust variables and conditions that evoke extremes not observed under routine circumstances. In addition, the experimenter can assemble combinations of variables for testing rather than having to search for their fortuitous appearance in the study environment.

Fourth, **replication**—repeating an experiment with different subject groups and conditions—leads to the discovery of an average effect of the independent variable across people, situations, and times. Fifth, researchers can use naturally occurring events and, to some extent, field experiments to reduce subjects' perceptions of the researcher as a source of intervention or deviation in their everyday lives.

Disadvantages

The artificiality of the laboratory is arguably the primary disadvantage of the experimental method. However, many subjects' perceptions of a contrived environment can be improved by investment in the facility. Second, generalization from nonprobability samples can pose problems despite random assignment. The extent to which a study can be generalized from college students to managers or executives is open to question. And when an experiment is unsuccessfully disguised, volunteer subjects are often those with the most interest in the topic. Third, despite the low costs of experimentation, many applications of experimentation far outrun the budgets for other primary data collection methods. Fourth, experimentation is most effectively targeted at problems of the present or immediate future. Experimental studies of the past are not feasible, and studies about intentions or predictions are difficult. Finally, management research is often concerned with the study of people. There are limits to the types of manipulation and controls that are ethical.

Conducting an Experiment[2]

In a well-executed experiment, researchers must complete a series of activities to carry out their craft successfully. Although the experiment is the premier scientific methodology for establishing causation, the resourcefulness and creativeness of the researcher are needed to make the experiment live up to its potential. In this section, we discuss seven activities the researcher must accomplish to make the endeavor successful:

MANAGEMENT Tip

1. Select relevant variables.
2. Specify the level(s) of the treatment.
3. Control the experimental environment.
4. Choose the experimental design.
5. Select and assign the subjects.
6. Pilot-test, revise, and test.
7. Analyze the data.

Selecting Relevant Variables

Throughout the book we have discussed the idea that a research problem can be conceptualized as a hierarchy of questions starting with a management problem. The researcher's task is to translate an amorphous problem into the question or hypothesis that best states the objectives of the research. Depending on the complexity of the problem, investigative questions and additional hypotheses can be created to address specific facets of the study or data that need to be gathered. Further, we have mentioned that a **hypothesis** is a relational statement because it describes a relationship between two or more variables. It must also be **operationalized,** a term we used earlier in discussing how concepts are transformed into variables to make them measurable and subject to testing.

Consider the following research question as we work through the seven points listed above:

> Does a sales presentation that describes product benefits in the introduction of the message lead to improved retention of product knowledge?

Since a hypothesis is a tentative statement—a speculation—about the outcome of the study, it might take this form:

> Sales presentations in which the benefits module is placed in the introduction of a 12-minute message produce better retention of product knowledge than those where the benefits module is placed in the conclusion.

The researchers' challenges at this step are to

MANAGEMENT Tip

1. Select variables that are the best operational representations of the original concepts.
2. Determine how many variables to test.
3. Select or design appropriate measures for them.

The researchers would need to select variables that best operationalize the concepts *sales presentation, product benefits, retention,* and *product knowledge.* The product's classification and the nature of the intended audience should also be defined. In addition, the term *better* could be operationalized statistically by means of a significance test.

The number of variables in an experiment is constrained by the project budget, the time allocated, the availability of appropriate controls, and the number of subjects being tested. For statistical reasons, there must be more subjects than variables.[3]

The selection of measures for testing requires a thorough review of the available literature and instruments. In addition, measures must be adapted to the unique needs of the research situation without compromising their intended purpose or original meaning.

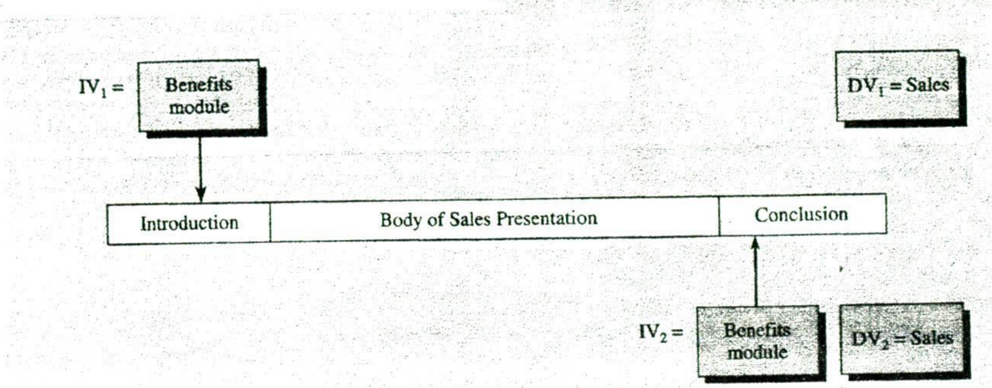
Specifying the Levels of Treatment

The **treatment levels** of the independent variable are the distinctions the researcher makes between different aspects of the treatment condition. For example, if salary is hypothesized to have an effect on employees exercising stock purchase options, it might be divided into high, middle, and low ranges to represent three levels of the independent variable.

The levels assigned to an independent variable should be based on simplicity and common sense. In the sales presentation example, the experimenter should not select 8 minutes and 10 minutes as the starting points to represent the two treatment levels if the average message about the product is 12 minutes long. Similarly, if the benefits module is placed in the first and second minutes of the presentation, observable differences may not occur because the levels are too close together. Thus, in the first trial, the researcher is likely to position the midpoint of the benefits module the same interval from the end of the introduction as from the end of the conclusion (see Exhibit 14–1).

Under an entirely different hypothesis, several levels of the independent variable may be needed to test order-of-presentation effects. Here we use only two. Alternatively, a **control group** could provide a base level for comparison. The control group is composed of subjects who are not exposed to the independent variable(s), in contrast to those who receive the **experimental treatment** (manipulation of the independent variable[s]).

EXHIBIT 14–1 Benefits Module Effectiveness Based on Timing of Inclusion



Controlling the Experimental Environment

Chapter 2 discussed the nature of extraneous variables and the need for their control.

In our sales presentation experiment, extraneous variables can appear as differences in age, gender, race, dress, communications competence, and many other characteristics of the presenter, the message, or the situation. These have the potential for distorting the effect of the treatment on the dependent variable and must be controlled or eliminated. However, at this stage, we are principally concerned with **environmental control,** holding constant the physical environment of the experiment. The introduction of the experiment to the subjects and the instructions would likely be videotaped for consistency. The arrangement of the room, the time of administration, the experimenter's contact with the subjects, and so forth, must all be consistent across each administration of the experiment.

Dorrie described a double-blind study in the opening vignette.

Other forms of control involve subjects and experimenters. When subjects do not know if they are receiving the experimental treatment, they are said to be **blind.** When the experimenters do not know if they are giving the treatment to the experimental group or to the control group, the experiment is said to be **double blind.** Both approaches control unwanted complications such as subjects' reactions to expected conditions or experimenter influence.

Choosing the Experimental Design

Many of the experimental designs are diagrammed and described later in this chapter.

Unlike the general descriptors of research design that were discussed in Chapter 6, experimental designs are unique to the experimental method. They serve as positional and statistical plans to designate relationships between experimental treatments and the experimenter's observations or measurement points in the temporal scheme of the study. In the conduct of the experiment, the researchers apply their knowledge to select one design that is best suited to the goals of the research. Judicious selection of the design improves the probability that the observed change in the dependent variable was caused by the manipulation of the independent variable and not by another factor. It simultaneously strengthens the generalizability of results beyond the experimental setting.

Selecting and Assigning Subjects

The subjects selected for the experiment should be representative of the population to which the researcher wishes to generalize the study's results. This may seem self-evident, but we have witnessed several decades of experimentation with college sophomores that contradict that assumption. In the sales presentation example, corporate

SNAPSHOT

Effect of Magazine Advertising on Sales

For the first time, the Magazine Publishers of America (MPA) has a definitive study demonstrating that magazine advertising does positively affect not only the incidence of sales but also the dollar value and quantity of sales. ACNielsen sent 50,000 households in its ACNielsen Household Scanner Panel™ a four-color questionnaire featuring the covers of April, May, and June issues of 14 magazines. Panelists scanned the bar codes of the covers of the magazines they had read. The scanned information was uploaded to ACNielsen, where demographically matched panels of 4,000 households each were constructed. Half of each panel had been exposed to magazine ads for 1 of the 10 brands being tracked, while the other half had not. Actual sales data drawn from records of scanned purchases were compared. Households exposed to magazine ads were more likely to purchase those brands, and dollar sales also increased among 8 of the 10 brands studied. You can link to the complete MPA study report from our text website or learn more about the ACNielsen Household Scanner Panel™.

www.magazine.org

www.acnielsen.com

buyers, purchasing managers, or others in a decision-making capacity would provide better generalizing power than undergraduate college students *if* the product in question was targeted for industrial use rather than to the consumer.

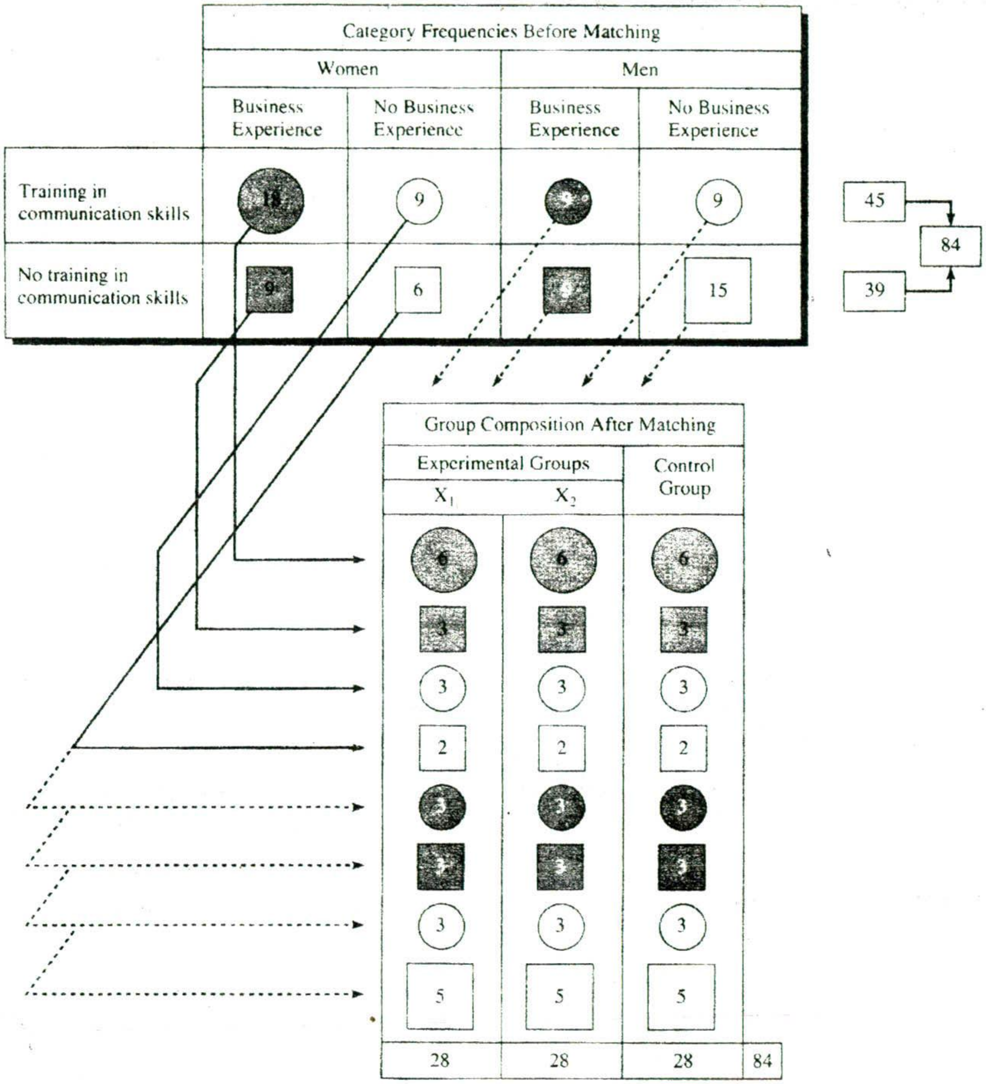
We discussed random sampling in Chapter 7.

The procedure for random sampling of experimental subjects is similar in principle to the selection of respondents for a survey. The researcher first prepares a sampling frame and then assigns the subjects for the experiment to groups using a randomization technique. Systematic sampling may be used if the sampling frame is free from any form of periodicity that parallels the sampling ratio. Since the sampling frame is often small, experimental subjects are recruited; thus they are a self-selecting sample. However, if randomization is used, those assigned to the experimental group are likely to be similar to those assigned to the control group. **Random assignment** to the groups is required to make the groups as comparable as possible with respect to the dependent variable. Randomization does not guarantee that if a pretest of the groups was conducted before the treatment condition, the groups would be pronounced identical; but it is an assurance that those differences remaining are randomly distributed. In our example, we would need three randomly assigned groups—one for each of the two treatments and one for the control group.

When it is not possible to randomly assign subjects to groups, **matching** may be used. Matching employs a nonprobability quota sampling approach. The object of matching is to have each experimental and control subject matched on every characteristic used in the research. This becomes more cumbersome as the number of variables and groups in the study increases. Since the characteristics of concern are only those that are correlated with the treatment condition or the dependent variable, they are easier to identify, control, and match.[4] In the sales presentation experiment, if a large part of the sample was composed of businesswomen who had recently completed communications training, we would not want the characteristics of gender, business experience, and communication training to be disproportionately assigned to one group.

Some authorities suggest a **quota matrix** as the most efficient means of visualizing the matching process.[5] In Exhibit 14–2, one-third of the subjects from each cell of the matrix would be assigned to each of the three groups. If matching does not alleviate the assignment problem, a combination of matching, randomization, and increasing the sample size would be used.

EXHIBIT 14–2 Quota Matrix Example



Pilot-Testing, Revising, and Testing

The procedures for this stage are similar to those for other forms of primary data collection. Pilot testing is intended to reveal errors in the design and improper control of extraneous or environmental conditions. Pretesting the instruments permits refinement before the final test. This is the researcher's best opportunity to revise scripts, look for control problems with laboratory conditions, and scan the environment for factors that might confound the results. In field experiments, researchers are sometimes caught off guard by events that have a dramatic effect on subjects: the test marketing of a competitor's product announced before an experiment, or a reduction in force, reorganization, or merger before a crucial organizational intervention. The experiment should be timed so that subjects are not sensitized to the independent variable by factors in the environment.

Analyzing the Data

If adequate planning and pretesting have occurred, the experimental data will take an order and structure uncommon to surveys and unstructured observational studies. It is not that data from experiments are easy to analyze; they are simply more conveniently arranged because of the levels of the treatment condition, pretests and post-tests, and the group structure. The choice of statistical techniques is commensurately simplified.

Researchers have several measurement and instrument options with experiments. Among them are

- Observational techniques and coding schemes.
- Paper-and-pencil tests.
- Self-report instruments with open-ended or closed questions.
- Scaling techniques (e.g., Likert scales, semantic differentials, Q-sort).
- Physiological measures (e.g., galvanic skin response, EKG, voice pitch analysis, eye dilation).

Validity in Experimentation

Even when an experiment is the ideal research design, it is not without problems. There is always a question about whether the results are true. We have previously defined validity as whether a measure accomplishes its claims. While there are several different types of validity, here only the two major varieties are considered: **internal validity**—do the conclusions we draw about a demonstrated experimental relationship truly imply cause?—and **external validity**—does an observed causal relationship generalize across persons, settings, and times?[6] Each type of validity has specific threats we need to guard against.

Internal Validity

Among the many threats to internal validity, we consider the following seven:

- History
- Maturation
- Testing
- Instrumentation
- Selection
- Statistical regression
- Experimental mortality

History During the time that an experiment is taking place, some events may occur that confuse the relationship being studied. In many experimental designs, we take a control measurement (O_1) of the dependent variable before introducing the manipulation (X). After the manipulation, we take an after-measurement (O_2) of the dependent variable. Then the difference between O_1 and O_2 is the change that the manipulation has caused.

A company's management may wish to find the best way to educate its workers about the financial condition of the company before this year's labor negotiations. To assess the value of such an effort, managers give employees a test on their knowledge of the company's finances (O_1). Then they present the educational campaign (X) to these employees, after which they again measure their knowledge level (O_2). This design, known as a pre-experiment because it is not a very strong design, can be diagrammed as follows:

O_1	X	O_2
Pretest	Manipulation	Post-test

Between O_1 and O_2, however, many events could occur to confound the effects of the education effort. A newspaper article might appear about companies with financial problems, a union meeting might be held at which this topic is discussed, or another occurrence could distort the effects of the company's education test.

Maturation Changes also may occur within the subject that are a function of the passage of time and are not specific to any particular event. These are of special concern when the study covers a long time, but they may also be factors in tests that are as short as an hour or two. A subject can become hungry, bored, or tired in a short time, and this condition can affect response results.

Testing The process of taking a test can affect the scores of a second test. The mere experience of taking the first test can have a learning effect that influences the results of the second test.

Instrumentation This threat to internal validity results from changes between observations in either the measuring instrument or the observer. Using different questions at each measurement is an obvious source of potential trouble, but using different observers or interviewers also threatens validity. There can even be an instrumentation problem if the same observer is used for all measurements. Observer experience, boredom, fatigue, and anticipation of results can all distort the results of separate observations.

Selection An important threat to internal validity is the differential selection of subjects for experimental and control groups. Validity considerations require that the groups be equivalent in every respect. If subjects are randomly assigned to experimental and control groups, this selection problem can be largely overcome. Additionally, matching the members of the groups on key factors can enhance the equivalence of the groups.

Statistical Regression This factor operates especially when groups have been selected by their extreme scores. Suppose we measure the output of all workers in a department for a few days before an experiment and then conduct the experiment with only those workers whose productivity scores are in the top 25 percent and bottom 25 percent. No matter what is done between O_1 and O_2, there is a strong tendency for the average of the high scores at O_1 to decline at O_2 and for the low scores at O_1 to increase. This tendency results from imperfect measurement that, in effect, records some persons abnormally high and abnormally low at O_1. In the second measurement, members of both groups score more closely to their long-run mean scores.

Experiment Mortality This occurs when the composition of the study groups changes during the test. Attrition is especially likely in the experimental group and with each dropout, the group changes. Because members of the control group are not affected by the testing situation, they are less likely to withdraw. In a compensation incentive study, some employees might not like the change in compensation method and may withdraw from the test group; this action could distort the comparison with the control group that has continued working under the established system, perhaps without knowing a test is under way.

All the threats mentioned to this point are generally, but not always, dealt with adequately in experiments by random assignment. However, five additional threats to internal validity are independent of whether or not one randomizes.[7] The first three have the effect of equalizing experimental and control groups.

1. **Diffusion or imitation of treatment.** If people in the experimental and control groups talk, then those in the control group may learn of the treatment, eliminating the difference between the groups.
2. **Compensatory equalization.** Where the experimental treatment is much more desirable, there may be an administrative reluctance to deprive the control group members. Compensatory actions for the control groups may confound the experiment.
3. **Compensatory rivalry.** This may occur when members of the control group know they are in the control group. This may generate competitive pressures, causing the control group members to try harder.
4. **Resentful demoralization of the disadvantaged.** When the treatment is desirable and the experiment is obtrusive, control group members may become resentful of their deprivation and lower their cooperation and output.
5. **Local history.** The regular history effect already mentioned impacts both experimental and control groups alike. However, when one assigns all experimental persons to one group session and all control people to another, there is a chance for some idiosyncratic event to confound results. This problem can be handled by administering treatments to individuals or small groups that are randomly assigned to experimental or control sessions.

External Validity

Internal validity factors cause confusion about whether the experimental treatment (X) or extraneous factors are the source of observation differences. In contrast, external validity is concerned with the interaction of the experimental treatment with other factors and the resulting impact on the ability to generalize to (and across) times, settings, or persons. Among the major threats to external validity are the following interactive possibilities.

The Reactivity of Testing on X The reactive effect refers to sensitizing subjects via a pretest so they respond to the experimental stimulus (X) in a different way. A before-measurement of a subject's knowledge about the ecology programs of a company will often sensitize the subject to various experimental communication efforts that might be made about the company. This before-measurement effect can be particularly significant in experiments where the IV is a change in attitude.

Interaction of Selection and X The process by which test subjects are selected for an experiment may be a threat to external validity. The population from which one selects subjects may not be the same as the population to which one wishes to generalize results. Suppose you use a selected group of workers in one department for a test of the piecework incentive system. The question may remain as to whether you can extrapolate those results to all production workers. Or consider a study in which you ask a cross-section of a population to participate in an experiment, but a substantial number refuses. If you conduct the experiment only with those who agree to participate (self-selection), can the results be generalized to the total population?

Other Reactive Factors The experimental settings themselves may have a biasing effect on a subject's response to *X*. An artificial setting can obviously produce results that are not representative of larger populations. Suppose the workers who are given the incentive pay are moved to a different work area to separate them from the control group. These new conditions alone could create a strong reactive condition.

If subjects know they are participating in an experiment, there may be a tendency to role-play in a way that distorts the effects of *X*. Another reactive effect is the possible interaction between *X* and subject characteristics. An incentive pay proposal may be more effective with persons in one type of job, with a certain skill level, or with a certain personality trait.

Problems of internal validity can be solved by the careful design of experiments, but this is less true for problems of external validity. External validity is largely a matter of generalization, which, in a logical sense, is an inductive process of extrapolating beyond the data collected. In generalizing, we estimate the factors that can be ignored and that will interact with the experimental variable. Assume that the closer two events are in time, space, and measurement, the more likely they are to follow the same laws. As a rule of thumb, first seek internal validity. Try to secure as much external validity as is compatible with the internal validity requirements by making experimental conditions as similar as possible to conditions under which the results will apply.

Experimental Research Designs

The many experimental designs vary widely in their power to control contamination of the relationship between independent and dependent variables. The most widely accepted designs are based on this characteristic of control: (1) pre-experiments, (2) true experiments, and (3) field experiments (see Exhibit 14–3).

Pre-Experimental Designs

All three pre-experimental designs are weak in their scientific measurement power—that is, they fail to control adequately the various threats to internal validity. This is especially true of the one-shot case study.

One-Shot Case Study This may be diagrammed as follows:

X	O	
Treatment or manipulation of independent variable	Observation or measurement of dependent variable	**(1)**

An example is an employee education campaign about the company's financial condition without a prior measurement of employee knowledge. Results would reveal only how much the employees know after the education campaign, but there is no way to judge the effectiveness of the campaign. How well do you think this design would meet the various threats to internal validity? The lack of a pretest and control group makes this design inadequate for establishing causality.

EXHIBIT 14–3 Key to Design Symbols

X	An X represents the introduction of an experimental stimulus to a group. The effects of this independent variable(s) are of major interest.
O	An O identifies a measurement or observation activity.
R	An R indicates that the group members have been randomly assigned to a group.

The X's and O's in the diagram are read from left to right in temporal order.

$O \; X \; O \; O$

Time →

X's and O's vertical to each other indicate that the stimulus and/or observation take place simultaneously.

$X \; O$

$\quad\;\; O$

Time →

Parallel rows that are not separated by dashed lines indicate that comparison groups have been equalized by the randomization process.

$X \; O$

$\quad\;\; O$

Those separated with a dashed line have not been so equalized.

$O \; X \; O$

- - - - -

$\qquad\;\; O$

One-Group Pretest–Post-Test Design This is the design used earlier in the educational example. It meets the various threats to internal validity better than the one-shot case study, but it is still a weak design. How well does it control for history? Maturation? Testing effect? The others?

O	X	O	
Pretest	Manipulation	Post-test	(2)

Static Group Comparison This design provides for two groups, one of which receives the experimental stimulus while the other serves as a control. In a field setting, imagine this scenario. A forest fire or other natural disaster is the experimental treatment, and psychological trauma (or property loss) suffered by the residents is the measured outcome. A pretest before the forest fire would be possible, but not on a large scale (as in the California fires). Moreover, timing of the pretest would be problematic. The control group, receiving the post-test, would consist of residents whose property was spared.

$$\begin{array}{cc} X & O_1 \\ \text{- - - - - - - -} & \text{- - - - - - -} \\ & O_2 \end{array} \qquad (3)$$

The addition of a comparison group creates a substantial improvement over the other two designs. Its chief weakness is that there is no way to be certain that the two groups are equivalent.

True Experimental Designs

The major deficiency of the pre-experimental designs is that they fail to provide comparison groups that are truly equivalent. The way to achieve equivalence is through matching and random assignment. With randomly assigned groups, we can employ tests of statistical significance of the observed differences.

It is common to show an X for the test stimulus and a blank for the existence of a control situation. This is an oversimplification of what really occurs. More precisely, there is an X_1 and an X_2, and sometimes more. The X_1 identifies one specific independent variable while X_2 is another independent variable that has been chosen, often arbitrarily.

SNAPSHOT

Vanguard Experiments with Philips Electronics' 401k Savings Rates

Vanguard, a major provider of retirement benefit programs, is conducting an experiment within Philips Electronics North America to determine whether employees can be encouraged to increase the amount they save in their 401k retirement plans. When asked if they could increase their savings, most employees indicated that they live "paycheck to paycheck" and therefore cannot save more. Yet financial planners know that most people can save 1 percent, 3 percent, or even 5 percent more of their income over time and not notice a difference in their standard of living. The Vanguard/Philips experiment attempts to overcome this "painful to save" barrier by having workers agree to save more in the *future*—not today. In the experiment, which began in February 2002, about 800 workers in two geographically separate and distinct divisions of Philips (D_1 and D_2) have been invited to join the SmartSave program. Under the program, they have the choice of increasing their 401k savings rate by 1, 2, or 3 percent drawn from a future pay increase. The rate change will occur on April 1 of each year, at the time of future merit increases. Whatever rate they choose, that increase will occur each April during the life of the experiment, unless they decide to discontinue or increase their savings rate.

SmartSave is being introduced with lots of fanfare, including a newsletter, two teaser postcards, SmartSave posters in the workplace, a required-attendance meeting on company time in which the program will be explained, and company raffles for participants. Additionally, workers in D_2 are being offered one-on-one meetings with a local financial planner. Vanguard and Philips will analyze several pre- and post-metrics:

- Number of people enrolled in the Philips 401k plan.
- Distribution of SmartSave participants at the 1 percent, 2 percent, and 3 percent levels.
- The average 401k savings rate.
- The SmartSave participation rate.
- The number of SmartSave participants who in April choose to abandon, continue, or increase their rate increase.

The experiment involves fewer than 10 percent of Philips employees, but SmartSave will be expanded if savings and participation rates increase. If successful, Vanguard will have a tool to boost the assets it manages in retirement plans, while helping thousands of Americans enjoy a more secure retirement—a win-win situation from any perspective. Can you diagram the Vanguard/Phillips experiment?

www.philips.com

www.vanguard.com

as the control case. Different levels of the same independent variable may also be used, with one level serving as the control.

Pretest–Post-Test Control Group Design This design consists of adding a control group to the one-group pretest–post-test design and assigning the subjects to either of the groups by a random procedure (R). The diagram is:

$$\begin{array}{cccc} R & O_1 & X & O_2 \\ R & O_3 & & O_4 \end{array} \qquad (4)$$

The effect of the experimental variable is

$$E = (O_2 - O_1) - (O_4 - O_3)$$

In this design, the seven major internal validity problems are dealt with fairly well, although there are still some difficulties. Local history may occur in one group and not the other. Also, if communication exists between people in test and control groups, there can be rivalry and other internal validity problems.

Maturation, testing, and regression are handled well because one would expect them to be felt equally in experimental and control groups. Mortality, however, can be a problem if there are different dropout rates in the study groups. Selection is adequately dealt with by random assignment.

The record of this design is not as good on external validity, however. There is a chance for a reactive effect from testing. This might be a substantial influence in attitude change studies where pretests introduce unusual topics and content. Nor does this design ensure against reaction between selection and the experimental variable. Even random selection may be defeated by a high decline rate by subjects. This would result in using a disproportionate share of people who are essentially volunteers and who may not be typical of the population. If this occurs, we will need to replicate the experiment several times with other groups under other conditions before we can be confident of external validity.

Post-Test-Only Control Group Design In this design, the pretest measurements are omitted. Pretests are well established in classical research design but are not really necessary when it is possible to randomize. The design is:

$$\begin{array}{ccc} R & X & O_1 \\ R & & O_2 \end{array} \qquad (5)$$

The experimental effect is measured by the difference between O_1 and O_2. The simplicity of this design makes it more attractive than the pretest–post-test control group design. Internal validity threats from history, maturation, selection, and statistical regression are adequately controlled by random assignment. Since the subjects are measured only once, the threats of testing and instrumentation are reduced, but different mortality rates between experimental and control groups continue to be a potential problem. The design reduces the external validity problem of testing interaction effect, although other problems remain.

Extensions of True Experimental Designs

True experimental designs have been discussed in their classical forms, but researchers normally use an operational extension of the basic design. These extensions differ from the classical design forms in (1) the number of different experimental stimuli that are considered simultaneously by the experimenter and (2) the extent to which assignment procedures are used to increase precision.

Researchers know that as many as 60 percent of purchase decisions are made in the store. Thus marketers aggressively seek in-store space to place temporary displays, shelf-talkers, and instant coupons, as well as ceiling signs and banners. Even the floor is contested real estate. So the ability to demonstrate the effectiveness of promotional materials is critical. FLOORgraphics, Inc., uses a longitudinal design, tracking sales of products in matched groups of stores (test and control groups). After test stores receive the FLOORad, relative sales in both groups are again compared to pre-ad performance and to each other. Research shows the FLOORad effect (the percentage sales increase directly due to the FLOORad) can lift sales 20–40 percent depending on the product category. www.floorgraphics.com

Before we consider the types of extensions, some terms that are commonly used in the literature of applied experimentation must be introduced. **Factor** is widely used to denote an independent variable. Factors are divided into treatment levels, which represent various subgroups. A factor may have two or more levels, such as (1) male and female; (2) large, medium, and small; or (3) no training, brief training, and extended training. These levels should be operationally defined.

Factors also may be classified by whether the experimenter can manipulate the levels associated with the subject. **Active factors** are those the experimenter can manipulate by causing a subject to receive one level or another. Treatment is used to denote the different levels of active factors. With the second type, the **blocking factor**, the experimenter can only identify and classify the subject on an existing level. Gender, age group, customer status, and organizational rank are examples of blocking factors, because the subject comes to the experiment with a pre-existing level of each.

Check this website for examples of industrial experiments: ***http://www.statsoft.com/textbook/stathome.html.***

Up to this point, the assumption is that experimental subjects are people, but this is often not so. A better term for subject is **test unit;** it can refer equally well to an individual, organization, geographic market, animal, machine type, mix of materials, and innumerable other entities.

Completely Randomized Design The basic form of the true experiment is a completely randomized design. To illustrate its use, and that of more complex designs, consider a decision now facing the pricing manager at the Top Cannery. He would like to know what the ideal difference in price is between the Top's private brand of canned vegetables and national brands such as Del Monte and Stokely's.

SNAPSHOT A Nose for Problem Odors

Ever wonder how consumer product companies test the effectiveness of their creations? At Hill Top Research, Inc., founded in 1947 and the largest consumer product testing firm in the world, they use a variety of devices—including the human nose. In one deodorant study subjects were brought to a test site that contained a *hot room*. Researchers applied the product being tested to each subject's armpit, followed by the insertion of a cotton pad under each arm which subjects retained by pressing their arms to their sides. Researchers then led subjects to the *hot room*—where temperatures are warm enough to make anyone sweat. When the subjects exit the room after the defined period of time, the cotton pad is removed for analysis, then the odor detective does his or her job. A cup with a small hole in the bottom is placed against the subject's armpit (to assure uniform distance between nose and pit), and then the detective positions her nose near the hole and inhales. With a successful formulation, the odor detective does not detect a strong or offensive odor. What are some of the variables a researcher would need to control in this study? What sources of error must be controlled?

www.hill-top.com

It is possible to set up an experiment on price differentials for canned green beans. Eighteen company stores and three price spreads (treatment levels) of 7 cents, 12 cents, and 17 cents between the company brand and national brands are used for the study. Six of the stores are assigned randomly to each of the treatment groups. The price differentials are maintained for a period, and then a tally is made of the sales volumes and gross profits of the canned green beans for each group of stores.

This design can be diagrammed as follows:

$$\begin{array}{llll} R & O_1 & X_1 & O_2 \\ R & O_3 & X_3 & O_4 \\ R & O_5 & X_5 & O_6 \end{array} \qquad (6)$$

Here, O_1, O_3, and O_5 represent the total gross profits for canned green beans in the treatment stores for the month before the test. X_1, X_3, and X_5 represent 7-cent, 12-cent, and 17-cent treatments, while O_2, O_4, and O_6 are the gross profits for the month after the test started.

It is assumed that the randomization of stores to the three treatment groups was sufficient to make the three store groups equivalent. Where there is reason to believe this is not so, we must use a more complex design.

Randomized Block Design When there is a single major extraneous variable, the randomized block design is used. Random assignment is still the basic way to produce equivalence among treatment groups, but something more may be needed for two reasons. The more critical reason is that the sample being studied may be so small that it is risky to depend on random assignment alone to guarantee equivalence. Small samples, such as the 18 company stores, are typical in field experiments because of high costs or because few test units are available. Another reason for blocking is to learn whether treatments bring different results among various groups of subjects.

Consider again the canned green beans pricing experiment. Assume there is reason to believe that lower-income families are more sensitive to price differentials than are higher-income families. This factor could seriously distort our results unless we stratify the stores by customer income. Therefore, each of the 18 stores is assigned to one of three income blocks and randomly assigned, within blocks, to the price difference treatments. The design is shown in the accompanying table.

In this design, one can measure both main effects and interaction effects. The **main effect** is the average direct influence that a particular treatment has independent of other factors. The **interaction effect** is the influence of one factor on the effect of another. The main effect of each price differential is secured by calculating the impact of each of the three treatments averaged over the different blocks. Interaction effects occur if you find that different customer income levels have a pronounced influence on customer reactions to the price differentials. (See Chapter 17, "Hypothesis Testing.")

Active Factor—Price Difference		*Blocking Factor—Customer Income* High	Medium	Low	
7 cents	R	X_1	X_1	X_1	
12 cents	R	X_2	X_2	X_2	(7)
17 cents	R	X_3	X_3	X_3	

Note: The O's have been omitted. The horizontal rows no longer indicate a time sequence, but various levels of the blocking factor. However, before-and-after measurements are associated with each of the treatments.

Whether the randomized block design improves the precision of the experimental measurement depends on how successfully the design minimizes the variance within blocks and maximizes the variance between blocks. If the response patterns are about the same in each block, there is little value to the more complex design. Blocking may be counterproductive.

Latin Square Design The Latin square design may be used when there are two major extraneous factors. To continue with the pricing example, assume we decide to block on the size of store and on customer income. It is convenient to consider these two blocking factors as forming the rows and columns of a table. Each factor is divided into three levels to provide nine groups of stores, each representing a unique combination of the two blocking variables. Treatments are then randomly assigned to these cells so that a given treatment appears only once in each row and column. Because of this restriction, a Latin square must have the same number of rows, columns, and treatments. The design looks like the table below.

Store Size	*Customer Income* High	Medium	Low	
Large	X_3	X_1	X_2	(8)
Medium	X_2	X_3	X_1	
Small	X_1	X_2	X_3	

Treatments can be assigned by using a table of random numbers to set the order of treatment in the first row. For example, the pattern may be 3, 1, 2 as shown above. Following this, the other two cells of the first column are filled similarly, and the remaining treatments are assigned to meet the restriction that there can be no more than one treatment type in each row and column.

The experiment is carried out, sales results are gathered, and the average treatment effect is calculated. From this, we can determine the main effect of the various price spreads on the sales of company and national brands. With cost information, we can discover which price differential produces the greatest margin.

A limitation of the Latin square is that we must assume there is no interaction between treatments and blocking factors. Therefore, we cannot determine the interrelationships among store size, customer income, and price spreads. This limitation exists because there is not an exposure of all combinations of treatments, store sizes, and customer income groups. To do so would take a table of 27 cells, while this one has only 9. This can be accomplished by repeating the experiment twice to furnish the number needed to provide for every combination of store size, customer income, and treatment. If one is not especially interested in interaction, the Latin square is much more economical.

Factorial Design One commonly held misconception about experiments is that the researcher can manipulate only one variable at a time. This is not true; with factorial designs, you can deal with more than one treatment simultaneously. Consider again the pricing experiment. The president of the chain might also be interested in finding the effect of posting unit prices on the shelf to aid shopper decision making. The accompanying table can be used to design an experiment that includes both the price differentials and the unit pricing.

	Price Spread		
Unit Price Information?	7 Cents	12 Cents	17 Cents
Yes	X_1Y_1	X_1Y_2	X_1Y_3
No	X_2Y_1	X_2Y_2	X_2Y_3

(9)

This is known as a 2×3 factorial design in which we use two factors: one with two levels and one with three levels of intensity. The version shown here is completely randomized, with the stores being randomly assigned to one of six treatment combinations. With such a design, it is possible to estimate the main effects of each of the two independent variables and the interactions between them. The results can help to answer the following questions:

1. What are the sales effects of the different price spreads between company and national brands?
2. What are the sales effects of using unit-price marking on the shelves?
3. What are the sales-effect interrelations between price spread and the presence of unit-price information?

We discuss the statistical aspects of covariance analysis when we present analysis of variance (ANOVA) in Chapter 17.

Covariance Analysis We have discussed direct control of extraneous variables through blocking. It is also possible to apply some degree of indirect statistical control on one or more variables through analysis of covariance. Even with randomization, one may find that the before-measurement shows an average knowledge level difference between experimental and control groups. With covariance analysis, one can adjust statistically for this before-difference. Another application might occur if the canned green beans pricing experiment were carried out with a completely randomized design, only to reveal a contamination effect from differences in average customer income levels. With covariance analysis, one can still do some statistical blocking on average customer income even after the experiment has been run.

SNAPSHOT

Business Experiments on the Web?

We all know the Internet is useful for data collection. But experiments? Web samples are usually not representative of the populations we want to make inferences about. There are other concerns: Web access, uninvited participants, multiple trials by the same individual, "team" responses, distracting environments, and the lack of a probability sample for statistical inference. All of these issues currently stir debate. However, for some business studies, there are advantages to using a Web experiment.

Eric DeRosia, a marketing PhD student at the University of Michigan, has devised the "e-Experiment." According to DeRosia, "The software's purpose_is to facilitate primary research over the Web in fields such as psychology and consumer behavior." He claims advantages and disadvantages over traditional laboratory experiments. In a pilot study of 125 marketing students who were randomly assigned to a supervised group lab and an outside lab on the Web, responses to stimuli from attitude and brand honesty scales revealed strikingly similar reliability coefficients. In some ways, DeRosia notes, Web experiments provide more control over stimulus timing, response code verification (out-of-range responses), and participants who peek ahead or change previous answers. By randomly assigning participants to experimental treatments and by controlling which questions are presented and their order, many objections to Web experiments are tackled. For e-Experiment to work properly, it needs a small CGI program on a Web server. Such programs pose security risks for universities and businesses, thus requiring a third-party vendor to host the research. Future programming will make this unnecessary.

In the meantime, other software is being released that continues the promise of e-Experiment: rapid data collection, graphical interface, open- and closed-question programming, randomization, estimation of participant loss rates, response time calculation, and authentication of participation (for rewards or incentives).

www-personal.umich.edu/~ederosia/e-exp/

Field Experiments: Quasi- or Semi-Experiments[8]

Under field conditions, we often cannot control enough of the extraneous variables or the experimental treatment to use a true experimental design. Because the stimulus condition occurs in a natural environment, a **field experiment** is required.

A modern version of the bystander and thief field experiment, mentioned at the beginning of the chapter, involves the use of electronic article surveillance to prevent shrinkage due to shoplifting. In a proprietary study, a shopper came to the optical counter of an upscale mall store and asked the salesperson to see special designer frames. The salesperson, a confederate of the experimenter, replied that she would get them from a case in the adjoining department and disappeared. The "thief" selected two pairs of sunglasses from an open display, deactivated the security tags at the counter, and walked out of the store.

Thirty-five percent of the subjects (store customers) reported the theft upon the return of the salesperson. Sixty-three percent reported it when the salesperson asked about the shopper. Unlike previous studies, the presence of a second customer did not reduce the willingness to report a theft.

This study was not possible with a control group, a pretest, or randomization of customers, but the information gained was essential and justified a compromise of true experimental designs. We use the pre-experimental designs previously discussed or quasi-experiments to deal with such conditions. In a quasi-experiment, we often cannot know when or to whom to expose the experimental treatment. Usually, however, we can decide when and whom to measure. A quasi-experiment is inferior to a true experimental design but is usually superior to pre-experimental designs. In this section, we consider a few common quasi-experiments.

Nonequivalent Control Group Design This is a strong and widely used quasi experimental design. It differs from the pretest–post-test control group design, because the test and control groups are not randomly assigned. The design is diagrammed as follows:

$$\begin{array}{ccc} O_1 & X & O_2 \\ \hdashline O_3 & & O_4 \end{array} \qquad (10)$$

There are two varieties. One is the *intact equivalent design,* in which the membership of the experimental and control groups is naturally assembled. For example, we may use different classes in a school, membership in similar clubs, or customers from similar stores. Ideally, the two groups are as alike as possible. This design is especially useful when any type of individual selection process would be reactive.

The second variation, the *self-selected experimental group design,* is weaker because volunteers are recruited to form the experimental group, while nonvolunteer subjects are used for control. Such a design is likely when subjects believe it would be in their interest to be a subject in an experiment—say, an experimental training program.

Comparison of pretest results ($O_1 - O_3$) is one indicator of the degree of equivalence between test and control groups. If the pretest results are significantly different, there is a real question about the groups' comparability. On the other hand, if pretest observations are similar between groups, there is more reason to believe internal validity of the experiment is good.

Separate Sample Pretest-Post-Test Design This design is most applicable when we cannot know when and to whom to introduce the treatment but we can decide when and whom to measure. The basic design is:

$$\begin{array}{llll} R & O_1 & (X) & \\ R & & X & O_2 \end{array} \qquad (11)$$

The bracketed treatment (X) is irrelevant to the purpose of the study but is shown to suggest that the experimenter cannot control the treatment.

This is not a strong design because several threats to internal validity are not handled adequately. History can confound the results but can be overcome by repeating the study at other times in other settings. In contrast, it is considered superior to true experiments in external validity. Its strength results from its being a field experiment in which the samples are usually drawn from the population to which we wish to generalize our findings.

We would find this design more appropriate if the population were large, if a before-measurement were reactive, or if there were no way to restrict the application of the treatment. Assume a company is planning an intense campaign to change its employees' attitudes toward energy conservation. It might draw two random samples of employees, one of which is interviewed about energy use attitudes before the information campaign. After the campaign the other group is interviewed.

Group Time Series Design A time series design introduces repeated observations before and after the treatment and allows subjects to act as their own controls. The single treatment group design has before-after measurements as the only controls. There is also a multiple design with two or more comparison groups as well as the repeated measurements in each treatment group.

The time series format is especially useful where regularly kept records are a natural part of the environment and are unlikely to be reactive. The time series approach is also a good way to study unplanned events in an ex post facto manner. If the federal government would suddenly begin price controls, we could still study the effects of this action later if we had regularly collected records for the period before and after the advent of price control.

The internal validity problem for this design is history. To reduce this risk, we keep a record of possible extraneous factors during the experiment and attempt to adjust the results to reflect their influence.

Close-Up

A Job Enrichment Quasi-Experiment[9]

One theory of job attitudes holds that "hygiene" factors, which include working conditions, pay, security, status, interpersonal relationships, and company policy, can be a major source of dissatisfaction among workers but have little positive motivational power. This theory says that the positive motivator factors are intrinsic to the job; they include achievement, recognition for achievement, the work itself, responsibility, and growth or advancement.[10]

A study of the value of job enrichment as a builder of job satisfaction was carried out with laboratory technicians, or "experimental officers" (EOs), at British Chemical. The project was a multiple group time series quasi-experiment. The project is diagrammed at the end of this "Close-Up."

Two sections of the department acted as experimental groups and two sections acted as control groups. It is not clear how these groups were chosen, but there was no mention of random assignment. One of the experimental groups and one of the control groups worked closely together, while the other two groups were separated geographically and were engaged in different research. Hygiene factors were held constant during the research, and the studies were kept confidential to avoid the tendency of participants to act in artificial ways.

A before-measurement was made using a job reaction survey instrument. This indicated the EOs typically had low morale, and many wrote of their frustrations. All EOs were asked to write monthly progress reports, and these were used to assess the quality of their work. The assessment was made against eight specifically defined criteria by a panel of three managers who were not members of the department. These assessors were never told which laboratory technicians were in the experimental group and which were in the control group.

The study extended over a year, with the treatments introduced in the experimental groups at the start of the 12-month study period. Changes were made to give experimental group EOs important chances for achievement; these changes also made the work more challenging. Recognition of achievement was given, authority over certain aspects was increased, new managerial responsibilities were assigned to the senior EOs, added advancements were given to others, and the opportunity for self-initiated work was provided. After about six months, these same changes were instituted with one of the control groups, while the remaining group continued for the entire period as a control. Several months of EO progress reports were available as a prior base line for evaluation. The results of this project are shown in Exhibit 14–4.

O O O X O O O O O O O O O O O O

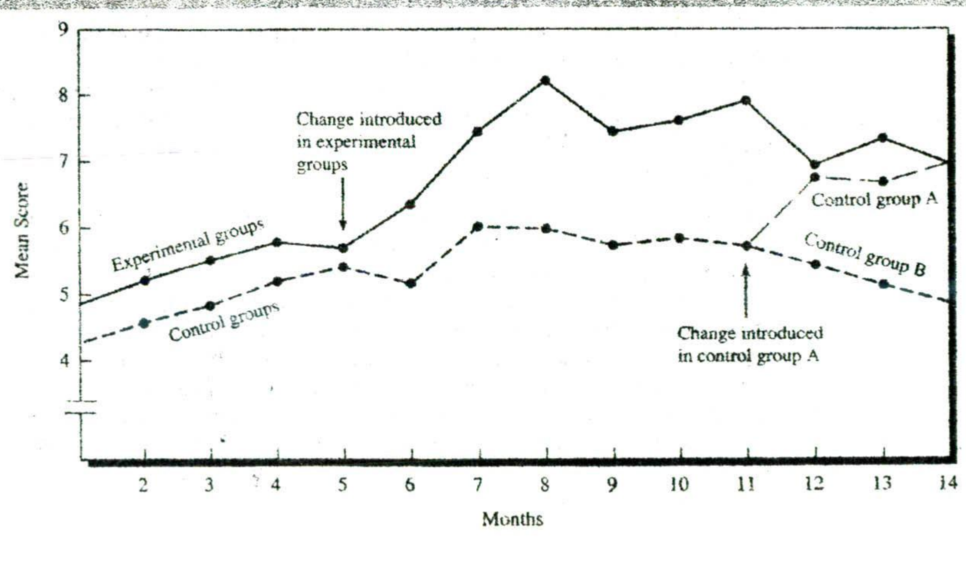
- -

O O O O O O O O O X O O O O O O

- -

O O O O O O O O O O O O O O O O

EXHIBIT 4–4 Assessment of EOs' Monthly Reports



SUMMARY

1 Experiments are studies involving intervention by the researcher beyond that required for measurement. The usual intervention is to manipulate a variable (the independent variable) and observe how it affects the subjects being studied (the dependent variable).

An evaluation of the experimental method reveals several advantages: (1) the ability to uncover causal relationships, (2) provisions for controlling extraneous and environmental variables, (3) convenience and low cost of creating test situations rather than searching for their appearance in business situations, (4) the ability to replicate findings and thus rule out idiosyncratic or isolated results, and (5) the ability to exploit naturally occurring events.

2 Some advantages of other methods that are liabilities for the experiment include: (1) the artificial setting of the laboratory, (2) generalizability from nonprobability samples, (3) disproportionate costs in select business situations, (4) a focus restricted to the present and immediate future, and (5) ethical issues related to the manipulation and control of human subjects.

3 Consideration of the following activities is essential for the execution of a well-planned experiment:

1. Select relevant variables for testing.
2. Specify the levels of treatment.
3. Control the environmental and extraneous factors.
4. Choose an experimental design suited to the hypothesis.
5. Select and assign subjects to groups.
6. Pilot-test, revise, and conduct the final test.
7. Analyze the data.

4 We judge various types of experimental research designs by how well they meet the tests of internal and external validity. An experiment has high internal validity if one has confidence that the experimental treatment has been the source of change in the dependent variable. More specifically, a design's internal validity is judged by how well it meets seven threats. These are history, maturation, testing, instrumentation, selection, statistical regression, and experiment mortality.

External validity is high when the results of an experiment are judged to apply to some larger population. Such an experiment is said to have high external validity regarding that population. Three potential threats to external validity are testing reactivity, selection interaction, and other reactive factors.

5 Experimental research designs include (1) pre-experiments, (2) true experiments, and (3) quasi-experiments. The main distinction among these types is the degree of control that the researcher can exercise over validity problems.

Three pre-experimental designs were presented in the chapter. These designs represent the crudest form of experimentation and are undertaken only when nothing stronger is possible. Their weakness is the lack of an equivalent comparison group; as a result, they fail to meet many internal validity criteria. They are the (1) one-shot control study, (2) one-group pretest–post-test design, and (3) static group comparison.

Two forms of the true experiment were also presented. Their central characteristic is that they provide a means by which we can assure equivalence between experimental

and control groups through random assignment to the groups. These designs are (1) pretest–post-test control group and (2) post-test-only control group.

The classical two-group experiment can be extended to multigroup designs in which different levels of the test variable are used as controls rather than the classical nontest control. In addition, the true experimental design is extended into more sophisticated forms that use blocking. Two such forms, the randomized block and the Latin square, were discussed. Finally, the factorial design was discussed in which two or more independent variables can be accommodated.

Between the extremes of pre-experiments, with little or no control, and true experiments, with random assignment, there is a gray area in which we find quasi-experiments. These are useful designs when some variables can be controlled, but equivalent experimental and control groups usually cannot be established by random assignment. There are many quasi-experimental designs, but only three were covered in this chapter: (1) nonequivalent control group design, (2) separate sample pretest–post-test design, and (3) group time series design.

KEY TERMS

active factors 439
blind 429
blocking factors 439
control group 428
dependent variable (DV) 425
double blind 429
environmental control 429
experimental treatment 428
experiments 425
external validity 432
factor 439
field experiment 443
hypothesis 428
independent variable (IV) 425
interaction effect 441
internal validity 432
main effect 441
matching 430
operationalized 428
quota matrix 430
random assignment 430
replication 427
test unit 439
treatment levels 428

EXAMPLES

Company	Scenario	Page
ACNielsen	Ad recall survey and purchasing behavior data drawn from its Household Scanner Panel™ served as the test and control groups for the Magazine Publishers of America study on ad effectiveness.	430
British Chemical*	A study of the value of job enrichment as a builder of job satisfaction.	445
FLOORgraphics, Inc.	Using test and control groups to determine the effectiveness of FLOORads.	439
Hill Top Research, Inc.	Human odor detectives employed to test the effectiveness of underarm deodorants.	440
Magazine Publishers of America (MPA)	Conducted a study to prove the sales lift following magazine advertising using the purchasing behavior of a panel.	430
Philips Electronics North America	Sponsoring a research study to encourage employees to increase their 401k savings rate.	437
Research House Inc.	A Canadian firm offering a variety of facilities for different types of studies.	426

Top Cannery*	An experiment to determine the ideal price difference between private and national brands.	439
Vanguard	Conducting a study to determine whether employees will save more in 401k savings if their savings increase comes from future raises rather than current earnings.	437

*Due to the confidential and proprietary nature of most research, the names of some companies have been changed.

DISCUSSION QUESTIONS

Terms in Review

1. Distinguish between the following:
   a. Internal validity and external validity.
   b. Pre-experimental design and quasi-experimental design.
   c. History and maturation.
   d. Random sampling, randomization, and matching.
   e. Active factors and blocking factors.
   f. Environmental variables and extraneous variables.
2. Compare the advantages of experiments with the advantages of survey and observational methods.
3. Why would a noted business researcher say, "It is essential that we always keep in mind the model of the controlled experiment, even if in practice we have to deviate from an ideal model"?
4. What ethical problems do you see in conducting experiments with human subjects?
5. What essential characteristics distinguish a true experiment from other research designs?

Making Research Decisions

6. A lighting company seeks to study the percentage of defective glass shells being manufactured. Theoretically, the percentage of defectives is dependent on temperature, humidity, and the level of artisan expertise. Complete historical data are available for the following variables on a daily basis for a year:
   a. Temperature (high, normal, low).
   b. Humidity (high, normal, low).
   c. Artisan expertise level (expert, average, mediocre).

   Some experts feel that defectives also depend on production supervisors. However, data on supervisors in charge are available for only 242 of the 365 days. How should this study be conducted?
7. Describe how you would operationalize variables for experimental testing in the following research question: What are the performance differences between 10 microcomputers connected in a local-area network (LAN) and one minicomputer with 10 terminals?
8. A pharmaceuticals manufacturer is testing a drug developed to treat cancer. During the final stages of development the drug's effectiveness is being tested on individuals for different (1) dosage conditions and (2) age groups. One of the problems is patient mortality during experimentation. Justify your design recommendations through a comparison of alternatives and in terms of external and internal validity.
   a. Recommend the appropriate design for the experiment.
   b. Explain the use of control groups, blinds, and double blinds if you recommend them.
9. You are asked to develop an experiment for a study of the effect that compensation has on the response rates secured from personal interview subjects. This study will involve 300

people who will be assigned to one of the following conditions: (1) no compensation, (2) $1 compensation, and (3) $3 compensation. A number of sensitive issues will be explored concerning various social problems, and the 300 people will be drawn from the adult population. Describe how your design would be set up if it were (a) a completely randomized design, (b) a randomized block design, (c) a Latin square, and (d) a factorial design (suggest another active variable to use). Which would you use? Why?

10. What type of experimental design would you recommend in each of the following cases? Suggest in some detail how you would design each study:

a. A test of three methods of compensation of factory workers. The methods are hourly wage, incentive pay, and weekly salary. The dependent variable is direct labor cost per unit of output.

b. A study of the effects of various levels of advertising effort and price reduction on the sale of specific branded grocery products by a retail grocery chain.

c. A study to determine whether it is true that the use of fast-paced music played over a store's public address system will speed the shopping rate of customers without an adverse effect on the amount spent per customer.

Bringing Research to Life

11. Identify an experiment done with subjects who are twins that would meet the criteria that Dorrie feels must be met for experimentation.

From Concept to Practice

12. Using Exhibit 14–3, diagram an experiment described in one of the Snapshots featured in this chapter using research design symbols.

13. For experiments and surveys on the Web, visit http://www.psych.upenn.edu/links.html#webexpts and participate in an online experiment. Prepare a short paper describing your experience and make suggestions for improving the experimental design.

WWW Exercises

Visit our website for Internet exercises related to this chapter at www.mhhe.com/business/cooper8

CASES*

ENVIROSELL, INC.

RETAILERS UNHAPPY WITH DISPLAYS FROM MANUFACTURERS

GOODYEAR'S AQUATRED

*All cases indicating a video icon are located on the Instructor's Videotape Supplement. All nonvideo cases are in the case section of the textbook. All cases indicating a CD icon offer a data set, which is located on the accompanying CD.

REFERENCE NOTES

1. Bibb Latane and J. M. Darley, *The Unresponsive Bystander: Why Doesn't He Help?* (New York: Appleton-Century-Crofts, 1970), pp. 69–77. Research into the responses of bystanders who witness crimes was stimulated by an incident in New York City where Kitty Genovese was attacked and killed in the presence of 38 witnesses who refused to come to her aid or summon authorities.
2. This section is largely adapted from Julian L. Simon and Paul Burstein, *Basic Research Methods in Social Science*, 3rd ed. (New York: Random House, 1985), pp. 128–33.
3. For a thorough explanation of this topic, see Helena C. Kraemer and Sue Thiemann, *How Many Subjects? Statistical Power Analysis in Research* (Beverly Hills, CA: Sage Publications, 1987).
4. Kenneth D. Bailey, *Methods of Social Research*, 2nd ed. (New York: Free Press, 1982), pp. 230–33.
5. The concept of a quota matrix and the tabular form for Exhibit 14–2 were adapted from Earl R. Babbie, *The Practice of Social Research*, 5th ed. (Belmont, CA: Wadsworth, 1989), pp. 218–19.

6. Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1963), p. 5.
7. Thomas D. Cook and Donald T. Campbell, "The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings," in *Handbook of Industrial and Organizational Psychology*, ed. Marvin D. Dunnette (Chicago: Rand McNally, 1976), p. 223.
8. For an in-depth discussion of many quasi-experiment designs and their internal validity, see ibid., pp. 246–98.
9. William J. Paul, Jr., Keith B. Robertson, and Frederick Herzberg, "Job Enrichment Pays Off," *Harvard Business Review* (March–April 1969), pp. 61–78.
10. Frederick J. Herzberg, "One More Time: How Do You Motivate Employees?" *Harvard Business Review* (January–February 1968), pp. 53–62.

REFERENCES FOR SNAPSHOTS AND CAPTIONS

e-Experiments

Eric DeRosia, "True Experiments on the Web," Working Paper 99-021 (Ann Arbor, MI: University of Michigan Business School) (http://www-personal.umich.edu/~ederosia/e-exp).

FLOORad

Antonia DeMatto, FLOORgraphics, Inc.

"Raising the Roof with Floor Ads," *Business Week*, September 16, 1999 (http://www.businessweek.com:/smallbiz/news/coladvice/reallife/r1990916.htm?scriptFramed).

"Floor Show: Savvy Ideas to Boost Sales," *Entrepreneur* (http://www.entrepreneur.com/Magazines/MA_SegArticle/0,1539,227019---2-,00.html).

Hill Top Research

"Good Morning America," ABC News, March 3, 2000 (www.hill-top.com/consumer.html).

Vanguard/Philips

Richard H. Thaler and Shlomo Benartzi, "Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving," Working Paper, August 2001.

Louis Uchitelle, "Economic View: Why It Takes Psychology to Make People Save," *The New York Times*, January 13, 2002, p. 4, sec. 3.

Steve Utkus, Vanguard, interview on February 4, 2002.

MPA

Lorraine Calvacca, "Making a Case for the Glossies," *American Demographics*, July 1999, pp. 36–37 (http://www.magazine.org/resources/downloads/Sales_Scan_Highlights.pdf).

CLASSIC AND CONTEMPORARY READING

Campbell, Donald T., and M. Jean Russo. *Social Experimentation.* Thousand Oaks, CA: Sage Publishing, 1998. The evolution of the late Professor Campbell's thinking on validity control in experimental design.

Campbell, Donald T., and Julian C. Stanley. *Experimental and Quasi-Experimental Designs for Research.* Chicago: Rand McNally, 1963. A universally quoted discussion of experimental designs in the social sciences.

Cook, Thomas D., and Donald T. Campbell. "The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings." In *Handbook of Industrial and Organizational Psychology.* 2nd ed., edited by Marvin D. Dunnette and Leaetta M. Hough. Palo Alto, CA: Consulting Psychologists Press, 1990. *Quasi-Experimentation: Design and Analysis Issues for Field Settings.* Chicago: Rand McNally, 1979. Major authoritative works on both true and quasi-experiments and their design. Already classic references.

Edwards, Allen. *Experimental Design in Psychological Research.* 4th ed. New York: Holt, Rinehart & Winston, 1972. A complete treatment of experimental design with helpful illustrative examples.

Green, Paul E., Donald S. Tull, and Gerald Albaum. *Research for Marketing Decisions.* 5th ed. Englewood Cliffs, NJ: Prentice-Hall, 1988. A definitive text with sections on the application of experimentation to marketing research.

Kirk, Roger E. *Experimental Design: Procedures for the Behavioral Sciences.* 3rd ed. Belmont, CA: Brooks/Cole, 1994. An advanced text on the statistical aspects of experimental design.

Krathwohl, David R. *Social and Behavioral Science Research: A New Framework for Conceptualizing, Implementing, and Evaluating Research Studies.* San Francisco: Jossey-Bass, 1985. Chapters 3, 4, and 5 present a convincing argument for reformulating internal and external validity into broader concepts. A conceptually refreshing approach.